ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

COMPUTER SCIENCE AND ENGINEERING

Ciclo 38

Settore Concorsuale: 01/B1 - INFORMATICA

Settore Scientifico Disciplinare: INF/01 - INFORMATICA

FROM CARS TO CLINICS: INTERPRETABLE MACHINE LEARNING FOR LUXURY AUTOMOTIVE LIFECYCLE AND CLINICAL CASE STUDIES

Presentata da: Alessandro Ghibellini

Coordinatore Dottorato

Paola Salomoni

Supervisore

Maurizio Gabbrielli

Co-supervisori

Luciano Bononi
Mauro Coletto

Esame finale anno 2026

Finanziato dall'Unione europea- Next Generation EU, Missione 4, Componente 2, Investimento 3.3 (D.M. 117/2023) CUP J33C22001400009

Acknowledgement

Intraprendere un dottorato di ricerca significa porsi l'obiettivo di raggiungere le frontiere della conoscenza: una sfida ambiziosa, carica di oneri e onori, che per me è stata fonte di immense soddisfazioni e, inevitabilmente, di profondi pensieri.

Il merito più grande dei miei supervisori, **Maurizio** e **Luciano**, è stato quello di aver saputo guidare la mia curiosità verso i temi portanti di questa tesi senza mai caricarmi di eccessive pressioni, permettendomi di apprezzare appieno ogni traguardo raggiunto. La loro esperienza è stata per me una bussola costante. Un pilastro fondamentale di questo percorso è stata **Ferrari**, l'azienda che ha co-finanziato la mia ricerca. Il mio grazie va innanzitutto a **Mauro**, per avermi accolto nel suo team, e ai colleghi **Andrea**, **Paolo** e **Simone**, compagni di viaggio preziosi. Oltre alla sfida tecnica legata al *Machine Learning*, ho avuto l'opportunità di immergermi in un business unico come quello dell'auto di lusso. Imparare da professionisti esperti è stato un privilegio: ringrazio **Claudia** e **Andrea M.** per il Marketing; **Andrea S.** per la pazienza e le ore dedicate a spiegarmi le dinamiche del mercato *Pre-owned*; e l'intero team *Personalizzazioni* che mi ha ospitato. In particolare, un ringraziamento a **Marco** e **Andrea B.** per l'intuizione iniziale, e a **Marialaura**, **Eleonora** e **Jessica** per il supporto quotidiano.

Il versante medico della ricerca ha beneficiato della competenza del team R&D di **Isokinetic**. Grazie a **Francesco DV** e **Stefano** per la costante fiducia e supporto, e ad **Eileen** per la cura premurosa con cui ha seguito il mio lavoro (per i tanti tè e tisane). Ringrazio inoltre il team di **Neurologia dell'ospedale di Reggio Emilia** partendo da **Francesco C.** e, in particolare, **Giulia** per le innumerevoli ore spese insieme a sviluppare i progetti che hanno arricchito questo percorso.

L'ultima fase di questo viaggio mi ha portato oltreoceano. Grazie a **Tommaso** per aver reso possibile il mio periodo alla **Cornell Tech** di New York, e a tutto il team **Ferrari North America (FNA)** per il supporto totale. Un grazie speciale a **Ivano** per l'accoglienza, a **Luca** per la fiducia, a **Carlotta** per gli insegnamenti e a **Marco** e **Federica**, a cui devo moltissimo. Grazie anche a **Patrizio** ed **Enzo** per i preziosi consigli e a **Lorenzo** per le corse del sabato a Central Park! Infine, il ringraziamento più importante va alla mia **famiglia**, che mi ha sopportato e supportato costantemente, facendomi sentire la sua vicinanza anche dall'altra parte dell'oceano. Soprattutto **Mamma**, grazie per esserci sempre stata. Grazie alla mia **ragazza**, che ha saputo condividere con me questo percorso nonostante le distanze. Grazie ai miei **amici di sempre**, che non sono mai mancati anche quando la testa era altrove, e ai **nuovi amici** incontrati a New York, che hanno saputo farmi sentire a casa in una città così diversa dalle mie abitudini.

Sono certo di aver dimenticato qualcuno in queste righe, e me ne scuso, ma spero di avere presto l'occasione di ringraziarvi di persona. Grazie a tutti voi.

Abstract

In the luxury automotive market, strategic success increasingly depends on three tightly connected dimensions of the customer lifecycle: personalization, retention, and residual value. Personalization strengthens emotional attachment and reinforces exclusivity, but its benefits endure only if customers remain engaged throughout the long waiting periods typical of bespoke production. Strong retention, in turn, sustains residual value by stabilizing demand and preserving brand desirability. High residual values then further stimulate new purchases: existing clients can reinvest the value of their previous cars into new models, and potential buyers are attracted by the brand's proven long-term value. Together, these mechanisms form a self-reinforcing cycle that ties customer experience, market perception, and financial performance.

This thesis was developed within that ecosystem, thanks to the opportunity to work directly with a leading automotive manufacturer and to experience operations within one of its largest markets, the United States. This dual perspective, combining corporate-level exposure with day-to-day business observation in a complex, competitive environment, provided access to extensive first-party data and a practical understanding of how analytical insights translate into strategic and operational decisions across marketing, personalization, and after-sales domains.

Our first contribution is a hybrid recommender system that adapts collaborative and content-based learning to the domain of highly configurable vehicles. Trained on over 10,000 customer profiles and 900 configuration options, it achieves Precision@5 = 0.65, NDCG@5 = 0.66, and Hit Rate 1.00, with a Mean Inter-List Diversity 0.81. The scalable XGBoost architecture, validated through real-world deployments in Atelier sessions, confirms its robustness and practical impact.

The second contribution introduces an interpretable retention model that predicts customer churn during long waiting periods. A CatBoost classifier trained on approximately 50,000 historical contracts, achieves F1 0.85 (0.93 hold-out), with SHAP explanations identifying key drivers such as waiting time, inflation, and country market. Beyond performance, the framework takes advantage of an LLM to translate predictions into actionable insights for dealerships.

The third contribution focuses on residual value estimation, integrating vehicle attributes with macroeconomic indicators (e.g., CPI, interest rates, unemployment) considering a period of 10 years and more than 20k transactions. The model achieves a mean absolute percentage error of 5.6%, describes the impact of all features on RV, and introduces the concept of a standard vehicle to disentangle genuine market trends from sample variability.

Beyond the automotive domain, the same interpretable and domain-adapted methodology is extended to Clinical Case studies, where data share similar challenges: limited sample sizes, high heterogeneity, and sparse feature spaces. Just as luxury markets involve few but highly distinctive customers, clinical datasets capture rare conditions and individual variability.

In neurology, we apply the framework to the prediction of GBA1 mutation status in Parkinson's disease. The model combines XGBoost with SHAP explanations to achieve approximately 73% accuracy, trained on a matched cohort of more than 100 patients. It reveals clinical and cognitive patterns that differentiate genetic subtypes. These interpretable outputs allow clinicians to identify potential mutation carriers based on routine clinical data, supporting more targeted genetic testing and personalized patient management.

In sports science, we analyze footballers' change-of-direction biomechanics using unsupervised clustering and supervised classification. The dataset comprises motion-capture and force-plate recordings from more than 1,000 athletes, collected through high-speed cameras and a calibrated force platform. The resulting model identifies four distinct movement phenotypes, ranging from high-force/low-control to high-force/high-control profiles, achieving a macro-F1 of about 0.92. This condensed, interpretable representation enables real-time monitoring of performance and injury risk, providing actionable insights for coaches and medical staff in the prevention of anterior cruciate ligament (ACL) injuries.

Taken together, these contributions show that interpretable, deployment-oriented ML can turn first-party data into actionable decision support across the customer lifecycle: personalization, retention, and residual value, while generalizing to small, heterogeneous, high-stakes datasets in clinical and sports contexts. The common recipe is learning under scarcity: encode domain structure, constrain model complexity, and pair predictions with transparent, human-in-the-loop interfaces. Beyond headline metrics, the emphasis on explanation and workflow integration makes outputs auditable and operational. Future work will prioritize prospective validation and causal evaluation of interventions (e.g., dealer outreach, targeted genetic testing, phenotype-specific training) and unify the modules related to the customer journey into a single decision-support layer that closes the loop between data, insight, and outcome.

Contents

List of Figures

List of Tables

Part I

Background

Chapter 1

Introduction

The *luxury market* refers to the segment of the economy comprising high-end goods and experiences that are non-essential yet highly desirable, characterized by exceptional quality, exclusivity, craftsmanship, and often a premium price. According to Bain & Company, the overall luxury market encompasses both luxury goods and experiences across nine segments, including luxury cars, personal luxury goods, luxury hospitality, fine wines and spirits, gourmet food and fine dining, high-end furniture and housewares, fine art, private jets and yachts, and luxury cruises, together accounting for approximately 80% of the total market by value [1]. The *luxury car market* encompasses the industry segment focused on sales of premium passenger vehicles, such as luxury sedans, SUVs, and compact luxury cars, characterized by superior materials, advanced technology, enhanced comfort, exceptional performance, and elevated brand image. According to McKinsey, luxury cars (defined as vehicles valued at over USD 150,000) represent only a small share of total vehicle sales but wield outsized influence on brand prestige and profitability [2]. Sales of luxury cars, the largest segment of the luxury industry, fell by 5% at current exchange rates in 2024, totaling approximately €579 billion, yet still remaining above 2019 levels. The market exhibited notable polarization across price tiers: demand softened in the upper-premium bracket as local competitors intensified competition in Asian markets, while the absolute luxury segment held firm, powered by strong consumer demand for ultra-personalization [3]. A support of this, Ferrari, a top-tier player in the luxury market, saw personalization account for approximately 20 % of total revenue in 2024 [4], as emphasized by

CEO Benedetto Vigna, who attributed the company's 11.8 % net revenue growth to a strong product mix and growing demand for personalization [5]. Beyond personalization, luxury automakers also strengthen their brand equity through deliberate production constraints. As KPMG highlights, exclusivity and scarcity are foundational to the definition of luxury [6]. Maintaining limited production runs and waitlists creates an aura of rarity, exactly the kind of exclusivity that branding experts endorse: "a smaller, coveted, exclusive brand" [7]. Marketing analyses further explain that artificial scarcity not only drives up secondary market prices but also reinforces the prestige of luxury brands [8]. Academic evidence from the automobile industry confirms its effectiveness: vehicles with intentionally low introductory inventory generate higher consumer preference, suggesting that scarcity is a deliberate lever rather than an incidental outcome [9]. As a direct consequence of deliberate scarcity and the resulting long waiting lists, luxury automakers face the distinct challenge of retaining customer interest throughout the pre-delivery period. Research from S&P Global Mobility suggests that loyalty rates among luxury consumers lag behind mainstream brands, indicating heightened vulnerability to churn during extended wait times [10]. To mitigate this, forward-looking OEMs and dealerships are turning to personalized, data-driven engagement strategies [11]. Another central dimension in the luxury automotive lifecycle is the cost of ownership, particularly when vehicles are financed through leasing agreements. In this context, residual value (RV), the forecasted market value of a vehicle at the end of its lease or ownership cycle, plays a decisive role. A higher RV reduces the total cost of ownership for buyers and lowers monthly lease payments, making vehicles more financially accessible while preserving brand prestige [12]. Consulting analyses emphasize that residual value management has become a key competitive lever in the automotive sector: manufacturers that succeed in sustaining strong RVs not only improve customer satisfaction but also increase the attractiveness of leasing offers, thereby expanding their client base [13]. Moreover, high RVs reinforce the perception of long-term brand value, as luxury vehicles that depreciate less are regarded as both more desirable and more financially rational investments [14]. Beyond market dynamics, Artificial Intelligence (AI) can be broadly defined as the capability of machines and computer systems to perform tasks that typically require human intelligence, including perception, reasoning, learning, and decision-making [15]. In the automotive sector, AI is increasingly viewed as a transformative force with the potential to reshape the entire value chain. According to McKinsey, AI-driven applications are set to unlock up to $400 billion of annual value in the mobility industry by 2030, spanning areas from product personalization to predictive maintenance and connected services [16]. Deloitte similarly emphasizes that AI will fundamentally alter the customer journey, empowering automakers to optimize residual value management, enhance loyalty during long waiting periods, and deliver hyper-personalized experiences at scale [17]. More broadly, BCG notes that the integration of AI and advanced analytics represents one of the most disruptive shifts facing the luxury automotive market, changing how vehicles are designed, marketed, financed, and retained across their lifecycle [18]. In this context, the thesis focuses on applying Machine Learning (ML) techniques to the most important journeys of the luxury market: personalization, retention, and residual value. ML can be defined as a subset of AI that enables systems to automatically learn from data and improve their performance over time without being explicitly programmed [19]. Prior research has highlighted that configuring a high-end vehicle is inherently a highly personalized process, as nearly every car is uniquely tailored from hundreds of optional features [20]. Recommender systems, while widely adopted in domains such as e-commerce and media [21, 22], have been less explored in the context of big-ticket configurable products like luxury automobiles. Early works on recommendation-based product configurators demonstrated that data-driven suggestions can simplify complex choice processes and increase conversion rates [20, 23]. Luxury-sector case studies, ranging from bespoke fashion to premium automotive, similarly report higher customer engagement when personalized option advice is available [24, 25]. A major challenge in this context stems from the disjoint nature of option catalogs across vehicle models and the sparsity of historical interaction data, which causes classical collaborative filtering to underperform [26, 27]. Hybrid approaches that combine collaborative and content-based techniques have emerged as the state-of-the-art remedy, enabling cross-model generalization, alleviating cold-start issues, and respecting configuration constraints [28, 29]. More recent advances, including neural collaborative filtering [30] and collaborative deep learning [31], further strengthen the paradigm by improving accuracy in sparse data regimes while preserving scalability. For what concern customer retention, cancellation prediction has been studied extensively in domains such as hospitality, where machine learning and statistical models have been employed to forecast hotel booking cancellations [32, 33, 34]. Techniques explored include traditional ML classifiers [33, 34], Bayesian networks [35], and more recently deep-learning frameworks [36]. While these studies demonstrate strong predictive accuracy, they often trade off interpretability, making them difficult to adapt to other sectors where decision processes and products differ significantly [37]. In the luxury automotive context, where demand consistently outstrips production, effective management of waiting lists becomes crucial for maintaining customer trust and brand equity. Prior studies on customer retention stress the importance of engaging clients throughout the waiting period, using interventions such as personalized communication, targeted events, or proactive contract management to reduce cancellation risk [38]. Finally, RV estimation has been widely studied given its central role in leasing margins, ownership costs, and strategic decision-making in the automotive sector. Early works relied on hedonic pricing models and system-dynamics approaches [39, 40], which provided foundational insights but struggled to capture complex, nonlinear depreciation dynamics. Later methods explored genetic algorithms and neuro-fuzzy inference systems for used-car pricing [41, 42], though these approaches were limited by small datasets and lack of temporal validation. More recent studies consistently demonstrate that machine learning outperforms linear baselines. Random forests and neural networks deliver superior predictive accuracy in large used-car datasets [43, 44, 45], while techniques such as Multivariate adaptive regression spline (MARS) and feature selection further improve forecasts in emerging markets [46, 47]. Recent advances also emphasize interpretability: Bergmann and Feuerriegel [48] applied SHapley Additive exPlanations (SHAP) [49] analysis to granular equipment data, showing which optional features most strongly influence RV, though their study was limited to German mid-market vehicles. The review of the literature across personalization, retention, and residual value reveals several important gaps that motivate the present thesis:

- **Personalization (P-RQ1–P-RQ2).** While recommender systems are well established in e-commerce and media, their use for highly configurable, big-ticket products such as luxury vehicles remains underexplored. Sparse, long-tailed customer–option interactions, fragmented option catalogs, and cold-start clients/options make accuracy, diversity, and maintainability hard to achieve. This motivates **P-RQ1**: can a hybrid recommender that couples collaborative signals with content metadata operate reliably in an Atelier/Configurator under severe sparsity, producing accurate, diverse, and non-obvious suggestions? And **P-RQ2**: can we temper popularity bias while preserving personalization and scaling across model-year/option churn, with explanations that improve dealer–client dialogue?

- **Waiting-List Retention (WL-RQ1–WL-RQ3).** Prior cancellation-prediction work, mostly in hospitality and services, often favors black-box accuracy over interpretability and governance, limiting transfer to luxury automotive where long waits and high client value raise the stakes. This gap motivates **WL-RQ1**: within GDPR/EU-AI-Act constraints, can an interpretable model deliver early, actionable lift on cancellation risk (sufficient for resource-aware intervention)? **WL-RQ2**: which drivers, e.g. lead recency/quality, prior ownership and garage similarity, event engagement, elapsed wait—shape risk, and can global/local attributions support human judgment? **WL-RQ3**: how should outputs be operationalized to improve triage speed, outreach timing, and customer experience in daily dealership workflows?

- Classic econometrics struggles with nonlinear depreciation and interaction effects; recent ML/DL improves accuracy but rarely targets the luxury segment, integrates macro drivers, or separates true trends from shifting sample mix. This motivates **RV-RQ1**: Can non-linear ML models (e.g., CatBoost, TabNet) achieve superior RV estimation accuracy over a 10-year, multi-region luxury dataset compared to linear baselines? Can our model consistently maintain an MAPE below the industry-accepted 10% threshold? **RV-RQ2**:Which features, internal (age, mileage, options) versus macroeconomic (CPI, unemployment, 10-year bond yield), most influence RV, and in which manner do they significantly alter depreciation? **RV-RQ3**: Does the "Standard Vehicle" approach significantly reduce noise in quarterly RV trends caused by sample variability, yielding smoother, more interpretable market shifts?

This thesis contributes to the understanding and advancement of machine learning applications in the luxury automotive sector by addressing three critical dimensions of the customer journey:

- A novel framework is developed for guiding customers through the complex configuration of luxury vehicles, where option catalogs are vast, fragmented, and highly individual. The contribution lies in adapting hybrid machine learning approaches to provide accurate and diverse recommendations across models, while ensuring interpretability and scalability for industrial deployment.

- An interpretable machine learning framework is introduced to proactively identify at-risk customers during extended waiting periods and translate predictions into actionable interventions for dealerships. The contribution is twofold: first, advancing the modeling of cancellation risk in a domain where long delays make churn particularly costly; and second, embedding explainability and compliance into the predictive process, enabling dealerships to strengthen customer engagement and safeguard brand loyalty.

- A comprehensive approach is proposed for estimating and interpreting vehicle residual values in the luxury segment. Beyond achieving state-of-the-art predictive accuracy, the contribution is in quantifying depreciation thresholds and introducing mechanisms to isolate genuine market signals from sample variability. This allows manufacturers and financial stakeholders to better anticipate long-term value dynamics and incorporate external macroeconomic factors into strategic decision-making.

In addition to the automotive setting, this thesis extends its methodological framework to two domains that share a surprisingly similar challenge: clinical sports science and neurology. Like the luxury automotive context, where each client and configuration is unique, clinical datasets are small, heterogeneous, and costly to acquire. In both cases, predictive modeling must learn under scarcity while preserving interpretability, ensuring that results remain actionable for experts rather than opaque to them.

In clinical sports science, one of the most pressing challenges is the prevention of anterior cruciate ligament (ACL) injuries, a problem that persists despite decades of biomechanical research and improved awareness. These injuries often result from a complex interplay of intrinsic and extrinsic risk factors, such as neuromuscular control, joint biomechanics, and indirect contact mechanisms during dynamic movements like cutting or landing [50, 51, 52]. The high incidence of ACL injuries in professional football, particularly among young athletes and females, highlights the urgent need for novel approaches. Machine learning provides a powerful way to analyze high-dimensional biomechanical data and detect latent movement patterns associated with elevated injury risk, offering practical insights for clinicians, physiotherapists, and trainers. This motivates **S-RQ1**: can unsupervised phenotyping of change-of-direction/landing mechanics, followed by a lightweight supervised classifier, distill high-dimensional kinematics and kinetics into *interpretable force-control phenotypes* linked to ACL-injury risk—supporting near-real-time feedback in clinical workflows without sacrificing transparency or practicality? In clinical neurology, genetic variants in the *GBA1* gene represent the most common known genetic risk factor for Parkinson's disease, accounting for 5–30% of cases depending on population and age at onset [53]. Patients with GBA1-related Parkinson's disease (GBA1-PD) tend to show earlier onset, faster progression, and a higher frequency of motor fluctuations and non-motor symptoms compared to non-mutated cases [54]. Yet, the clinical overlap between GBA1-PD and idiopathic PD complicates reliable diagnosis. Conventional statistical models, which assume linear and additive relationships, are poorly suited to capture the complex, nonlinear phenotypic signatures of GBA1 carriers. In this context, machine learning offers an opportunity to model intricate feature interactions, stratify patients with higher accuracy, and provide interpretable outputs that can support clinicians in prioritizing genetic testing and tailoring patient management. This motivates **C-RQ1**: can an interpretable ML pipeline trained on routinely collected clinical measures detect genotype-specific subtypes (e.g., *GBA1*-related PD) in small, heterogeneous cohorts—and return clinician-usable attributions that help prioritize genetic testing and guide stratified management? This thesis contributes to the understanding and advancement of machine learning applications in the Sport Science and Nurology by addressing critical dimensions of the patient journey:

- A complete pipeline was developed for assessing change-of-direction technique in footballers, with the aim of identifying biomechanical phenotypes linked to anterior cruciate ligament (ACL) injury risk. The contribution lies in condensing laboratory-grade biomechanics into a real-time, interpretable decision-support tool through a combination of unsupervised clustering, supervised classification, and GDPR-compliant deployment.
- A lightweight, interpretable ML pipeline was proposed for predicting genotype-specific subtypes of Parkinson's disease, in particular the identification of GBA1 mutation carriers. The contribution consists in coupling XGBoost with SHAP explanations to deliver both predictive performance and clinician-ready interpretability. By prioritizing patients for genetic testing based on routinely collected data, the system illustrates how explainable ML can make precision medicine more cost-effective and accessible.

The remainder of this thesis is organized as follows. Chapter 1 introduces the research context, objectives, and overall structure of the work. Chapter 2 provides the methodological background, including an overview of machine learning (Section 2.1), approaches to model selection and problem formulation (Section 2.2), interpretability methods such as tree-based models with SHAP and dimensionality reduction (Section 2.3), and considerations for model deployment (Section 2.4). Chapter 3 presents the original contributions of the thesis across two domains. The first set of studies focuses on applications in the luxury automotive sector: personalization in luxury car configurators (Section 3.1.1), prediction and mitigation of customer churn during waiting lists (Section 3.1.2), and a methodology for residual value estimation that integrates predictive accuracy, interpretability, and macroeconomic factors (Section 3.1.3). The second set comprises clinical case studies: prediction of GBA1 mutation status in Parkinson's disease patients (Section 3.2.1) and the identification of ACL biomechanical phenotypes with real-time classification models (Section 3.2.2). Finally, Chapter 4 concludes by synthesizing insights across these domains, discussing industrial and clinical implications, and outlining avenues for future research (Section 4.1).

Chapter 2

Machine Learning

Artificial intelligence (AI) can be broadly defined as the discipline that studies the design of agents capable of perceiving their environment and taking actions that maximize the likelihood of achieving specific goals [56]. Machine learning (ML), a subfield of AI, studies computational methods that learn patterns from data to improve performance on well-defined tasks [56]. Deep learning, in turn, is a subset of ML that employs multilayer artificial neural networks to automatically learn hierarchical representations of data, achieving state-of-the-art results in domains such as vision and language [57] (Figure 2.1).

Long before the formalization of ML, Alan Turing anticipated the possibility of intelligent machines in his seminal 1950 paper, where he introduced what is now known as the "Turing Test", a behavioral criterion for assessing machine intelligence based on indistinguishability from human responses [58]. To sidestep vague philosophical debates, Turing proposed an operational criterion in which a

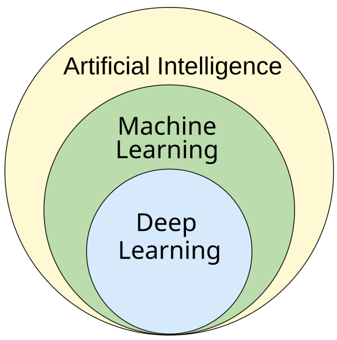

Figure 2.1: Machine learning as a subset of Artificial Intelligence [55]

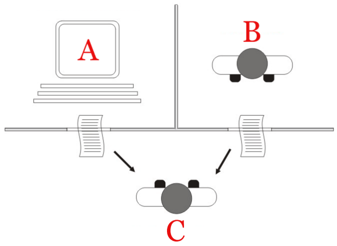

Figure 2.2: Turing Test Diagram [59]

human evaluator interacts through text with both a machine and another human. If the evaluator cannot reliably distinguish between the two, the machine is said to exhibit intelligent behavior (Figure 2.2). The origins of ML trace back to the 1950s. Early pioneers such as Arthur Samuel developed programs that improved their performance in checkers through experience [60]. In the same period, the perceptron model introduced by Rosenblatt formalized one of the first artificial neural networks as a statistical model of learning [61]. These early ideas showed the promise of learning systems, but also highlighted severe limitations, such as the inability of single-layer perceptrons to solve problems like XOR, a criticism famously detailed by Minsky and Papert [62].

In parallel, the Dartmouth Conference of 1956 is often recognized as the symbolic birth of AI as a field, where machine learning was framed as a central challenge [63]. Initial enthusiasm led to inflated expectations, but the technological limitations of the time, restricted computational power, small datasets, and overly optimistic assumptions about symbolic reasoning, triggered periods of disillusionment, known as the "AI winters". The first AI winter in the mid-1970s followed critical evaluations such as the Lighthill Report in the United Kingdom, which highlighted the limited progress and scalability of AI research [64]. In the aftermath of the first AI winter, research and funding shifted toward expert systems, which promised practical applications by codifying human knowledge into rule-based decision frameworks. These systems, such as MYCIN in medical diagnosis, combined a knowledge base with an inference engine to mimic the reasoning of domain specialists [65]. The approach attracted significant governmental and industrial investment during the 1980s, including DARPA's Strategic Computing Initiative and Japan's Fifth Generation Computer Systems project. Although initially successful in narrow domains, expert systems revealed several fundamental limitations. Knowledge acquisition proved slow and costly, leading to the so-called "knowledge acquisition bottleneck." Systems were brittle, failing outside of their narrowly defined rule sets, and their maintenance became prohibitively expensive as knowledge bases grew. Moreover, they lacked the ability to generalize or learn from new data, making them ill-suited to dynamic real-world environments. These structural limitations, combined with poor economic returns, led to waning industrial interest and contributed to the onset of the second AI winter [66]. By the mid-1990s a new paradigm of data-driven approaches triggered what is often called the "machine learning boom." Unlike symbolic systems, these methods learned directly from data and were grounded in solid statistical principles.

Algorithms such as CART [67] and C4.5 [68] became popular for their interpretability and practical performance. By recursively splitting the data according to feature values, decision trees produced human-readable models that were widely applied in classification and regression tasks. The instability of single decision trees motivated the development of ensemble methods. Bagging, introduced by Breiman, reduced variance through bootstrap aggregation, while boosting methods such as AdaBoost sequentially trained weak learners to improve bias. The random forest method [69] combined these ideas, offering robustness, scalability, and competitive accuracy across diverse domains.

In parallel, probabilistic approaches offered a principled way to reason under uncertainty. Bayesian networks provided a graphical representation of dependencies between variables [70], while hidden Markov models became the standard for sequence modeling and speech recognition [71]. These frameworks established the statistical foundations for learning and inference in structured data.

The introduction of support vector machines (SVMs) in the mid-1990s represented a major advance in classification [72]. By maximizing the margin between classes and using kernel functions to capture nonlinear relationships, SVMs achieved state-of-the-art results in text categorization, image recognition, and bioinformatics throughout the 1990s and early 2000s.

While statistical learning methods dominated the 1990s and early 2000s, the mid-2000s marked the resurgence of neural networks in the form of "deep learning". A key breakthrough was Hinton's work on deep belief networks, which demonstrated how unsupervised pretraining could initialize multilayer networks effectively [73]. This development coincided with the increasing availability of graphical processing units (GPUs), which offered massive speedups for the linear algebra operations central to neural network training [74]. At the same time, the explosive growth of digital data provided the scale necessary to train large models.

A defining moment came in 2012, when Krizhevsky, Sutskever, and Hinton achieved a dramatic improvement in the ImageNet competition with their convolutional neural network “AlexNet” [75]. This result demonstrated that deep architectures could outperform all traditional machine learning methods in computer vision. Shortly thereafter, recurrent neural networks and their gated variants, such as LSTMs, achieved breakthroughs in speech and sequence modeling [76, 77]. Finally, the introduction of the Transformer architecture [78] revolutionized natural language processing, enabling large-scale pretraining and the foundation of today’s language models.

Machine learning has proven to be a versatile paradigm, demonstrating its effectiveness across a broad range of application domains. Its adaptability to heterogeneous data sources and problem structures makes it a general-purpose methodology with significant impact:

- **Healthcare:** ML has enabled breakthroughs in early diagnosis and personalized treatment, with deep learning models achieving dermatologist-level performance in skin cancer detection from images [79], and predictive models improving the management of complex diseases such as cardiovascular and neurodegenerative disorders [80].

- **Finance:** Supervised learning is widely used for credit scoring, fraud detection, and risk assessment, often outperforming traditional statistical models. For instance, ensemble methods and gradient boosting have been benchmarked as state-of-the-art solutions for credit scoring [81]. In asset pricing, machine learning methods have also shown superior predictive power compared to linear factor models [82].

- **Transportation:** In autonomous driving, reinforcement learning and computer vision models provide the backbone for perception, decision-making, and control [83]. In addition, traffic prediction and flow optimization tasks benefit from both classical ML and deep learning, enabling more efficient urban mobility [84].

- **Agriculture:** ML techniques are applied to precision farming, crop yield prediction, soil and water monitoring, and early detection of plant diseases, contributing to higher efficiency and sustainability in food production [85].
- **Climate and Energy:** Recent works have demonstrated how machine learning can address climate challenges, including renewable energy forecasting, climate modeling, and emissions monitoring [86].

The cross-domain adaptability of machine learning is also reflected in this thesis. The pipelines originally developed for the luxury automotive sector, covering tasks such as residual value estimation, waitlist cancellation risk, and customer recommendations, were designed with interpretability, robustness, and regulatory compliance in mind. These same methodological choices, proved transferable to the clinical domain. When applied to the classification of genetic Parkinson's disease (GBA1-PD) versus non-mutated cases, the models not only achieved competitive predictive performance but also highlighted clinically meaningful variables, clarifying the direction and strength of their contributions. A further case study in sports science demonstrated how the approach could be adapted to cluster athletes based on change-of-direction metrics, supporting the identification of movement phenotypes associated with higher injury risk. Taken together, these results show that methodologies validated in the complex business environment of the luxury automotive market can be successfully repurposed to advance knowledge and decision-making in healthcare and sports performance.

2.1 Model selection and Problem formulation

Selecting the appropriate methodology is a fundamental step that depends on the nature of the problem, the availability and quality of data, and the desired trade-off between predictive performance and interpretability. Comparative studies emphasize that the choice of algorithm should not rely solely on benchmark accuracy but also on considerations such as computational efficiency, robustness to overfitting, and transparency of the decision process [87]. Moreover, model

Type	Simple	Complicated	Complex
Example	Sit on a chair	Object recognition	Personal hiring
Can success be objectively defined?	Yes	Yes	No
Are the success factors known?	Yes	Yes	No
Is accuracy required?	No	Yes	Non applicabile

Figure 2.3: Type of probelm

selection depends strongly on the type of problem being addressed. As illustrated in Figure 2.3 (inspired from a speech of Benedetto Vigna, 2024), problems can be classified as *simple*, *complicated*, or *complex*. Simple problems, such as "sitting on a chair," are well defined: success can be objectively specified, all relevant factors are known, and accuracy is not critical. For such tasks, straightforward algorithms or even rule-based systems are sufficient.

Complicated problems, such as object recognition, still allow an objective definition of success, and the relevant factors are known, but accuracy is essential. These settings are typical domains for ML/DL, where statistical models and deep learning architectures excel.

Complex problems, such as personal hiring decisions, differ in that success cannot be objectively defined, nor are all success factors known. Accuracy is not directly applicable. These tasks involve ambiguity and context-specific judgment, making them challenging for traditional ML and deep learning methods, and better suited to hybrid human–AI decision-support frameworks.

At the most general level, learning problems can be approached in three main ways: supervised learning, unsupervised learning, and, more recently, reinforcement learning [88]. Supervised learning problems, including regression and classification, involve labeled data and allow both linear and nonlinear models to be applied. Unsupervised learning addresses structure discovery without labels, where clustering or dimensionality reduction methods are commonly used. Reinforcement learning, in contrast, focuses on sequential decision-making under uncertainty, with deep learning and generative models enabling recent breakthroughs in scalability and performance [89]. Concerning model choices, three broad families can be recognized. Linear models, such as linear regression or logistic regression, remain widely used due to their simplicity, efficiency, and interpretability [90]. Classical machine learning models, including tree-based ensembles and kernel methods, provide more flexibility by capturing nonlinear patterns without requiring extensive computational resources [90]. Deep learning methods, based on multi-layer neural networks, achieve state-of-the-art performance in many domains but at the cost of higher complexity, reduced interpretability, and significant computational demands [57]. Beyond traditional families of models, recent years have seen the rapid rise of generative AI, a subset of deep learning that focuses on creating new data instances rather than only performing classification or regression. Generative adversarial networks (GANs), variational autoencoders (VAEs), and large language models (LLMs) exemplify this trend, enabling applications such as synthetic data generation, text generation, and image synthesis [91]. A guiding principle in model selection is to begin with simpler models, which are not only easier to interpret but also more efficient in terms of computation and implementation. Linear and tree-based methods, for instance, generally require fewer resources and provide faster training and inference compared to deep learning architectures. As Domingos [92] highlights, simpler approaches often serve as strong baselines, reducing unnecessary complexity and resource consumption when they achieve comparable performance. Moreover, when the amount of training data is limited, simpler models often generalize better than complex ones, because they are less likely to overfit and require fewer parameters to estimate accurately [93]. Another important consideration in model selection is deployment cost and inference/training speed. Deep learning models often require GPU or accelerator hardware and incur large energy, maintenance, and cloud infrastructure costs. For example, Cottier et al. [94] report that the most compute-intensive DL models (such as GPT-4, Gemini) now cost tens of millions of dollars merely for a single training run. In contrast, many classical ML models (e.g., logistic regression, random forest, SVM) can be trained in minutes or hours on CPU or cheaper hardware, with far lower infrastructure overheads [95]. Our workflow for developing solutions presented in the thesis follows a structured sequence. First, we consult domain experts to formulate the problem precisely and determine its nature. If the task is trivial or well defined, a simple rule-based approach is either adopted or the problem is discarded if no value is added by further modeling. Otherwise, we classify the problem as primarily supervised or unsupervised and establish the overall objectives.

Second, we specify the dataset requirements, including size, quality, and relevance, and agree on the budget in terms of both development time and computational costs for training and inference. The formulation stage also includes defining the evaluation criteria and selecting performance metrics that reflect the real-world objectives.

For modeling, we adopt a progressive strategy: starting with transparent and lightweight models (e.g., linear or logistic regression) and introducing additional complexity only when necessary. More advanced methods (e.g., ensemble learning or deep learning) are considered only if simpler approaches cannot meet the performance requirements. Once satisfactory performance is achieved, no additional complexity is introduced.

Finally, an essential requirement throughout the workflow is model interpretability. While complex models can deliver high predictive performance, their opacity may limit their applicability, particularly in sensitive domains such as healthcare or finance. We therefore prioritize methods that provide transparent explanations of their outputs or can be complemented by post-hoc interpretation techniques. This includes quantifying the contribution of input features, highlighting decision boundaries, and ensuring that model behavior aligns with domain knowledge.

2.2 Interpretability

2.2.1 Tree-based models and SHAP

Model interpretability refers to the extent to which a human can understand the reasoning behind a model's predictions. It is not enough for a machine learning model to achieve high accuracy if its decision process remains opaque. Interpretability provides the ability to examine how input features contribute to the output, making the model's behavior more transparent and trustworthy. Lipton [96] describes interpretability as a collection of related ideas, such as transparency, post-hoc explanations, and fairness, that together help humans form a mental model of how a system works. As Ribeiro et al. [97] argue, explanations are critical: without them, even highly accurate systems may not be adopted, since stakeholders cannot verify or contest their outcomes.

Approaches to interpretability are commonly divided into two broad categories. **Intrinsic interpretability** refers to models whose structure and parameters are directly understandable by humans. Linear regression, logistic regression, and decision trees exemplify this class, since their functional form or branching rules allow a straightforward mapping from inputs to outputs [98]. Such models achieve interpretability by design, although often at the cost of predictive power on highly complex tasks.

By contrast, **post-hoc interpretability** methods aim to explain models that are not intrinsically transparent. These techniques: such as feature importance measures, surrogate models, and local explanation tools like LIME or SHAP, provide approximations or visualizations that help users make sense of black-box predictors [99]. Post-hoc methods are indispensable for modern deep learning and ensemble models, where accuracy is high but the internal decision process is opaque.

Table 2.1: Comparison between intrinsic and post-hoc interpretability approaches.

Approach	Advantages	Limitations
Intrinsic Interpretability	Simple and transparent by design (e.g., linear/logistic regression, decision trees). Easy to communicate to stakeholders.	Limited flexibility and often lower predictive accuracy on highly complex tasks. May not capture nonlinear relationships well.
Post-hoc Interpretability	Allows explanation of black-box models (e.g., ensembles, deep neural networks). Enables use of highly accurate models without sacrificing some level of interpretability.	Explanations are approximations, not necessarily faithful to the true internal logic. Requires additional computation and expertise; may produce inconsistent explanations.

In this dissertation, the focus is on domains where interpretability is not a secondary feature but a fundamental requirement. In luxury automotive applications, tasks such as **personalization**, **customer retention**, and **residual value estimation** directly affect strategic decisions and customer trust. Similarly, in

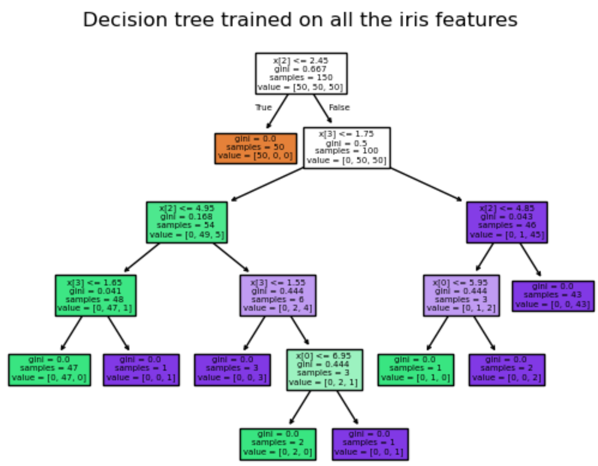

Figure 2.4: Example from Decision tree documentation (Scikit-learn)

sports science, preventing anterior cruciate ligament (ACL) injuries involves high-stakes decisions about athlete health, while in medicine, distinguishing between Parkinson's disease phenotypes represents a domain where black-box predictions without explanations are unacceptable.

For these fields, maximizing **explainability** is prioritized over pursuing marginal improvements in accuracy with opaque methods. Following the guidance of Rudin [98], this thesis emphasizes intrinsically interpretable approaches wherever possible, and applies post-hoc explanation techniques only when necessary to ensure stakeholder trust and regulatory compliance.

In light of this objective, we focus on **tree-based models**. Decision trees are among the earliest and most intuitive machine learning algorithms, representing predictions as a sequence of human-readable rules. Their structure makes them intrinsically interpretable, as it is possible to trace exactly how input features influence the output. However, individual trees often suffer from high variance and limited predictive accuracy when applied to complex real-world data. To overcome these limitations, ensemble methods such as bagging and boosting extend the decision tree paradigm by combining multiple trees into a stronger learner, improving both stability and generalization while retaining interpretability through tools like feature importance and SHAP values. This family of models therefore offers an effective compromise between accuracy and transparency, aligning with the thesis goal of providing reliable and explainable decision-support systems.

Decision trees are among the most intuitive and interpretable predictive models. They partition the input space recursively by selecting the feature and threshold that best split the data according to an impurity measure (e.g. Gini index, entropy). At each node t, the impurity $I(t)$ can be defined, for classification, as:

$$I_{\text{Gini}}(t) = \sum_{k=1}^{K} p_k(1 - p_k), \quad I_{\text{Entropy}}(t) = -\sum_{k=1}^{K} p_k \log p_k, \tag{2.1}$$

where p_k is the proportion of class k among the samples reaching node t. Splitting continues until a stopping criterion is met (e.g. maximum depth, minimum samples per leaf). The resulting model is easy to interpret but prone to high variance.

Random Forests extend decision trees through bagging [69]. Given B bootstrap samples, B trees are grown independently and their predictions averaged (for regression) or aggregated by majority vote (for classification):

$$\hat{f}_{\text{RF}}(x) = \frac{1}{B} \sum_{b=1}^{B} f_b(x), \tag{2.2}$$

where $f_b(x)$ is the prediction of the b-th tree. This reduces variance and improves generalization, while feature importance remains accessible.

XGBoost implements gradient boosting, building trees sequentially to correct errors of the previous ensemble [100]. At iteration t, the model is updated as:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i), \tag{2.3}$$

where η is the learning rate and f_t is the newly added regression tree fitted to the gradient of the loss function. This yields high predictive performance and fine control over regularization.

CatBoost further refines gradient boosting by introducing ordered boosting and efficient encoding of categorical variables [101]. Its main contribution is reducing target leakage in categorical feature handling and mitigating overfitting. Like XGBoost, it can be interpreted using SHAP values and built-in feature importance measures. Decision trees provide a natural way to estimate feature importance, since each split is associated with a reduction in an impurity measure such as the Gini index or entropy. The idea is that the more a feature reduces impurity across the tree, the more important it is for the predictive task.

Formally, let $I(t)$ be the impurity of node t and $N(t)$ the number of samples reaching it. When a node t is split into left (L) and right (R) child nodes using feature j, the *impurity decrease* is:

$$\Delta I_j(t) = N(t)\, I(t) - N(L)\, I(L) - N(R)\, I(R). \tag{2.4}$$

The global importance of feature j in a single tree is the sum of $\Delta I_j(t)$ over all nodes where j is used for splitting, normalized by the total number of samples:

$$FI_j = \frac{1}{N} \sum_{t \in T_j} \Delta I_j(t), \tag{2.5}$$

where T_j is the set of nodes split on feature j and N the total number of training samples.

In **Random Forests**, feature importance is averaged over all trees, reducing variance and yielding a more stable estimate. In boosting algorithms such as **XGBoost** and **CatBoost**, feature importance is computed in the same spirit but weighted by the contribution of each tree in the additive ensemble. Beyond impurity-based metrics, modern implementations often rely on permutation importance and SHAP values for more robust attributions, mitigating biases that can occur with correlated or high-cardinality features. Suppose we train a decision tree classifier on a dataset with two features (x_1, x_2) and a binary target. Assume that the root node has $N = 100$ samples with a Gini impurity of:

$$I_{\text{root}} = 1 - (0.5^2 + 0.5^2) = 0.5.$$

The root node is split using x_1, producing:

- Left child ($N_L = 40$): class distribution $(30, 10)$, impurity $I(L) = 1 - (0.75^2 + 0.25^2) = 0.375$.
- Right child ($N_R = 60$): class distribution $(20, 40)$, impurity $I(R) = 1 - (0.33^2 + 0.67^2) \approx 0.444$.

The impurity decrease attributed to x_1 at the root is:

$$\Delta I_{x_1} = N \cdot I_{\text{root}} - N_L \cdot I(L) - N_R \cdot I(R),$$

$$\Delta I_{x_1} = 100 \cdot 0.5 - 40 \cdot 0.375 - 60 \cdot 0.444 \approx 50 - 15 - 26.64 = 8.36.$$

Later in the tree, suppose x_2 is used for a split in one child node with $N = 40$, $I = 0.375$, producing two pure leaves ($I = 0$). Then:

$$\Delta I_{x_2} = 40 \cdot 0.375 - 20 \cdot 0 - 20 \cdot 0 = 15.$$

Thus, the total raw importances are:

$$FI_{x_1} = 8.36, \quad FI_{x_2} = 15.$$

After normalization:

$$FI_{x_1} = \frac{8.36}{8.36 + 15} \approx 0.36, \quad FI_{x_2} = \frac{15}{23.36} \approx 0.64.$$

So in this toy tree, x_2 is more important than x_1, because its splits reduce impurity more overall.

Overall, these ensemble methods provide a compromise between predictive accuracy and interpretability, making them well-suited for the case studies considered in this thesis:

- **Random Forest** is employed in the context of Biomechanical Phenotypes and Real-time Classification. Its ensemble of decision trees preserves a relatively transparent structure, where global and local feature importance can be directly extracted to identify biomechanical variables most associated with injury risk.

- **XGBoost** is used for personalization tasks in automotive and for genetic classification in Parkinson's disease (e.g., GBA1 vs non-mutation carriers). This model offers state-of-the-art performance on structured data and integrates well with SHAP-based interpretability, allowing clear attribution of predictions to clinical or behavioral features.

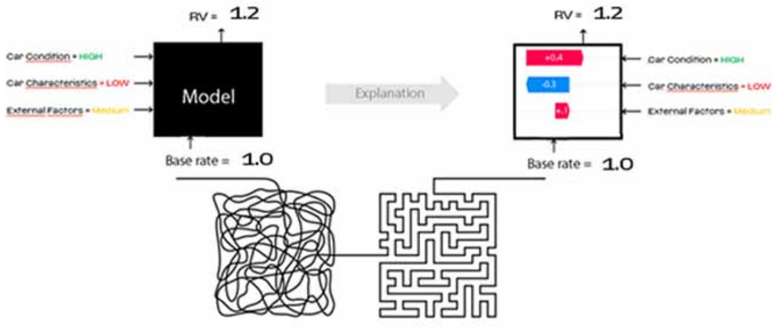

Figure 2.5: SHAP simplified application on the RV study

- **CatBoost** is adopted for residual value estimation and customer retention analysis in the luxury automotive domain. In addition to high accuracy, CatBoost natively handles categorical variables and provides efficient methods for calculating feature importance, making it particularly suited to business scenarios where transparent justification of recommendations is essential.

For what concern SHAP [49], it is a post-hoc interpretability method grounded in cooperative game theory. The central idea is to treat the model output for a specific instance as a “payout” that must be fairly distributed among the input features. Each feature receives a Shapley value, which represents its average marginal contribution to the prediction across all possible subsets of features. Formally, the Shapley value for feature i is defined as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \Big(f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S) \Big), \tag{2.6}$$

where F is the set of all features, S is a subset of F not containing i, f_S denotes the model restricted to features in S, and x_S are the corresponding feature values. The weighting term ensures that contributions are averaged across all possible coalitions, providing a principled and consistent attribution.

SHAP values have desirable theoretical properties:

- **Efficiency**: the sum of contributions equals the difference between the model output and its expected value.

- **Symmetry**: features with identical effects receive equal contributions.

- **Additivity**: attributions are consistent across models.

At the end, SHAP provides a principled way to compute **feature importance**, both at the global level (which variables drive the model overall) and at the local level (which features influenced a single prediction). This makes it directly applicable to the tree-based ensemble methods adopted in this thesis: Random Forest, XGBoost, and CatBoost, where SHAP values allow each prediction to be decomposed into feature contributions in a consistent and theoretically sound manner. In order to be clear we will present a small example. Consider a linear model $f(x) = w_1x_1 + w_2x_2 + b$ and an instance $x = (x_1, x_2)$. For clarity, set the baseline to $x^{(0)} = (0, 0)$ and $b = 0$ (i.e., $f(x^{(0)}) = 0$). Let $w_1 = 2$, $w_2 = 1$, $x_1 = 3$, $x_2 = 4$. Then $f(x) = 2 \cdot 3 + 1 \cdot 4 = 10$.

Define the value function $v(S)$ as the model output with features in S set to their instance values and all other features fixed at baseline, minus the baseline output:

$$v(S) = f\Big(x_S, x_{\bar{S}}^{(0)}\Big) - f\Big(x^{(0)}\Big).$$

For two features, Shapley values reduce to averaging the two possible insertion orders:

$$\phi_1 = \tfrac{1}{2}\Big[(v(\{1\}) - v(\varnothing)) + (v(\{1,2\}) - v(\{2\}))\Big], \quad \phi_2 = \tfrac{1}{2}\Big[(v(\{2\}) - v(\varnothing)) + (v(\{1,2\}) - v(\{1\}))\Big].$$

Compute the coalition values:

$$\begin{aligned} v(\varnothing) &= 0, \\ v(\{1\}) &= f(x_1, 0) = 2 \cdot 3 = 6, \\ v(\{2\}) &= f(0, x_2) = 1 \cdot 4 = 4, \\ v(\{1,2\}) &= f(x_1, x_2) = 10. \end{aligned}$$

Plug in:

$$\begin{aligned} \phi_1 &= \tfrac{1}{2}\Big[(6-0) + (10-4)\Big] = \tfrac{1}{2}(6+6) = 6, \\ \phi_2 &= \tfrac{1}{2}\Big[(4-0) + (10-6)\Big] = \tfrac{1}{2}(4+4) = 4. \end{aligned}$$

Additivity (efficiency) holds:

$$\phi_1 + \phi_2 = 6 + 4 = 10 = f(x) - f(x^{(0)}).$$

SHAP values were systematically employed to enhance interpretability across the different studies:

- **Genetic Parkinson's disease (GBA) classification:** SHAP was used to determine not only the magnitude but also the *direction* of each feature's contribution, clarifying whether a clinical variable increased or decreased the likelihood of being classified as GBA1-PD.
- **Residual value modeling:** SHAP values allowed the identification of thresholds beyond which certain features (e.g., mileage, vehicle age, macroeconomic indicators) started to exert a negative impact on price retention, offering actionable insights into when depreciation accelerates.
- **Cancellation risk analysis:** SHAP was crucial in highlighting the point at which individual drivers (such as contract duration, waitlist status, or customer profile) shifted from mitigating to amplifying the probability of contract withdrawal, supporting proactive retention strategies.

2.2.2 Dimensionality reduction

High-dimensional datasets often make interpretation difficult, since patterns, clusters, or structure are not easily observable. Dimensionality reduction techniques project data into a lower-dimensional space while preserving as much relevant structure as possible, allowing visual inspection and improved understanding of model behavior.

Preserving *local structure* means that data points which are close neighbors in the original high-dimensional space remain close in the reduced space. This is particularly important for detecting fine-grained subgroups or clusters. Preserving *global structure*, on the other hand, refers to maintaining long-range relationships and the overall geometry of the data manifold, ensuring that the relative placement of clusters is faithful to the original space [102].

Different dimensionality reduction methods emphasize different aspects. PCA is a linear and unsupervised technique that seeks orthogonal directions capturing maximum variance. Given a centered data matrix $X \in \mathbb{R}^{n \times p}$, the covariance matrix is

$$S = \frac{1}{n} X^\top X.$$

The first principal component is obtained by solving:

$$\max_w \; w^\top S w \quad \text{s.t.} \quad \|w\| = 1,$$

where w is an eigenvector of S. Projection onto the first k eigenvectors $W_k \in \mathbb{R}^{p\times k}$ yields the reduced representation

$$Z = XW_k.$$

Introduced by Pearson in 1901 and formalized by Hotelling in 1933, PCA remains one of the most widely used dimensionality reduction methods [103]. t-SNE is a nonlinear, unsupervised method focused on preserving *local structure.* Similarities in the original space are modeled with conditional probabilities:

$$p_{j|i} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k\neq i}\exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)},$$

and symmetrized as

$$P_{ij} = \frac{p_{i|j} + p_{j|i}}{2n}.$$

In the embedding space, similarities are defined using a Student-t kernel:

$$Q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k\neq l}(1 + \|y_k - y_l\|^2)^{-1}}.$$

The embedding minimizes the Kullback–Leibler divergence:

$$C = \sum_{i\neq j} P_{ij} \log \frac{P_{ij}}{Q_{ij}}.$$

t-SNE was proposed by van der Maaten and Hinton as an improvement over Stochastic Neighbor Embedding [104]. By contrast, UMAP is nonlinear and supports both unsupervised and supervised variants. It constructs a weighted k-nearest neighbor graph where the edge probability is

$$p_{ij} = \exp\left(-\frac{\max(0, d(x_i, x_j) - \rho_i)}{\sigma_i}\right),$$

with $d(\cdot,\cdot)$ a distance metric, ρ_i a local connectivity radius, and σ_i a normalization factor. In the low-dimensional embedding, a similar fuzzy set is defined as

$$q_{ij} = \left(1 + a \left\|y_i - y_j\right\|^{2b}\right)^{-1},$$

with parameters a, b controlling tightness of embedding. The optimization minimizes the cross-entropy between fuzzy sets:

$$C = \sum_{i \neq j} \left[p_{ij} \log \frac{p_{ij}}{q_{ij}} + (1 - p_{ij}) \log \frac{1 - p_{ij}}{1 - q_{ij}} \right].$$

In the supervised variant, label information modifies the graph structure so that points from the same class are embedded closer together, improving class separation. UMAP was introduced by McInnes et al. as a scalable manifold learning approach grounded in Riemannian geometry and algebraic topology [105].

In this thesis, t-SNE was employed in two contexts where interpretability of high-dimensional spaces was critical:

- **Customer retention in the luxury automotive domain**, where behavioral and transactional features are numerous and heterogeneous. t-SNE allowed projecting customers into two-dimensional maps that highlight separable patterns between retained and churn-prone individuals.

- **Biomechanical Phenotypes and Real-time Classification**, where multiple biomechanical variables were measured simultaneously. t-SNE enabled the visualization of athletes' movement profiles, making it possible to identify natural groupings that correlate with higher or lower injury risk.

2.3 Model Deployment

Concerning production deployment, the process is organized into sequential stages that ensure robustness and reproducibility. Deployment is not merely a technical afterthought, but a vital phase. Many organizations that build ML prototypes fail to deploy their models in production or fail to maintain successful operation of deployed models due to novel operationalizing challenges of ML-enabled software systems compared to traditional software [106]. Structured pipelines (e.g. defining pipelines, using shared version-control for datasets, features, and models) enhance team visibility, reduce duplication, and improve maintainability and reliability [106].

Data are first collected from source systems and ingested into a centralized repository, typically a cloud data lake such as Amazon S3. This guarantees durability, versioning, and scalability for large volumes of raw information. Durability means the data are stored redundantly in object storage with high-availability mechanisms, protecting against data loss. Versioning and metadata tracking ensure that changes to datasets (new extracts, schema evolution, updates) are traceable and reversible, which is essential for reproducibility and governance.

Scalability refers both to volume (handling terabytes or more) and variety (structured, semi-structured, unstructured data). A properly designed data lake supports flexible ingestion in batch and streaming modes, without rigid schema requirements, thereby accommodating evolving data sources and formats. These properties set the foundation for downstream processes such as cleaning, feature engineering, and model training, by assuring that the raw data is reliable, versioned, and usable [107].

The raw data undergoes cleaning and validation. This includes handling missing values with appropriate imputers, detecting and correcting anomalies and outliers, and ensuring consistency of formats and units. These steps are critical to guarantee both correctness and quality prior to model development. As noted by Kwak and Kim, properly treating missing values and outliers in the pretreatment stage helps reduce biased results and prevents over or under-estimation in downstream modeling [108].

Once the dataset is prepared, domain knowledge is leveraged to design features that enhance model performance. Feature engineering often involves transformations, interaction terms, encoding, or ontology-based mappings when data are heterogeneous. For example, in a clinical setting with patient data, ontology-based feature engineering significantly improved performance for logistic regression, random forest, and gradient boosted trees by about 35.7%, 54.5%, and 33.3% respectively [109]. Models are then trained following the progressive strategy outlined earlier: beginning with interpretable, low-complexity methods before advancing to more sophisticated architectures when required.

To facilitate adoption, it is important to provide a user interface that allows domain experts to explore results and interact with the trained model. For rapid development of such interfaces, Streamlit provides a lightweight, Python-native framework that enables interactive apps with minimal boilerplate and no need for frontend development [110]. Its practicality is evidenced across domains, from geospatial analytics [111] to clinical decision support [112].

For production readiness, Docker has become the de facto standard for packaging and serving machine learning models. First, containerization ensures consistency across environments, allowing the same image to be used from development to production, thus reducing dependency conflicts and reproducibility issues [113]. Second, Docker offers efficiency and scalability, since containers are lightweight compared to virtual machines and allow models to be deployed and scaled rapidly across cloud or on-premise infrastructure [114]. Third, Docker images guarantee portability, enabling seamless migration across heterogeneous hardware and operating systems without modification of the code base [113]. Finally, Docker integrates natively with orchestration systems such as Kubernetes, making it central to modern orchestration and microservice architectures for ML pipelines [115].

In terms of deployment, the industrial projects (residual value estimation, configuration recommender, and customer retention) were developed and deployed in the cloud, using Amazon S3 as a centralized data lake and AWS SageMaker for data preparation, training, and serving. Cloud infrastructure was chosen here to guarantee scalability, elasticity of compute, and seamless integration with enterprise pipelines. By contrast, the Biomechanical Phenotypes and Real-time Classification project followed a different deployment paradigm, reflecting the specific needs of the healthcare domain where privacy and regulatory compliance are paramount. The graphical interface was implemented in Streamlit 1.30 (Python 3.12) and packaged together with the serialized Random Forest model into a lightweight Docker image. This design eliminated external service calls, as the container bundled both the front end (‘app.py‘) and the model artifact (‘model.pkl‘). Deployment occurred on an on-premise Ubuntu 22.04 server equipped with two Intel Xeon E5-2680 v4 CPUs (2.40 GHz) and 4 GB RAM, secured behind the clinic’s virtual private network (VPN). All computation and storage remained

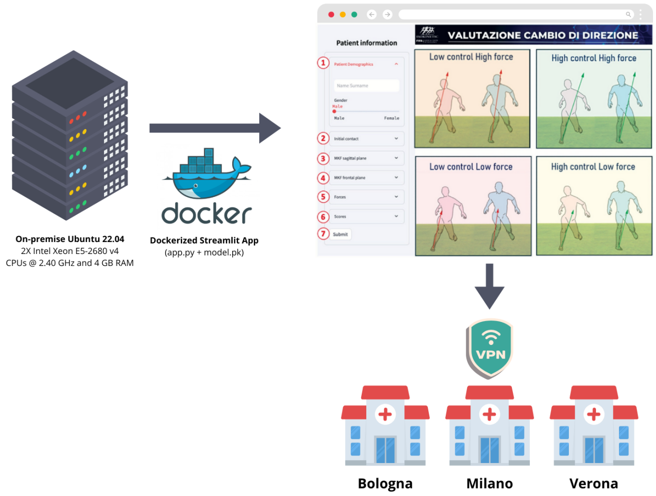

Figure 2.6: Example from Biomechanical Phenotypes and Real-time Classification project

within the clinical infrastructure, ensuring GDPR compliance without reliance on external cloud services (Figure 2.6).

Part II

Contributions

Chapter 3

Contributions in Luxury Automotive

Building on the broader notion of luxury as a field of non-essential yet highly desired goods and experiences distinguished by exceptional quality, craftsmanship, and a premium to mainstream alternatives, the *luxury automotive market* can be defined as the portion of the passenger-vehicle industry whose value proposition is anchored not only in superior engineering and materials but in the orchestration of scarcity, identity, heritage, and intensely personalized client experiences.[1, 3]. In this sense, "luxury car" denotes more than an equipment bundle or a price band: it describes an *experience platform* that fuses product excellence with symbolic and relational capital, typically delivered through controlled supply, bespoke configuration, brand rituals, and aftersales theater.

Among luxury categories, automobiles are structurally pivotal. Recent Bain–Altagamma monitoring shows that luxury cars are the largest single segment of the luxury economy by value; in 2024 the category amounted to roughly €579 bn at current exchange rates and declined modestly year-on-year while remaining above 2019 levels[3]. The same research also documents a polarization dynamic across the luxury landscape: experience-led spending continues to outpace pure products, and within experience-based goods (a bucket that includes luxury cars), contraction at the "entry-to-luxury" edge contrasts with durable demand at the absolute top from HNWIs[3]. Complementary evidence from McKinsey underscores the outsized strategic relevance of top-end vehicles: despite their small volume share, these cars punch far above their weight in brand equity, investor narratives, and halo effects that spill over into adjacent offerings and experiences [2].

In line with the luxury-brand literature, four families of value are particularly salient in automotive: (i) *functional value* (performance, refinement, longevity); (ii) *experiential value* (aesthetic pleasure, sensorial staging, service theater); (iii) *symbolic value* (status signaling and identity); and (iv) *financial value* (residual-value strength that supports access via leasing and signals enduring desirability)[116, 117, 118]. These dimensions interact in ways that are distinctive to cars. Engineering excellence and craftsmanship provide a credible substrate for symbolic value (e.g., motor-sport heritage, design codes, coachbuilding), while service rituals, delivery ceremonies, and communities (e.g., track programs) institutionalize experiential value that endures beyond the transaction.

Scarcity is not a by-product of operational constraints but an organizing principle of luxury [119]. In cars, it is implemented through disciplined capacity planning, sequenced allocations, limited series, and long waiting lists that curate rarity and protect identity. Empirical work in the automotive industry shows that lower introductory inventory is associated with higher consumer preference, consistent with supplier-induced scarcity mechanisms [120]. Transaction-level research likewise finds that consumers pay more when local inventory is low, evidencing *scarcity rents* in car retailing [121]. Contemporary examples illustrate this logic in practice: Porsche's 911 S/T was capped at 1,963 units to commemorate the model line's 60th anniversary, explicitly encoding rarity into the product [122]. Leading marques complement quantity discipline with qualitative selectivity (curation of allocations, invite-only models), reinforcing both social signaling and secondary-market performance.

Personalization has become a growth and margin engine across top-end automotive. Ferrari reports that personalizations account for roughly one-fifth of revenues from cars and spare parts, and management has emphasized the mix and customization tailwinds behind recent results [123, 124]. Programs such as Ferrari's Atelier and Tailor Made, Lamborghini's Ad Personam, and Bentley Mulliner revive the coachbuilding tradition within a modern industrial framework, enabling clients to co-create unique specifications under strict brand governance

[125, 126, 127]. Beyond factory programs, brands stage immersive experiences, private design studios, factory visits, and curated delivery rites, that convert configuration into narrative and social capital. These practices are economically rational: personalization sustains margins, elevates attachment (thus reducing defection risk during wait times), and further differentiates constrained supply.

As luxury OEMs re-architect distribution around experience and data ownership, we observe the emergence of "brand houses" and track ecosystems that extend the ownership ritual. Porsche's "Destination Porsche" format reframes dealerships as community hubs, while Porsche Track Experience and Ferrari's Corse Clienti/XX Programme convert clients into lifelong participants in the brand's culture and skills [128, 129, 130, 131]. Digital clienteling complements these physical rituals: apps such as MyFerrari tighten ties throughout the order-to-delivery journey, provide transparency, and create targeted touchpoints for content, invitations, and post-sale care, a relevant capability when scarcity extends waitlists [132]. Industry loyalty data stress the stakes: luxury-brand loyalty, while strong, tends to trail mainstream segments, implying that high expectations and abundant choice raise switching risk; proactive, personalized engagement is thus commercially material[10, 11].

Residual value, the forecasted market value of a vehicle at the end of its finance/lease horizon, is unusually central in luxury. High RVs reduce total cost of use for clients (e.g., lower lease payments), support disciplined pricing, and transmit a credible signal of enduring desirability [13, 133]. Two reinforcing mechanisms operate here. First, managed scarcity and strong brand equity support used-market prices; independent measures regularly highlight top-end sports cars (e.g., Porsche 911) as best-in-class for value retention [134]. Second, channel hygiene matters: minimizing price haggling and discount dispersion protects RVs, which is especially critical in premium segments [135]. Brands institutionalize RV protection through certified pre-owned (CPO) programs, refurbishment standards, and warranty extensions that stabilize secondary-market outcomes while extending the service theater into the second and third owners [136, 137]. In finance, more sophisticated RV forecasting and risk-transfer structures are spreading as volatility in used-vehicle prices persists [13].

Luxury cars are intensely semiotic objects. Classic and contemporary research shows that consumers use luxury goods to signal status and group membership, sometimes with conspicuous *brand prominence*, sometimes with deliberately quiet cues [118]. In automotive, those cues are embedded in design codes (proportions, surfacing), heritage markers (motorsport liveries, anniversary editions), and community participation (rallies, track time). The signaling function helps explain why high-end limited series often appreciate in the secondary market and why collectible-car culture, now a multi-hundred-billion-euro asset class, acts as a powerful amplifier of brand myth and desirability [138].

Unlike fashion, where iteration cycles are rapid, the car is a durable product whose value unfolds over years. Luxury OEMs therefore curate a *service stack*, concierge maintenance, OTA-enabled feature evolution consistent with brand codes, exclusive events, that protects the aura of *timelessness* while still delivering novelty. KPMG frames the sector's current tension as a calibration between exclusivity/scarcity and new forms of accessibility enabled by technology; top-end automotive addresses this not by commoditizing access but by ritualizing it (appointments, track programs, curated clubs) and by tightly governing the brand's data layer [6, 119]. At the same time, regional differences matter: in China, luxury buyers place unusually high value on "smartification" (ADAS, connectivity, digital ecosystems), raising the bar for digital luxury without diluting scarcity [135].

The luxury automotive market is best understood as an ecosystem in which product excellence is necessary but insufficient. What distinguishes luxury is the deliberate management of *rarity*, *meaning*, and *memory*, through capacity discipline and limited series (scarcity), through codes, heritage, and client communities (symbolic capital), and through personalization, rituals, and service choreography (experiential capital). Economically, these practices generate high lifetime margins and protect residual values, creating a self-reinforcing loop: strong RVs lower effective cost of use, widen the addressable pool within the top-end, and validate the brand's long-term value narrative. Strategically, they make luxury cars the keystone of the luxury economy's experience-based turn while preserving the very exclusivity that defines the category [2, 3, 133].

3.1 Personalization in Luxury Car Configurators

Recommender systems are software tools and techniques that provide personalized suggestions by estimating the relevance of items a user may prefer within a large catalog. By tailoring options to individual preferences, they reduce information overload and improve decision making and user satisfaction [139].

Collaborative Filtering (CF) is one of the earliest and most widely adopted approaches in recommender systems. It generates predictions by exploiting preference patterns across many users, under the assumption that users who agreed in the past will also agree in the future. CF can be implemented through user-based or item-based similarity models, as well as latent factor methods such as matrix factorization [140].

The first practical implementation of Collaborative Filtering appeared in the *Tapestry* system, developed at Xerox PARC, which enabled users to filter email and documents by explicitly relying on the ratings and annotations of others [141]. This work demonstrated that leveraging collective human judgments could effectively reduce information overload. Subsequent systems such as GroupLens extended this paradigm by automating similarity computations across large communities, making CF scalable and widely applicable [142]. The main benefits of CF lie in its domain independence, since it does not require explicit knowledge about item content, and in its ability to uncover complex and unexpected preference patterns through aggregated user behavior. These advantages made CF the backbone of some of the most influential industrial recommender systems. Amazon's item-to-item CF scaled recommendations to millions of users and products by focusing on item similarity rather than user similarity [143], while the Netflix Prize further highlighted the effectiveness of latent factor CF methods, such as matrix factorization, in improving predictive accuracy on a large-scale dataset of movie ratings [144]. Over time, evaluation has moved beyond raw accuracy toward ranking-oriented and user-centric metrics. In addition to precision@k, recall@k, MAP, NDCG, MRR, and AUC, researchers have stressed the importance of measures that capture real user experience, including diversity, novelty, and serendipity. This shift reflects both foundational work on evaluation frameworks [145], critiques of accuracy as an insufficient benchmark [146], and empirical studies on top-N recommendation quality [147] as well as novelty, and diversity-aware metrics [148].

Despite its success, Collaborative Filtering faces several challenges that became increasingly evident as datasets and user communities expanded. A central issue is *data sparsity*: in large catalogs, most users interact with only a small fraction of items, making similarity computations unreliable. Closely related is the *cold-start problem*, where new users or new items lack sufficient ratings to be modeled accurately. Furthermore, scalability became a practical concern as the numbers of users and items grew into the millions, straining memory and computation in neighborhood-based approaches. These limitations motivated the development of latent factor models and, later, deep learning architectures that aim to capture user–item interactions more effectively in high-dimensional spaces [144, 149].

Hybrid recommender systems emerged to overcome the limitations of pure collaborative or pure content-based approaches. By combining multiple sources of evidence, such as user ratings, demographic attributes, and item metadata, hybrid methods improve accuracy, alleviate cold-start problems, and increase robustness. Such systems can be designed in different ways, including weighted or switching combinations, feature augmentation, or full model-level integration [150].

The growing availability of large-scale datasets and advances in machine learning further extended this trajectory. Recommenders began to incorporate supervised learning techniques such as decision trees, logistic regression, gradient boosting, and factorization machines, which enabled the inclusion of heterogeneous side information and the optimization of recommendation as a predictive task [151].

More recently, deep learning has been widely applied to recommender systems. Neural architectures enable the modeling of complex, non-linear interactions and sequential behaviors. Pioneering works such as Neural Collaborative Filtering [152], GRU4Rec for session-based recommendation [153], and YouTube's industrial-scale deep recommender architecture [154] demonstrate the flexibility of deep learning in capturing temporal dynamics, multimodal signals, and high-dimensional user–item relations.

The latest wave of research explores the integration of Large Language Models (LLMs) into recommender systems. Unlike earlier approaches that rely on structured user–item interactions, LLMs exploit natural language as a universal interface for both users and items. This enables recommendations to be contextualized by textual queries, reviews, or conversational dialogues, moving beyond traditional numeric ratings. Early studies show that pretrained transformers can serve as zero-shot recommenders by mapping items and preferences into shared semantic spaces, while fine-tuned variants show promise in tasks such as sequential recommendation and cross-domain transfer [155, 156]. At the industrial level, companies are beginning to experiment with LLM-based conversational recommenders, highlighting the potential of these models to unify recommendation, explanation, and personalization within a single framework.

Although Large Language Models offer new opportunities for personalization, they also raise challenges that differ from earlier approaches. One critical issue is the tendency of LLMs to *hallucinate*, that is, to generate fluent but factually incorrect or fabricated recommendations. This undermines trust and reliability in sensitive domains such as healthcare or finance. Furthermore, the black-box nature of LLMs exacerbates concerns about interpretability and accountability. To mitigate these risks, recent works emphasize the need for *guardrails*, including retrieval-augmented generation, factual grounding, and post hoc validation layers that constrain model outputs to verifiable knowledge sources [157, 158].

Unlike mainstream e-commerce or media domains, the luxury automotive market presents a distinctive set of challenges for recommender systems. First, the customer base is extremely limited, which means interaction data is sparse and user similarity signals are difficult to compute reliably. Second, the product space is vast and highly configurable: a single model can generate thousands of possible variants across trims, colors, interiors, and bespoke features. This combination of few customers and many items leads to severe data sparsity, exacerbating the cold-start problem and making collaborative filtering approaches less effective [159, 160].

Moreover, the infrequency of purchases in the luxury segment further limits the availability of behavioral data. Unlike digital platforms, where users generate feedback daily, luxury automotive decisions are rare, long-term, and emotionally charged. Studies in automotive product configuration highlight the need for hybrid and knowledge-based recommenders that incorporate explicit domain knowledge and user attributes to compensate for missing historical signals [161, 162]. This distinguishes the luxury automotive context as one of the most demanding application areas for recommender technologies, where traditional large-scale collaborative methods must be adapted or rethought.

Several industry deployments illustrate their increasing strategic importance. These cases show how recommender technologies are being adapted to the distinctive conditions of the sector, particularly in luxury and premium contexts.

One example is provided by a European premium car manufacturer that collaborated with Statworx to develop an in-car recommendation engine for digital services. The system processed more than 37 million events across roughly 1,000 behavioral and contextual features, implemented on Spark and Azure infrastructures. Its deployment led to a reported increase in conversion rates of up to 70%, underlining the tangible business value of personalization in after-sales and in-car services [163]. Porsche has integrated recommender technology into its car configurator, employing a large ensemble of approximately 270 neural networks trained on several million data points. The system achieved a reported 90% recommendation accuracy, thereby guiding users more effectively through complex vehicle configuration processes and reducing decision fatigue in high-variance product spaces [164].

In the broader market, Toyota has introduced data-driven personalization features across its digital platforms, leveraging browsing patterns and preference signals to suggest tailored vehicle options. Although details remain high level, this initiative illustrates how recommender systems are increasingly adopted even in high-volume contexts, indicating their growing relevance beyond purely luxury niches [165].

Finally, Ferrari has pursued a more experimental path by integrating generative AI into its digital ecosystem through AWS services. This initiative aims to enhance both design workflows and client-facing personalization in online and mobile channels, aligning technological innovation with the brand's emphasis on exclusivity and experience [166]. Although industrial applications highlight the promise of recommender systems in the automotive domain, a persistent obstacle remains: each deployment is fundamentally case specific and undermined by the scarcity of reliable data. Luxury auto markets are defined by a small customer base, rare purchase events, and immensely configurable product lines, creating severe interaction sparsity. This makes it infeasible to generalize models even across different brands or across product lines within the same brand. For example, Cai et al. demonstrate how sparse preference data in automotive design contexts necessitates hybrid and domain-aware methods such as clustering and nonlinear Bayesian Personalized Ranking to compensate for the lack of user history and group-level patterns [161]. Scholarly work also confirms that sparsity and cold-start issues remain among the most intractable challenges in recommender systems, especially when the user–item matrix contains overwhelmingly missing entries [167, 168]. Developing a recommender system for a high-end automotive company presents unique challenges. The purchase of a high-end vehicle is a highly personalised process that may involve hundreds of mutually compatible or exclusive options, so nearly every car is configured uniquely [20]. Smooth guidance through this combinatorial space is crucial to customer satisfaction [23].

Recommender systems have proved effective in e-commerce, media and social platforms [21, 149], but their application to big-ticket, configurable products remains comparatively under-explored. Early research on *recommendation-based product configurators* shows that data-driven suggestions can streamline complex choices and raise conversion rates [23]. Luxury-sector case studies (e.g. premium automotive and bespoke fashion) likewise report higher engagement when personalised option advice is available [24].

Two factors distinguish the automotive setting. First, every vehicle model possesses a largely disjoint catalogue of options; each model thus forms a separate domain with its own item set and customer cohort. This mirrors *cross-domain recommendation*, where knowledge must be transferred via shared descriptors or user links [24]. Second, historical interaction data are sparse and heavily skewed; many features are selected rarely, and new models continually introduce unseen items, triggering the well-known *cold-start* problem [26, 27]. Classical collaborative filtering alone therefore under-performs.

Hybrid approaches that fuse collaborative and content signals, originally surveyed by Burke [28] and refined in recent systematic reviews [169], are the de-facto remedy. By incorporating rich option metadata (material, aesthetic theme, package compatibility) into a collaborative backbone, hybrids can generalise across models, alleviate cold-start and respect configuration constraints. Recent advances in representation learning further strengthen this paradigm: neural collaborative filtering [152], collaborative deep learning [31] and implicit matrix-factorisation variants optimised for sparse data [170] yield state-of-the-art accuracy while preserving scalability for industrial catalogues.

In this contribution, we propose a hybrid method that combines collaborative filtering and content-based techniques, and we demonstrate its effectiveness through offline experiments (**P-RQ1**). The system has also been deployed in production and is currently used within the a high-end automotive company Atelier, a specialized facility where clients, assisted by experts, co-design vehicles precisely aligned with their personal tastes. Online pilot tests with real customers further confirm that this is a promising direction for enhancing the personalization and improving the user experience of high-end vehicle purchases (**P-RQ2**).

3.1.1 Methods

The goal of our methodology is to develop a recommendation engine capable of suggesting personalized and non-trivial optional features for high-end vehicle configurations. The approach must be scalable, robust, and suitable for deployment in a real-world industrial setting. To this end, we adopt a hybrid recommendation framework combining structured historical data with collaborative and content-based signals. In the following, we describe the dataset used, the model design, and the evaluation protocols.

The dataset utilized in this study comprises approximately 10,000 unique user profiles extracted from a Customer Relationship Management (CRM) system, alongside metadata for around 900 distinct car options, including features such as material, part of the vehicle (eg exterior, interior, wheels), and color. Historical ordering patterns were considered by analyzing approximately 14,000 past purchase records associated with recent vehicle models.

In particular the dataset used for the recommender system integrates **customer-level CRM attributes** with **option-level product features**. This hybrid structure enables the development of models that leverage both user profiles and item characteristics, thereby supporting hybrid recommendation strategies.

The CRM component captures socio-demographic and behavioral information about customers, while the option features describe the characteristics of

Name	Description	Type	Source
Gender	Customer gender information (male, female, other)	Categorical	CRM
Generation	Customer generation/age cohort (e.g., Gen X, Gen Y, Gen Z)	Categorical	CRM
Years as a customer	Tenure of the customer in years since first purchase	Numerical (continuous)	CRM
HUB	Geographical or organizational hub associated with the customer	Categorical	CRM
HQ Rank	Internal ranking of the customer within HQ scoring system	Ordinal	CRM
Customer Status	Current state of the relationship (active, prospect, former)	Categorical	CRM
Price	List or transaction price of the option/product	Numerical (continuous)	ERP
Model	Product model identifier (e.g., vehicle model)	Categorical	ERP
Description	Textual description of the option/product	Textual	ERP
Categories	Product segmentation (e.g., body type, trim level)	Categorical (multi-class)	ERP
Take Rate	Percentage of customers selecting a given option	Numerical (proportion)	ERP
Margin	Profit margin associated with the option/product	Numerical (percentage)	ERP

Table 3.1: Dataset structure: features included in the recommender system, with description, type, and source.

configurable products.

To facilitate deployment in an industrial setting, we rely on XGBoost [100], a robust and scalable gradient-boosted decision tree algorithm widely used for structured data. In our setup, XGBoost is trained to directly optimize the ranking of optional features for each customer, assigning higher scores to features the customer has selected in the past. While the input data includes all available user–feature interaction data, the model is trained to correctly rank *non highly-popular* optional features. This choice reflects the system's intended use case to provide personalized and non-trivial recommendations that go beyond frequently selected options. By concentrating the evaluation on less common features, we better assess the model's effectiveness in supporting meaningful personalization during vehicle configuration.

To build the training set, we combine collaborative with content-based information. In particular, we include as input features the scores generated by *Implicit Alternating Least Squares* (iALS) [171], a matrix factorization method that estimates user-item affinities from past interactions that is very effective [172], as well as of *Top Popular*, which assigns a popularity score to each optional based on global frequency.

The model also incorporates descriptive attributes of the optional features, such as category (e.g., interior trim, wheels) and material (e.g., carbon fiber, leather).

For instance, the system can learn that a customer who previously selected carbon fiber parts for one model may be interested in comparable performance-oriented options in other models, even if the exact feature is not available. This combination of collaborative and content-based information enables the model to generalize better across different vehicle models and customers, including cold-start scenarios.

Model evaluation was performed in two distinct phases:

1. **Offline Evaluation:** Historical data were used to assess the predictive performance of the model and verify alignment with past purchase patterns.

2. **Online Evaluation:** A controlled pilot test was conducted within the high-end automotive company Atelier, where trained dealers interacted with the tool under supervised conditions.

Offline Evaluation

To simulate realistic deployment conditions, historical orders were partitioned into three temporally ordered subsets. The earliest transactions were assigned to the **training set**, subsequent ones to the **validation set**, and the most recent to the **test set**. This chronological partitioning ensures that the system is evaluated on unseen future data, replicating the operational scenario where the model must anticipate customer choices based on past behavior.

Optional features are intrinsically tied to specific car models. As new models are released and old ones are discontinued, the concentration and availability of options vary significantly across time. Consequently, the training, validation, and test splits are not homogeneous in terms of feature distribution. The test set, in particular, contains a higher proportion of features associated with new models, allowing us to explicitly evaluate the system's capacity for **knowledge transfer across vehicle generations**, a critical capability in the luxury automotive domain where product innovation cycles are frequent.

Hyperparameter optimization was performed using **Optuna** [173], a state-of-the-art framework that leverages Bayesian optimization and adaptive sampling strategies to efficiently explore high-dimensional parameter spaces. Optuna's pruning mechanisms further reduce computational cost by early-stopping under-performing trials, thereby accelerating convergence toward optimal hyperparameter settings.

For evaluation, we adopted a suite of standard **top-n ranking metrics**:

- **Precision@n**: measures the fraction of relevant features among the top-n recommended options.
- **NDCG@n (Normalized Discounted Cumulative Gain)**: accounts for ranking position, rewarding systems that place relevant features higher in the recommendation list.
- **Hit Rate@n**: proportion of customers for whom at least one relevant feature is present among the recommendations.

To go beyond accuracy, we assess additional qualitative aspects:

- **Mean Inter-List Diversity (MIL Diversity)** [174]: quantifies personalization by measuring the average proportion of non-overlapping features between any two recommendation lists. Higher values indicate that the system generates more user-specific suggestions rather than repeating generic options.
- **Average Popularity**: evaluates the mean popularity of recommended features, normalized such that the most popular one has value 1. This metric assesses whether the model tends to favor universally popular features.
- **Popularity Ratio**: compares the frequency of recommendation of each feature to its frequency in the training data. Values below 1 suggest that the system avoids amplifying popularity bias, promoting a fairer distribution of recommendations.

This comprehensive evaluation framework ensures that the system is tested not only for predictive accuracy, but also for personalization and fairness, two dimensions that are essential for user satisfaction and business relevance.

Online Evaluation

The **operational workflow** of the recommender system is embedded within the dealership experience. When a customer books an appointment at an Atelier to configure a specific car model, the recommender engine is activated using three input sources: (i) the customer's CRM profile, (ii) the attributes of the selected car model, and (iii) contextual information about the dealership's local hub. Based on this multi-dimensional input, the engine generates a ranked list of personalized optional features tailored to both customer preferences and local dealership characteristics.

The generated recommendations are transmitted in advance to dealership staff, who use them to guide the upcoming configuration appointment. This design enables a **proactive personalization process**: dealers are equipped with contextualized suggestions before interacting with the customer, enhancing both the relevance and the quality of the advisory interaction.

Evaluating such an online system presents several methodological challenges. Unlike offline experiments, dealership interactions are inherently **context-dependent**, influenced by sales strategies, staff expertise, and local market conditions. Furthermore, technological and privacy constraints limit the possibility of fully logging and replicating these real-world interactions offline. As a result, offline metrics provide only a partial view of the system's utility.

To address these challenges, the online evaluation followed an exploratory field design. The recommender was introduced in a limited number of dealerships, where sales staff could access personalized configuration suggestions and report their impressions. Key outcomes of interest included configuration diversity, alignment with customer preferences, and perceived quality of the interaction as reported by sales staff. This setup allowed us to observe the recommender's contribution in realistic dealership settings without a formal randomized control structure.

3.1.2 Results and Discussion

In this section we summarize and discuss the results of both the offline and online evaluation.

Offline Evaluation

Table 3.2 shows the typical accuracy–novelty trade-off as the cutoff increases. At k=5, **Precision**=0.646 implies an expected $k \cdot \text{Precision} \approx 3.23$ relevant options among the top five, with **NDCG**=0.663 confirming that relevant items are concentrated near the top ranks. As k grows (10, 20, 30), precision decreases ($0.561 \rightarrow 0.503 \rightarrow 0.441$) as the list broadens, while **NDCG** remains high and even slightly rebounds at k=30 (0.567), indicating that although additional items include more exploratory content, the ranking still surfaces relevant items early.

Hit Rate is essentially saturated (0.998 at k=5 and 1.000 thereafter), meaning virtually every user receives at least one correct suggestion, which is desirable for showroom workflows where advisors need quick, reliable talking points.

On beyond-accuracy dimensions, **Mean Inter-List Diversity** is consistently around 0.80, i.e., about 80% of items in two users' lists differ on average. This supports the claim that the model produces *personalized* rather than one-size-fits-all outputs. At the same time, **Average Popularity** declines as k increases ($0.467 \rightarrow 0.322$), and the **Popularity Ratio** drops from 1.254 at k=5 to 0.865 at k=30. Together, these trends show that the system progressively explores the long tail as more recommendations are presented, while maintaining ranking quality (as reflected by NDCG). Operationally, this behavior is useful in a luxury context: short lists remain highly precise and easy to action, whereas longer lists reveal non-obvious, less popular features that can spark bespoke configurations.

Because optional features are model-dependent, the temporal split stresses *cross-model transfer*. High values on all three accuracy metrics with a temporally held-out test set (including newer model vintages) indicate the model extrapolates patterns learned on earlier models to later ones. This is consistent with a hybrid design in which user signals (history and profile) combine with content features (option descriptors, model attributes) to bridge gaps caused by shifting assortments.

Feature-importance analyses (e.g., permutation importance or SHAP for tree-based components) consistently highlight: (i) *user signals* (purchase/option history, tenure) as primary drivers, (ii) *item descriptors and popularity indicators* as complementary signals that stabilise ranking in sparse regions, and (iii) *dealership/context features* as secondary but useful tiebreakers (e.g., aligning to local taste). This aligns with the observed diversity and popularity ratios: personal history steers personalization, while content and mild popularity priors maintain precision at small k.

Against the commercial baseline, our model shows a superior accuracy personalization trade-off at k=5: Precision = 0.646 vs. 0.596 (+8.4% relative), NDCG = 0.663 vs. 0.643 (+3.1%), and MIL = 0.811 vs. 0.765 (+0.046 absolute). The baseline attains lower *Average Popularity* at higher k (e.g., 0.200 at k=30 vs. 0.322), suggesting a stronger novelty bias; however, this comes with lower precision and lower inter-user differentiation at small k. For Atelier operations, where short, high-confidence lists enable efficient appointments, our model offers a better overall balance.

For dealer-facing workflows we recommend k=5–10: high *Precision* and *NDCG* with strong *MIL*, and a *Popularity Ratio* close to 1 (minimal bias amplification). For inspiration phases or remote pre-reads, k=20–30 intentionally broadens the option set (lower average popularity) without collapsing ranking quality.

To quantify uncertainty around the reported metrics, we suggest non-parametric bootstrapping at the user level (e.g., 1,000 resamples for 95% CIs of Precision@k, NDCG@k, MIL, and popularity metrics). Further robustness checks include: (i) sliding temporal windows to test sensitivity to seasonality; (ii) ablations removing popularity signals to verify the source of gains; and (iii) calibration of popularity exposure via target ratios if business policy requires bounding bias (e.g., enforce Popularity Ratio $\in [0.8, 1.2]$).

Cutoff	**Precision**	**NDCG**	**Hit Rate**	**MIL Diversity**	**Popularity Avg.**	**Popularity Ratio**
5	0.646	0.663	0.998	0.811	0.467	1.254
10	0.561	0.597	1.000	0.800	0.408	1.096
20	0.503	0.560	1.000	0.799	0.367	0.984
30	0.441	0.567	1.000	0.793	0.322	0.865

Table 3.2: Evaluation metrics at different cutoff values.

Online Evaluation

The four knowledge scenarios separate how much prior information the system has on the *customer* and on the *model*. *Known Model & Customer* (Precision@5 = 0.23) performs better than *Cold-start Customer, Known Model* (0.19), underscoring the value of individual preferences. Interestingly, *Cold-start Model, Known Customer* attains the highest Precision@5 (0.27), suggesting that customer tastes generalise more reliably across new models than model-specific patterns generalise across new customers a desirable property for fast-moving product lines. Even the *Fully Cold-start* case reaches 0.40 on a small subset, indicating that the hybrid architecture can fall back on content features when no history exists.

The pilot covers 53 sessions; consequently, estimates have wide uncertainty and class imbalance across scenarios (e.g., fewer fully cold-start cases). We therefore interpret differences directionally, not as definitive. In future roll-outs, a cluster-randomised A/B design at the *dealer-day* level with stratification by model family would allow unbiased intent-to-treat estimates. We recommend reporting (i) Precision@k with 95% bootstrap CIs, (ii) session-level effect sizes (e.g., Cliff's delta) across scenarios, and (iii) business KPIs such as attachment rate of recommended options, configuration time, and margin lift.

Dealer interviews corroborate the offline picture: the lists are perceived as specific and actionable, improving conversation flow and helping surface non-trivial options. The stronger performance when the customer is known aligns with staff feedback that prior tastes are crucial for bespoke configurations.

For day-to-day use, a two-stage policy is advisable: (1) present a short, high-precision list (k=5–10) as the "primary" slate; (2) expose an auxiliary "explore" slate (e.g., another 10–20 items) ordered by novelty or margin for advisors who want breadth. If business policy emphasises novelty or margin, re-rank the tail using a constrained objective that keeps Popularity Ratio near 1 while optimising the secondary target.

Internal validity may be affected by staff selection effects (advisors who are more enthusiastic might rely more on the tool) and by non-random exposure of scenarios. External validity is bounded by the specific model cycles and customer mix observed in the pilot. We mitigate these by temporal validation offline, scenario stratification online, and plan to adopt cluster randomisation in subsequent phases.

Case	Precision	std	N. Cases (%)
Overall online test	0.21	0.16	100
Cold start Model & know customer	0.27	0.16	14
Know Model & customer	0.23	0.23	28
Cold start Customer & know model	0.19	0.13	52
Cold start Customer & model	0.40	0.20	6

Table 3.3: Online test evaluation (53 configurations), precision with a cutoff of five

3.1.3 Contributions

In this work, we presented a personalized recommendation system designed to support the configuration of high-end vehicles by suggesting non-trivial optional features customize to individual customer preferences. The system combines collaborative filtering with content-based signals within a scalable machine learning framework and has been successfully deployed in a real-world Atelier setting.

Our evaluation, both offline and in production, demonstrates that the model achieves high accuracy while also maintaining strong personalization and mitigating popularity bias. These results confirm the viability of applying recommender systems to this novel industrial use case and show the potential of data-driven tools to enhance the customer–dealer interaction during vehicle personalization.

3.2 Retention during waiting list

Customer retention is generally defined as a firm's ability to maintain its existing customers over time by encouraging repeat purchasing and fostering loyalty [175, 176].

The earliest approaches to customer retention were grounded in service marketing and relationship marketing traditions of the late 1980s and early 1990s. At that time, the dominant view of marketing was still transactional, emphasizing short-term sales over long-term relationships. The retention perspective emerged by stressing that loyal customers were more valuable than constant cycles of acquisition and replacement [177].

These first approaches framed retention as the outcome of customer satisfaction and service quality. Firms were encouraged to reduce defections by focusing on consistent delivery, trust-building, and the costs customers incur when switching providers [178]. Relationship marketing further established retention as a core construct, highlighting that trust and commitment form the foundation of enduring customer–firm exchanges. [179]. Bolton [180] analyzed customer retention in the telecommunications sector and found that satisfaction ratings, collected before any decision to cancel or remain loyal, were positively associated with the subsequent duration of the relationship.

With the diffusion of customer databases and the rise of customer relationship management (CRM) in the late 1990s, retention research moved from conceptual frameworks to more analytical and data-driven approaches. Rather than simply measuring satisfaction or tracking churn rates, firms began to apply statistical models to predict the likelihood of customer defection and to quantify the value of retaining a client.

The development of the customer lifetime value (CLV) framework provided the first systematic way to link retention with long-term profitability. Guota et Al. (2004) [176], find that a 1% improvement in retention has almost five times greater impact on firm value than a 1% change in discount rate or cost of capital.

As customer data became richer and more readily available in the 2000s, firms began adopting machine learning (ML) techniques to model retention and predict churn. Compared to classical statistical approaches such as logistic regression or survival analysis, ML offered the ability to capture nonlinearities, high-dimensional feature interactions, and complex temporal dynamics. Techniques like decision trees, random forests, and gradient boosting machines provided improved predictive accuracy while remaining interpretable enough for business use. Lemmens and Croux [181] (2006) applied bagging and boosting classification trees to a telecom churn dataset, reporting significant gains in predictive accuracy over logistic regression, as evidenced by improvements in validated Gini coefficients and top-decile lift. Neslin et al. [182] demonstrated that logistic regression and tree-based models were robust, interpretable, and profitable choices for churn prediction. In contrast, neural networks, discriminant analysis, and overly theory-driven approaches tended to underperform. Importantly, their benchmarking also showed that practitioners do not need to rebuild churn models on a monthly basis, since predictive performance remained stable for at least a three-month horizon. Coussement Van den Poel (2008) [183], showing that random forests outperformed logistic regression delivered the highest predictive accuracy in subscription churn settings. The most influential churn predictors were the length of the current subscription, the time since the last renewal, the month of contract expiration, the elapsed time since the last suspension, the recency of the last complaint, the age of the subscriber, and whether the subscription was initiated on the customer's own initiative rather than through marketing actions. Interestingly, and contrary to much of the CRM literature, monetary value and frequency did not emerge as significant churn drivers. On the other hand, Verbeke et al. (2012) [184] benchmarked 11 real-life churn datasets from telecom operators around the world. In their study, Decision Trees yielded the best overall performance in the experiments, although a large number of other techniques were not significantly outperformed. Hence, other properties of modeling techniques besides the predictive power have to be taken into account when choosing a classification technique, such as compre-hensibility and operational efficiency.

The last decade has seen the introduction of deep learning into customer retention modeling, paralleling broader advances in representation learning across domains. Suh et al. (2023) [185] developed a churn prediction model for a water-purifier rental service using real operational data (84 000 training customers; 250 000 validation contacts). The model achieved strong performance (AUC = 0.88, F1 = 0.93) and, importantly, attained an 80 % hit rate in identifying churners in live operations. Unlike prior cluster-level models, it generated customer-specific churn explanations, thus enabling individualized retention strategies. Joy et al. (2024) [186] proposed a hybrid deep learning model combining LSTM and GRU networks with LightGBM for churn prediction in streaming services. The model also included feature selection (Chi-square and Sequential Feature Selection), SMOTE for class imbalance, and explainability via SHAP. It achieved best-in-class performance with an AUC of 95.60% and an F1 score of 90.09%, outperforming both standalone ML and DL baselines.

The latest wave of AI research introduces large language models (LLMs) as potential tools for customer retention. Unlike earlier deep learning models optimized primarily for structured behavioral sequences or transactional data, LLMs can process and generate natural language, making them particularly valuable in domains where unstructured customer interactions: emails, chat logs, social media, or reviews, are central. Chajia et al.(2024) [187] investigated the use of LLM-derived embeddings, transforming textual and numerical customer data into semantic vectors, for customer churn prediction. They evaluated several classifiers and found that a logistic regression model built on LLM embeddings achieved the highest accuracy (89%), outperforming traditional feature encodings and demonstrating the added value of semantic representations in generalised churn modeling.

Recent industry deployments of AI illustrate its role in supporting customer retention strategies across sectors. Yum Brands has piloted AI-driven marketing campaigns using reinforcement learning to personalize email timing and content, reporting increased purchases and lower churn, while extending similar personalization to kiosks and mobile apps [188]. In the luxury sector, LVMH launched the generative AI platform *MaIA* to deliver personalized communications (e.g., Tiffany's), support marketing content creation, and even guide design and pricing decisions, with the aim of strengthening customer loyalty [189]. At the same time, industry analysts warn of a growing "trust recession": 91% of customers report frustrating AI service experiences, and 70% would switch brands after a single negative encounter, underscoring that AI must operate more as a “customer advocate” than a robotic gatekeeper [190]. Finally, AI-first customer service platforms such as Sobot demonstrate tangible results, with case studies citing up to 30% higher retention, 45% increased satisfaction, and 30% lower support costs; OPPO, for example, achieved an 83% chatbot resolution rate and a 57% repurchase boost [191].

While most retention research focuses on churn in subscription services, hospitality, or retail, the luxury automotive market presents distinctive dynamics. A defining feature is the long waiting period between order placement and vehicle delivery, often spanning months or even years for high-demand models. Unlike mass-market segments where retention efforts aim at repeat purchases or subscription renewals, luxury manufacturers must maintain the client's commitment during this interval.

In this context, retention strategies shift from preventing defection to sustaining engagement and reinforcing the emotional bond with the brand. The challenge is not simply to avoid cancellation of an order, but to transform the waiting period into part of the luxury experience itself. Studies of customer engagement in high-end services emphasize that communication, exclusivity, and experiential touchpoints are critical to sustaining loyalty in contexts where tangible product consumption is delayed [192, 193]. Recent literature indicates that luxury automotive firms are beginning to employ ML for advanced retention strategies. Predictive analytics, through decision trees, random forests, gradient boosting machines, and neural networks, is employed to forecast customer preferences, tailor marketing campaigns, and interpret customer sentiment from online feedback and social data [194]. However, few if any, studies address retention during the extended wait time between order placement and vehicle delivery, suggesting this remains a largely uncovered research opportunity. Predicting contract cancellations before delivery enables proactive and cost-effective interventions that preserve revenue and customer goodwill [33]. The value of early warning systems increases in settings where purchase and consumption are separated by long delays: the longer the wait, the greater the window for expectations to drift, external conditions to change, and the initial purchase intent to erode.

Within applied research on cancellation prediction, hospitality has served as the primary test bed, and a broad spectrum of statistical and machine-learning approaches has been explored. These include traditional ML [195, 196], Bayesian networks [35], and more recent deep-learning architectures [36, 197]. While these models often optimize for headline accuracy, they tend to trade off interpretability and portability. Although these works focus on maximizing classification accuracy, they often sacrifice interpretability and are difficult to transfer to other domains [198], given the sector-specific nature of customer decision processes and the fundamentally different products involved.

The luxury-automotive industry presents a complementary setting where demand frequently exceeds supply and the order-to-delivery interval can span several quarters. Managing this waiting period is operationally critical. Macroeconomic indicators (e.g., interest rates), market dynamics (e.g., new model launches), and brand-level activities (e.g., marketing events, test-drive opportunities) can all shift perceived value. To monitor large backlogs without overwhelming dealer staff, we propose a framework that flags high-priority contracts for human follow-up (e.g., personalized test drives), helping maintain engagement throughout the wait (Fig. 3.1).

This study makes three contributions, each aimed at balancing predictive performance with practical usability while addressing the research questions on waiting-list retention (**WL-RQ1–WL-RQ3**).

1. **Contract Risk Estimation.** A supervised classifier that combines external market conditions with internal contract and customer information to estimate the cancellation probability for each order (**WL-RQ1**).

2. **Interpretability of Predictions.** Model-agnostic explanations quantify how inputs influence the output, improving transparency and supporting accountable decision-making (**WL-RQ2**).

3. **Practical Tools for Dealerships.** A lightweight decision-support interface and an on-premises LLM that turn model outputs and pre-defined, market-specific rules into concise, contract-level recommendations suitable for daily operations (**WL-RQ3**).

Across internal data, the framework achieves a mean F1 score of 0.85; on a temporally out-of-sample hold-out collected after training, it retains strong generalization with F1 = 0.93. Explanations are provided with SHapley Additive exPlanations (SHAP) [49]. A retrieval-augmented LLM then converts the explanatory signals into plain-language guidance for dealership personnel in seconds. A high-end automotive partner positively evaluated these outcomes and is preparing to incorporate the tool's insights into decision processes.

3.2.1 Methods

Problem formulation

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i, t_i)\}_{i=1}^{n}$ denote the set of contracts observed during the training window, where $\mathbf{x}_i \in \mathbb{R}^d$ is the feature vector at decision time, $y_i \in \{0, 1\}$ indicates

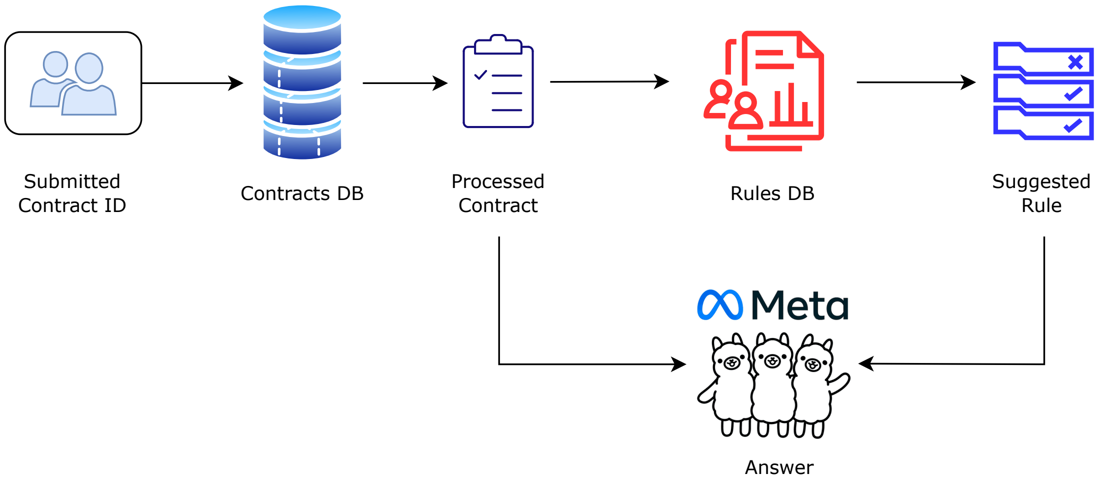

Figure 3.1: End-to-end framework: data ingestion, risk estimation, explanation, and recommendation delivery.

the outcome (*cancelled* vs *delivered*), and t_i indexes the contract's waiting-list (WL) trajectory. The goal is to estimate

$$p_{\text{cancel}}(\mathbf{x}) \triangleq \Pr(y = 1 \mid \mathbf{x}),$$

and to map it into a triage taxonomy for dealer use:

$$\text{tier}(\mathbf{x}) = \begin{cases} \text{low}, & p_{\text{cancel}}(\mathbf{x}) < 0.25, \\ \text{medium}, & 0.25 \leq p_{\text{cancel}}(\mathbf{x}) < 0.60, \\ \text{high}, & p_{\text{cancel}}(\mathbf{x}) \geq 0.60. \end{cases}$$

We optimize for balanced error using the F1 score, defined as $F1 = \frac{2\,\text{Prec}\cdot\text{Rec}}{\text{Prec}+\text{Rec}}$, considering the unequal distribution of classes. Explanations are required at both the global and local levels to justify actions and to facilitate process improvement.

Dataset

The corpus contains over 50,000 historical contracts, each either *delivered* or *cancelled*. WL durations span 90 to 1,000 days. At the data owner's request, precise observation windows and base cancellation rates are omitted for confidentiality.

Records that were still on the WL at the end of training are not used for training but are later revisited to form a strictly post-training hold-out (Sec. 3.3.2).

The initial schema comprised 35 attributes covering *Economic Landscape*, *Customer Characteristics*, *Customer Behaviour*, *Contract Attributes*, and *Vehicle Features*. After feature selection (Sec. 3.2.1), 15 variables were retained as the most informative set:

- `Market` (internal): geographical/economic area for the contract.
- `Customer Price Index (CPI)` (Bloomberg): market-level price changes [199].
- `Unemployment Rate` (Bloomberg): share of active labor force unemployed [199].
- `10-Year Bond Yield` (Bloomberg): long-term interest rate proxy [199].
- `Customer Status` (internal): account category (e.g., new, active).
- `HQ Rank` (internal): customer ranking from internal criteria.
- `Nr Cars in WL` (internal): concurrent WL contracts for the same customer.
- `New Cars delivered during WL` (internal): deliveries to the customer while on WL.
- `Total Set OPT` (internal): count of optionals in the contract.
- `WL Time` (internal): days from activation to cancellation, or to production-line entry for delivered contracts.
- `Coupé / Spider` (internal): body-type flag.
- `Share new clients` (internal): share of contracts for the model placed by new clients.
- `Residual value` (internal): quarterly mean estimated resale value by market and model.
- `Nr of Vehicles` (internal): customer's fleet size at start date.

- `Order collection speed` (internal): months to reach 80% of lifecycle volume for the model.

By construction, no direct personal identifiers are included in the retained feature set. WL-time conventions are aligned with operational milestones defined by the data owner.

Given the mix of numerical and categorical variables and the presence of high-cardinality categories, we employ CatBoost [101]. CatBoost's ordered statistics for categorical features and symmetric tree structures reduce the need for extensive encoding and help guard against target leakage during category handling. This choice prioritizes a strong accuracy–interpretability trade-off with minimal preprocessing burden. We apply Recursive Feature Elimination (RFE) [200] to identify a compact, high-signal subset. RFE iteratively trains the estimator, ranks features by importance, removes the least informative variables, and repeats until performance no longer improves materially. The resulting 15-feature set balances parsimony, stability, and ease of explanation. An elbow analysis indicated that adding or removing a few variables around this point produced negligible gains relative to the increased complexity. We tune core CatBoost hyperparameters (`depth`, `learning_rate`, `iterations`, `l2_leaf_reg`, `bagging_temperature`) using Optuna [173] with five-fold cross-validation. The objective is the mean F1 across folds. Final parameters are refit on the full training split and then held fixed for post-training evaluation. To approximate real-world deployment, we assemble a hold-out consisting of contracts still on WL at the training cutoff and re-label them based on subsequent outcomes during the next three months. This yields approximately 500 contracts and enforces strict temporal separation between training and evaluation. We use SHAP [49] to quantify global and local contributions. Global importance summarizes which factors matter most on average; local explanations decompose each prediction into feature-level additive effects. We report two dependence plots for the top drivers, `Total Set OPT` and `WL Time`, with axes suppressed per confidentiality constraints (Figs. 3.2–3.3).

Practical Solutions for Dealerships

We operationalize the pipeline with a browser-based interface built in Streamlit [201]. Contracts are embedded into two dimensions using t-SNE [104] and partitioned by k-means to create a visual map that lets staff quickly situate an order among peers. This view is an aid to orientation rather than a strict geometric representation.

An on-premises, open-weights LLM (LLaMA-3-8B [202]) is used as a controlled, retrieval-augmented layer: it fuses the classifier's output, SHAP-derived insights, and predefined market-specific rules into concise recommendations. On-premises inference avoids third-party processing and aligns with GDPR [203] and the AI Act [204] requirements. Figure 3.4 and 3.5 shows the workflow: a top card displays the tier and probability, a middle card lists the key drivers, and bottom cards propose context-aware actions drawn from codified playbooks.

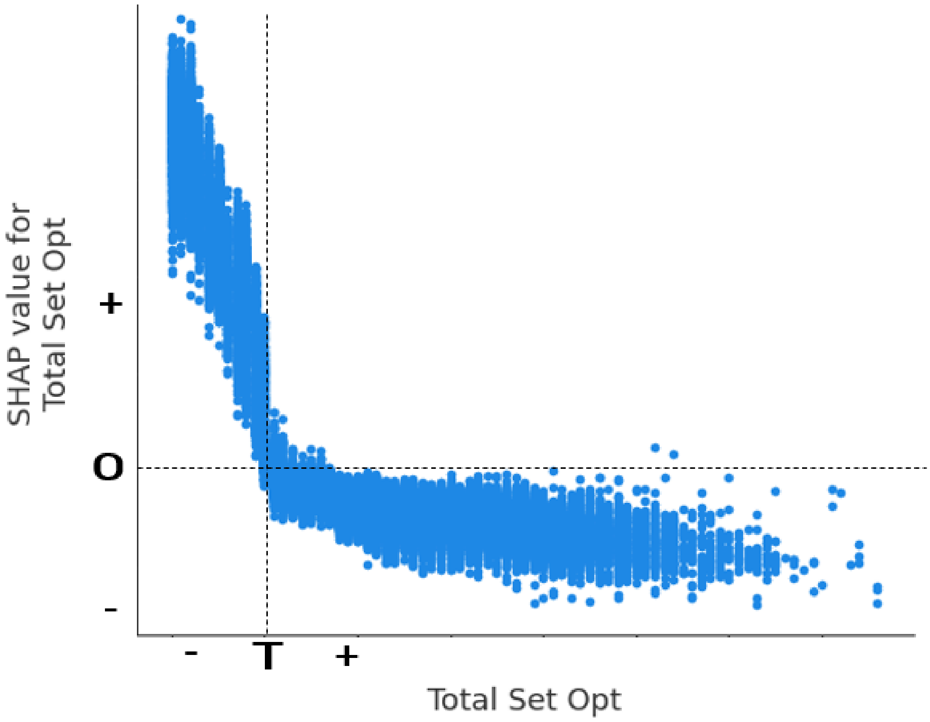

Figure 3.2: SHAP dependence for Total Set OPT. A data-driven threshold T (vertical dashed line, value withheld) marks the region in which scarcity of optionals is associated with a steep increase in risk; beyond T, the marginal effect flattens. Axes hidden for confidentiality.

The end-to-end interaction, from identifier input to recommendations, completes in about 10 seconds on an Apple M3 with 24 GB of GPU memory, supporting

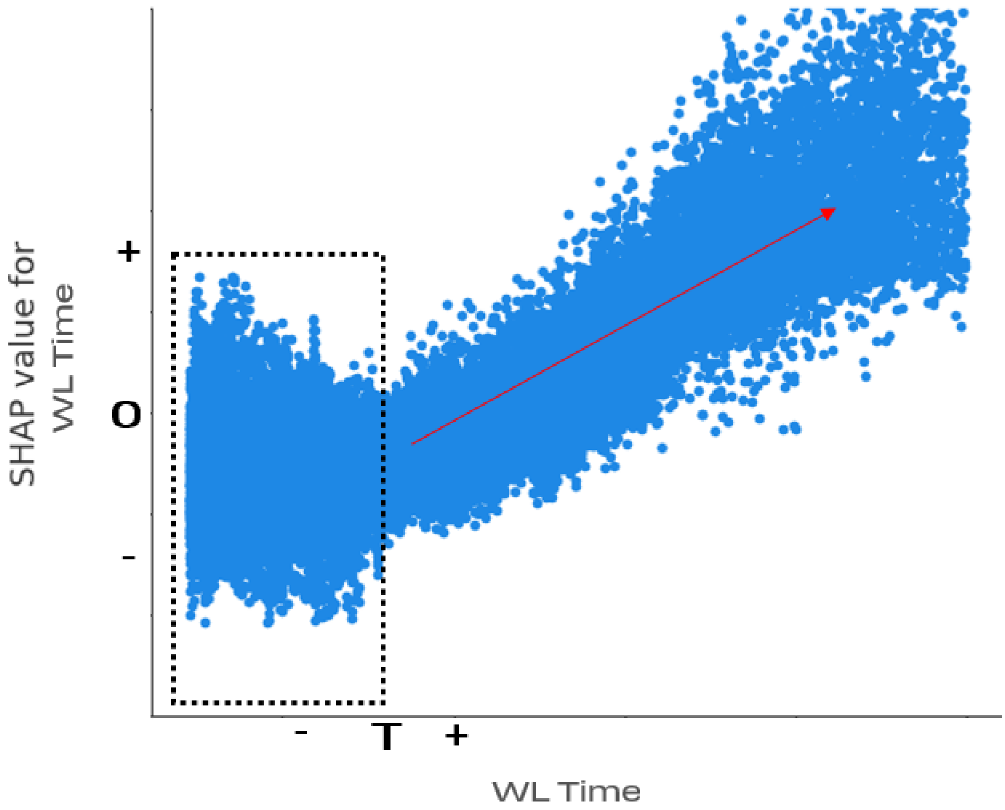

Figure 3.3: SHAP dependence for WL Time. The early WL phase exhibits low signal (dotted rectangle), followed by a roughly linear rise in feature impact after a second threshold T. Axes hidden for confidentiality.

real-time use at the desk.

3.2.2 Results and Discussion

Contract Risk Estimation and Interpretability

With the 15-feature subset and tuned hyperparameters, the model attains a mean F1 of 0.85 in five-fold cross-validation (Acc = 0.91, Prec = 0.91, Rec = 0.79). On the strictly post-training hold-out, it achieves F1 = 0.93 (Acc = 0.87, Prec = 0.96, Rec = 0.91), indicating robust temporal generalization to contracts that were unresolved at training cutoff.

Table 3.4 lists the dominant signals. The `Total Set OPT` dependence (Fig. 3.2)

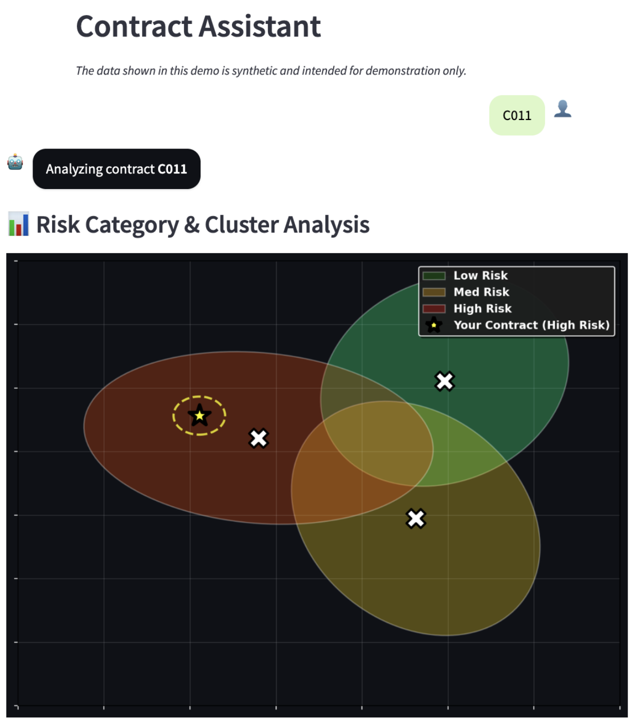

Figure 3.4: Example interface: contract query and 2D visualization for quick orientation.

shows a pronounced negative association with risk in the low-options regime, stabilizing beyond a threshold T. The `WL Time` dependence (Fig. 3.3) indicates a low-signal early phase followed by a roughly linear increase in influence beyond another T. These patterns are operationally actionable: early personalization and sustained engagement become progressively more important as WL grows.

Interface-level outcomes

In the dashboard, each flagged contract is shown with its tier, probability, and top SHAP drivers, reducing the cognitive load required to triage cases. The LLM layer consolidates these signals with predefined rules to surface short, concrete actions while keeping sensitive processing on-premises.

The classifier's performance (mean F1 = 0.85 in cross-validation; F1 = 0.93 post-training) is competitive with the best results reported in adjacent domains such as hospitality [197, 205], noting that exact cross-domain comparisons are

Figure 3.5: Example interface: risk assessment, SHAP-based drivers, and recommended actions.

imperfect due to different label balances, horizons, and covariates. Consistent with [35], ancillary purchase characteristics carry predictive value; here, `Total Set OPT` emerges as the top signal (Tab. 3.4).

Two interpretability aspects are worth underscoring. First, the strong effect of `Total Set OPT` in the low-option region suggests that thin configurations are fragile and benefit from earlier, more guided customization. Second, the non-monotonic profile for `WL Time` highlights a practical inflection: while the earliest weeks add little information, the salience of elapsed time increases thereafter. Monitoring how these thresholds T evolve by market provides a diagnostic for the effectiveness of WL management.

We treat contracts as the unit of analysis. Customers may appear in multiple contracts; cross-contract dependency is indirectly represented via features such as `Nr Cars in WL`, `Nr of Vehicles`, and `New Cars delivered during WL`. The model estimates association, not causation: importances and SHAP values describe

Feature	**Importance**
Total Set OPT	19.41
WL Time	12.53
Market	9.56
CPI %	8.65
HQ Rank	8.14
10Y Bond Yield	7.50
Residual Value	7.41
Unemployment %	6.07
% New Clients	4.98
Nr of Vehicles	4.58
Customer Status	3.27
Order Collection Speed	3.03
Nr New Cars Delivered During WL	1.89
Nr Cars in WL	1.58
Coupé / Spider	1.40

Table 3.4: Global feature importances from the fitted CatBoost model. Importances reflect model-specific contributions under the observed data distribution and should not be interpreted causally.

how the fitted decision function varies with inputs under the observed distribution.

Explanations are integral to auditability. Global charts guide process improvements (e.g., emphasizing configuration depth), while local attributions justify individual outreach. Running inference and the LLM on-premises avoids third-party data transfer and aligns with GDPR principles of data minimization and purpose limitation [203], as well as transparency and human oversight obligations emphasized in the AI Act [204].

Several factors limit generality. First, confidentiality constraints preclude release of base rates, time windows, and SHAP axes, which may affect external replication. Second, the data derive from a specific brand and product family; market mix and brand equity can shift signal strength across features such as `Market`, `Residual value`, and macro indicators. Third, the hold-out spans three months; longer horizons would further stress temporal robustness.

3.2.3 Contributions

We presented an interpretable, end-to-end framework for predicting cancellations in luxury automotive waiting lists. A CatBoost classifier, distilled to 15 variables via RFE and tuned with Optuna, achieves a mean F1 of 0.85 in cross-validation and 0.93 on a temporally separated hold-out, while SHAP yields transparent global and local attributions. A Streamlit interface and an on-premises, retrieval-augmented LLM transform these signals into compact, rule-grounded recommendations suitable for daily dealership operations.

Although still in an experimental phase, early validation indicates that the system can support faster triage, more timely outreach, and improved customer experience during extended WL periods. Future work will expand market coverage, diversify data sources, and explore compact transformer variants to test whether additional complexity can deliver incremental gains without sacrificing interpretability.

3.3 Residual value estimation

The automotive market is becoming increasingly complex due to several factors, including the proliferation of vehicle propulsion systems (internal combustion, hybrid, electric) and the multitude of available options, as well as rapid changes in the economic context and customer purchasing habits. To remain competitive, manufacturers must develop strategies that enable both reactive and proactive responses to market shifts [206]. In this respect, analyzing the pre-owned market dynamics, such as the distribution of pre-owned vehicles, their price evolution, and consumer valuation of used cars, can yield strategic insights for decision-makers [207]. When considering pre-owned vehicles, the most significant indicator is the residual value (RV), defined as the car's actual value at the end of ownership, typically expressed as:

$$\mathrm{RV} = \left(\frac{\text{Sale Price Preowned}}{\text{Sale Price New}}\right) \times 100. \tag{3.1}$$

In the used-car context, RV estimation is complicated by inherent market volatility and information asymmetries [208, 209]. For lessees and lessors, accurate

RV forecasts directly influence lease rates: overestimating RV can result in lower lease payments that fail to cover depreciation [210, 211], whereas underestimating RV leads to inflated lease rates, potentially deterring customers [212, 213]. Moreover, from a financial standpoint, RV estimates are crucial for quantifying the risk of deferred or defaulted leasing contracts [214, 215], and from an operational perspective, they enable continuous monitoring of negotiated buy-back prices versus expected market values, facilitating preemptive measures (e.g., targeted early buyout offers) to mitigate losses [216]. Residual value (RV) estimation methods have evolved from classical econometric approaches to machine learning (ML) techniques. Early hedonic and system-dynamics models provided foundational insights but often fell short when modeling complex, nonlinear depreciation patterns. For instance, Ohta & Griliches (1976) extended hedonic pricing for automobile prices, and Cowling & Cubbin (1972) developed hedonic indexes for UK cars. However, these methods require manual functional-form choices (e.g., logarithmic transformations for mileage) [217], and lack flexibility for modern, diverse datasets. Du et al. (2009) introduced a genetic-algorithm-based approach for reallocating vehicles to auctions, relying on a linear pricing model with manufacturer-suggested retail prices and segment-specific depreciation curves. They found that linear coefficients are difficult to estimate reliably for certain models and equipment segments, highlighting limitations of additive linear frameworks [218]. Wu et al. (2009) implemented an adaptive neuro-fuzzy inference system on a small dataset (192 used-car sales), showing improved performance when consolidating three equipment options into one predictor, but the model did not employ hold-out validation and used only four predictors [219]. More recent ML-based approaches consistently outperform linear models. Lessmann & Voß (2017) evaluated 19 regression methods on 454,618 German used-car transactions, finding that random forests achieved the lowest forecasting error when data are split randomly; however, they did not use chronological splits, differing from our sliding-window approach. Gleue et al. (2019) applied artificial neural networks to 250,000 transactions for a German OEM, demonstrating strong predictive power but noting that external factors (e.g., gasoline prices) did not improve forecasts. Liu et al. (2022) leveraged feature selection and particle swarm optimization on a 10,260-record Chinese used-car dataset, improving neural-network performance. Sharma & Mitra (2024) employed Multivariate Adaptive Regression Splines (MARS) on a 31,560-record Indian dataset, capturing nonlinear depreciation and demonstrating gains over linear models [220]. In the luxury segment, Ranjith et al. (2024) trained neural networks on the Indian pre-owned luxury market, outperforming econometric baselines [221]. However, most ML studies focus on general or region-specific markets, neglecting fine-grained model explanability and the integration of macroeconomic market variables. Huayi et al. (2022, 2023) addressed some of these gaps by using entity embeddings for categorical attributes and integrating macroeconomic variables (GDP, interest rate, exchange rate, CPI) on a New Zealand dataset, illustrating improved RV estimation in a single market context [222, 223]. In a similar way, Kim et al. (2023) assembled a proprietary dataset of 1.8 million used-car transactions across 600 models in Korea (2010–2022), considering vehicle characteristics and condition as well as market indicators [224]. Finally, Bergmann & Feuerriegel (2025) pioneered an automated, end-to-end preprocessing pipeline that maps nearly 50,000 German-language equipment identifiers into 18 high-level categories (e.g., navigation system, alloy rims). Their SHAP analysis revealed that features such as Android Auto/Apple CarPlay and keyless entry have the largest positive impact, whereas ubiquitous features (e.g., air conditioning, cruise control) have limited marginal value. Their work also highlighted the importance of chronological data splits (sliding-window cross-validation) to reflect real-world forecasting [225]. A fundamental limitation of the current literature is the lack of integrated modeling frameworks that simultaneously capture internal vehicle attributes (e.g., age, mileage, optional equipment costs) and macroeconomic indicators. Most contributions restrict attention to a single market and relatively short time spans (two to four years), which reduces the ability to generalize across geographies and to evaluate cyclical patterns. Incorporating longer horizons (e.g., ten years or more) and multiple markets would enable a more robust assessment of how both vehicle-specific and macroeconomic factors interact to shape residual value trajectories. Another limitation is the insufficient identification of depreciation thresholds. While most studies include variables such as vehicle age and mileage, they rarely quantify the specific inflection points at which residual values drop disproportionately. A further limitation concerns the absence of consistent baselines for isolating market-driven residual value trends. For example, shifts in the distribution of models, trims, or vintages included in the dataset can generate artificial variation in RV estimates.

This paper extends the literature by conducting the first large-scale, multi-region RV analysis for luxury cars using proprietary data from Ferrari S.P.A. covering Q1 2015 to Q2 2024. Under confidentiality, Ferrari provided 24,716 sales records across four anonymized geoclusters (A–D). Luxury cars are defined as high-prestige vehicles with MSRPs in the top market decile, built in limited volumes, featuring advanced materials, performance-oriented powertrains, and premium comfort or technology options. Our comprehensive approach has three objectives:

- **Residual value estimation:** Combining internal vehicle attributes (age, mileage, optional equipment cost, etc.) with three macroeconomic indicators (CPI, unemployment rate, 10-year bond yield) to build a robust model. By employing sliding-window validation over a ten-year span, including the pandemic period (2020–2021) (**RV-RQ1**).
- **Identification of impacts and depreciation thresholds:** Leveraging SHapley Additive exPlanation (SHAP) [226] to quantify how feature values, such as vehicle age and mileage, impact RV estimations, thereby identifying critical depreciation thresholds (**RV-RQ2**).
- **Trend Price Estimation:** Introducing the "Standard Vehicle" concept, defining a consistent approach (via linear regression on mileage and age, with fixed car characteristics) to isolate true market-driven RV trends from variations introduced by quarterly sample compositions (**RV-RQ3**).

After testing linear, machine learning (ML), and Deep Learning models (DL), CatBoost (ML), the best approach, achieved a mean absolute percentage error (MAPE) below 5%, comfortably under the industry acceptable threshold of around 10%. In addition, our SHAP analysis reveals that each continuous predictor exhibits a clear inflection threshold, but the precise point has not been revealed due to confidentiality reasons. For example, model age, months after the launch, has a positive effect on predicted RV during the early lifecycle, but each additional month sharply accelerates value loss, reaching a plateau after the threshold. Mileage and vehicle age behave similarly: low-use vehicles show a positive impact on RV until following a linear depreciation curve before and after the threshold is reached. Regarding macroeconomic variables, the Consumer Price Index (CPI) has a positive impact on RV for high values of inflation, whereas low values have neutral or slightly negative effects. While unemployment rate has an inverse impact with respect to CPI, for low values, it is beneficial for RV; for high values, the opposite. Finally, applying the Standard Vehicle methodology reduced spurious quarterly RV fluctuations by approximately 50%, resulting in a significantly smoother and more interpretable trend line. Nonetheless, isolating and then neutralizing the effect of sample composition variability (mileage, age, option rate) via the Standard Vehicle approach ensures that observed trend deviations genuinely reflect macro-driven market shifts (IR, CPI, Unemployment) rather than quarterly sampling noise. Ferrari S.P.A., the provider of the data, finds our tool accurate and useful and is now integrating the insights from our analysis of RV and preowned markets into the decision-making processes for product development, pricing, and car configurations.

3.3.1 Methods

In this section, the methodologies applied throughout our study are discussed. All the methodologies proposed are coded using Python (using Pandas, Scikit-Learn, CatBoost, XGBoost Python API, and TabNet's PyTorch implementation), and "no-code tools" are not used. For confidentiality reasons, hyperparameters, code, and data will not be publicly released. In the first subsection, a description of the data extraction process, highlighting how the proprietary data is structured and enriched with external economic indicators (Subsec. III.A). Then the focus is on the modeling techniques employed, outlining the selection process of various ML and DL models and explaining the sliding-window validation (Subsec. III.B). subsection III.C explores the model explainability approach, focusing on implementing SHAP to interpret feature impacts at both aggregate and individual levels. Lastly, in subsection III.D, a discussion of trend price estimation, detailing the “Standard Vehicle” method to isolate market effects from sample biases.

Data Extraction and Feature Selection The proprietary data encompasses a total of 24,716 sales, recorded from Q1 2015 through Q2 2024, spanning 16 distinct luxury car models. The data refers to four source regions, markets, covered in four anonymized geoclusters (A–D), as the data is granted under a confidentiality agreement. The selected features: *vehicle_age*, *distance*, *model_age*, *model*, *geocluster*, *val_car*, *val_opt*, *opt_ratio*, *cylinders*, and *Segment* are chosen based on their known relevance to residual value behavior, availability in the proprietary dataset, and validation through discussions with Ferrari S.P.A. industry experts. External macroeconomic variables: *CPI*, *IR10Y* (10-year bond yield), and *Unemployment* are collected from Federal Reserve Economic Data (FRED)[**mccracken2016fred**], aggregated to quarterly values. Table 3.5 lists all features and their types.

Table 3.5: Feature Description

Feature Name	**Description**	**Type**
Date	Day, month, and year of the sale	Date
vehicle_age	Age of the vehicle in number of months	Int
distance	Value in KM from odometer	Float
model_age	Age of the model in number of months	Int
model	Name of the model	Cat
geocluster	Market where the cars sold	Cat
val_car	The cost of the car without opt. cost in EUR	Float
val_opt	The cost of opt. features of the vehicle in EUR	Float
opt_ratio	val opt / val car	Float
cylinders	Number of cylinders: 12, 8 or 6	Int
Segment	SportCars or Pilot	Cat
RV	Residual Value of the car in percentage	Float
CPI	CPI harmonized value, index 2015	Float
IR10	Interest rate 10 year bond	Float
Unemployment	Unemployment Rate in percentage	Float

Model Selection and Evaluation Methodology We approach RV estimation as a regression problem. Building on prior findings that non-linear models capture threshold effects and feature interactions more effectively than linear models [227], we evaluate several techniques: Linear Regression (LR), Random Forest (RF), XGBoost (XGB), CatBoost (CB), and TabNet (TN), as detailed in Table 3.6. Mean Absolute Percentage Error (MAPE) served as our primary metric, consistent with industry standards and previous studies [227, 228]:

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{RV_i - \widehat{RV}_i}{RV_i}\right|. \tag{3.2}$$

Sliding-window validation is employed to ensure temporal consistency and prevent information leakage: each quarter from Q1 2019 through Q2 2024 served as a test set, with all preceding quarters as training data (e.g., Q1 2019 trained on Q1 2015–Q4 2018; Q2 2019 trained on Q1 2015–Q1 2019, etc.). This approach simulates real-world forecasting under evolving market regimes (e.g., pandemic shocks) and yields a robust performance distribution.

Table 3.6: List of Techniques

Model	**Type**	**Description**
Linear Regression (LR)	–	Linear model with constant marginal effects.
Random Forest (RF)	ML	Ensemble of decision trees via bootstrap aggregation.
XGBoost (XGB)	ML	Scalable tree boosting with regularization.
CatBoost (CB)	ML	Gradient boosting optimized for categorical features.
TabNet (TN)	DL	Attentive interpretable tabular networks.

Model Explainability via SHAP Once the ML/DL model demonstrating the best performance based on evaluation metrics is selected, SHAP is integrated to quantify the impact of each feature on the model's predictions. SHAP is a unified framework grounded in cooperative game theory that assigns each feature an importance value for a particular prediction. This approach facilitates both global and local interpretability of complex models by decomposing predictions into the individual contributions of each feature. The appropriate SHAP explainer is chosen according to the model type: for tree-based algorithms, TreeExplainer is employed; for neural networks, DeepExplainer is utilized. The SHAP library in Python is used to compute SHAP values for each feature across all instances in the test set. The entire dataset is input into the SHAP explainer to obtain the contribution of each feature to every individual prediction, thereby breaking down the model's output into additive components attributable to each feature. After computing SHAP values, a relationship between each feature's values and their corresponding percentage impact on RV is established. This is achieved by aggregating the SHAP values to determine the mean absolute SHAP value for each feature, providing a global perspective on feature importance across the entire dataset. To visualize these relationships, SHAP dependence plots are generated. Dependence plots illustrate how a feature's SHAP value changes with the feature's actual value, effectively highlighting both the magnitude and direction of the feature's impact on the prediction. For example, a dependence plot for mileage might reveal that, up to a certain threshold, mileage has a minimal negative impact on RV, but beyond that point, its negative influence becomes more pronounced. These plots are instrumental in identifying critical thresholds where a feature's influence shifts significantly. Variation in these threshold values across different models and markets provided a deeper understanding of which factors are most impactful in specific contexts. This analysis enabled the identification of potential risks or strengths within certain vehicle models or regions, thereby informing strategic decision-making. Additionally, an instance-level analysis is conducted by examining SHAP values for specific sales, allowing the precise influence of each feature on individual predictions to be understood. In the next section, the concept of thresholds is illustrated with concrete examples, demonstrating how specific feature values influence RV outcomes in practical scenarios.

Trend Price Estimation: The "Standard Vehicle" Concept A quarterly comparison is conducted between the mean residual values (RVs) derived from actual transaction prices and those estimated by the ML model for the same sample of vehicles. The primary objective of this comparison is to monitor short-term fluctuations and detect potential shifts in the market. By analyzing these discrepancies on a quarterly basis, the model's ability to accurately capture real market behavior and identify abrupt changes that might signify evolving trends or economic influences is assessed. This temporal analysis enabled evaluation of the model's responsiveness to market dynamics and its effectiveness in reflecting the underlying economic conditions affecting vehicle residual values. However, the model's estimations are inherently influenced by the characteristics of the specific sample of cars in each quarter, such as variations in mileage, vehicle age, and option rate (the proportion of optional features selected). These factors can exhibit significant variability across different samples, potentially skewing the observed trends in residual values. For instance, if a particular quarter's sample comprises a higher proportion of older vehicles or cars with elevated mileage, the mean residual values may appear lower, even in the absence of actual changes in market conditions. Consequently, the trends observed might not solely reflect true market shifts but could also be affected by the composition of the vehicle samples. To ensure that the observed trends genuinely reflect market shifts rather than fluctuations in sample composition, the model is assessed by analyzing the number of directional changes in the market. An excessive number of directional shifts can obscure meaningful patterns and hinder the readability of the results. Therefore, a smaller number of consistent directional changes is preferable, as it enhances the clarity of trend detection and provides a more accurate representation of the true market direction. This approach allows differentiation between genuine market movements and anomalies caused by sample variability. To mitigate the potential bias introduced by varying sample characteristics, an additional estimation is introduced, the “Standard Vehicle” concept. This concept is defined by a linear variation in key features, enabling trends in ML-estimated residual values to be tracked while isolating the impact of sample-specific attributes. For each vehicle model and geographical cluster, the standard vehicle’s attributes, such as mileage and vehicle age, are determined through linear regression analysis. Additionally, average values are applied for price and option rate to maintain consistency across different samples. This standardized approach ensures that residual value trends are reflective of true market conditions rather than being confounded by variations in the vehicle samples. By controlling for these key variables, the Standard Vehicle model provides a reliable baseline for assessing market trends and enhances the robustness of the analysis. Variation in threshold values across different models and markets provided deeper insights into which factors are the most impactful in specific contexts, enabling identification of potential risks or strengths within certain vehicle models or regions and informing strategic decision-making. Exploratory Data Analysis To explore the relationships between the variables in our dataset, a correlation matrix, shown in Figure 3.6, reveals three distinct clusters of highly correlated variables, each representing different aspects of the data. The first cluster, located in the top-left quadrant of the matrix, includes *vehicle_age*, *distance*, and *model_age*, which collectively represent the condition of the vehicle. The second cluster, situated in the center of the matrix, comprises *val_car*, *val_opt*, *opt_ratio*, and *cylinder*. These variables describe the intrinsic characteristics of the vehicle models, such as the proportion of optional

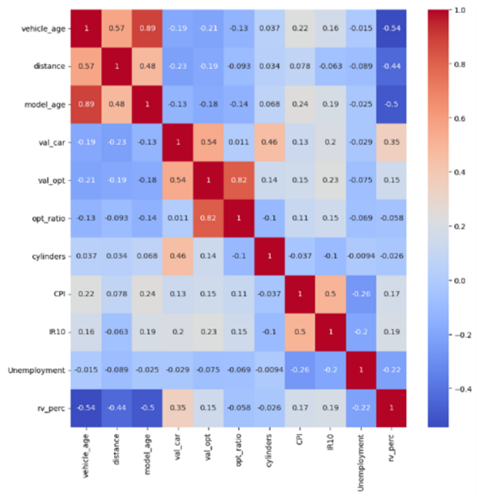

Figure 3.6: Correlation Matrix of Variables in the Dataset

features and engine specifications. The third cluster, found in the bottom-right quadrant, consists of macroeconomic indicators: *CPI* (Consumer Price Index), *IR10Y* (10-Year Interest Rate), and *unemployment*. These variables capture the broader economic environment. Additionally, the last row of the correlation matrix indicates the correlation of each feature with RV in percentage terms. The analysis reveals that vehicle condition variables (*vehicle_age* and *distance*) as well as *model_age* have the highest correlation with RV, underscoring their critical role in determining vehicle valuations. However, during the study period, the distribution of RV has evolved, influenced by shifts in the macroeconomic scenario and changes in model characteristics, as highlighted in the following analysis.

Market-Specific Residual Value Distribution Changes

Figures 3.7 and 3.8 illustrate the changes in RV distributions over time within two distinct market areas, designated as Market “A” and Market “B”[1]. These figures highlight how RVs are impacted differently across markets, particularly during the COVID-19 pandemic period, from 2020 to 2022.

In both markets, there is a noticeable decline in RV distributions during the COVID-19 period compared to the pre-pandemic period (Before COVID). However, the recovery patterns differed between the two markets:

- **Market “A”:** The RV distribution did not fully recover post-pandemic,

[1] Market names are hidden due to confidentiality agreement

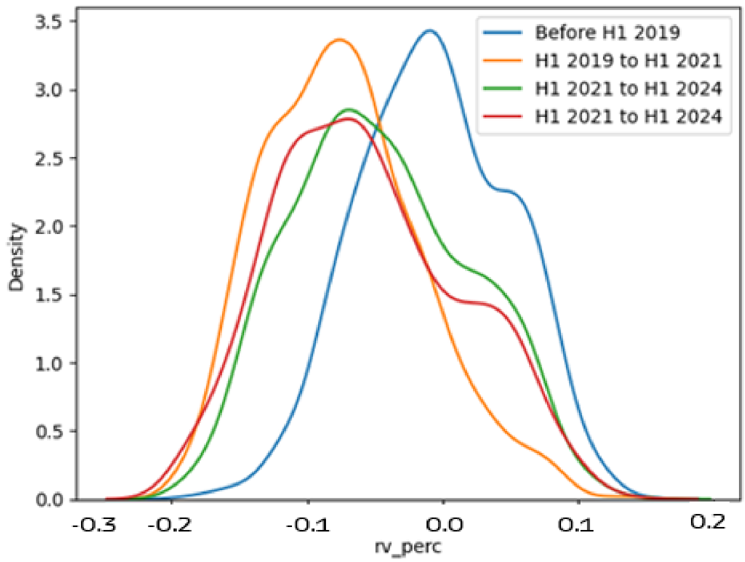

Figure 3.7: Change in Residual Value Distribution Due to COVID-19 in Market "A"

remaining significantly lower than pre-COVID levels. This persistent decline suggests that Market "A" may have been more severely affected by the pandemic, possibly due to region-specific economic challenges or shifts in consumer behavior that have not yet stabilized.

- **Market "B":** In contrast, the RV distribution post-pandemic in Market "B" shows a trend similar to the pre-COVID period, indicating a partial or complete recovery of residual values. This suggests that Market "B" is either less affected by the pandemic or is able to adapt more effectively to the changing conditions, allowing residual values to rebound.

These observations lead to two key insights:

1. **Changing RV Distribution Over Time:** The RV distribution has evolved over the years. This indicates that external factors are influencing the RV metric.
2. **Differential Impact Across Markets:** The pandemic's impact on RV varied between markets, likely due to region-specific conditions or differences

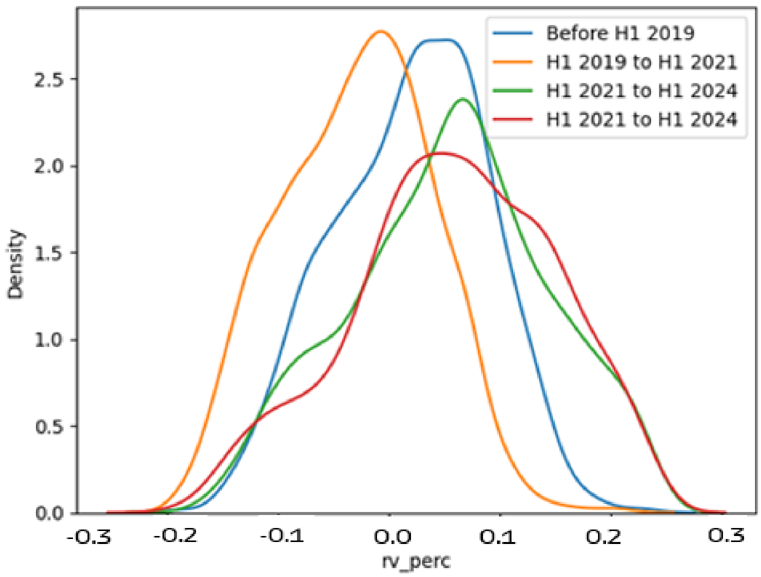

Figure 3.8: Change in Residual Value Distribution Due to COVID-19 in Market "B"

in customer behavior.

3.3.2 Results and Discussion

Models Comparison

Table 3.7 presents the Mean Absolute Percentage Error (MAPE) results for each of the DL/ML techniques evaluated in our study and presented in the next section. Among the models considered, the CatBoost (CB) model emerged as the top performer, achieving the lowest mean MAPE of 5.58%. The TabNet (TN) model also demonstrated strong performance, with a mean MAPE of 5.75%, closely trailing behind CB. This highlights the effectiveness and potential of transformer-like architectures when applied to tabular data. While Linear Regression (LR) showed a respectable mean MAPE of 7.57%, maintaining an overall error below 10%, further analysis revealed that over 27% of its predictions had a MAPE exceeding this threshold (Table 3.8). This indicates that although LR performs reliably on average, it encounters significant errors in specific segments of the data. These findings imply that more sophisticated models like CB and TN are better suited for capturing the complexities inherent in the dataset, providing more consistent results across different vehicle samples. The Random Forest (RF) and XGBoost (XGB) models exhibited moderate performance, with mean MAPEs of 6.51% and 6.13% respectively. Their results positioned them between the high-performing CB and TN models and the more basic LR model. Notably, RF and XGB outperformed LR in terms of both mean MAPE and the percentage of errors exceeding 10%, indicating a balanced trade-off between model complexity and predictive accuracy.

Table 3.7: MAPE Distribution

Model	**Mean**	**Min**	**Max**	**Std**
LR	7.57	5.51	11.70	± 1.96
RF	6.51	4.78	11.36	± 1.65
XGB	6.13	5.04	10.14	± 1.31
CB	5.58	4.28	8.14	± 0.99
TN	5.75	4.80	8.15	± 0.96

Table 3.8: Percentage of Predictions with MAPE Above 10%

Model	**Percentage of Errors (%)**
Linear Regression (LR)	27.34%
Random Forest (RF)	20.03%
XGBoost (XGB)	21.10%
CatBoost (CB)	15.04%
TabNet (TN)	15.24%

The superior performance of CB and TN techniques can be attributed to their ability to capture complex, non-linear relationships within the data, as well as their robustness in handling various feature interactions. In contrast, LR, while simple and interpretable, lacks the capacity to model such intricate patterns, leading to higher error rates in specific segments. Furthermore, the percentage of predictions with MAPE exceeding 10% provides additional insight into each model's reliability. CB and TN techniques not only achieved lower mean errors but also maintained a more consistent performance across the dataset, with only approximately 15% of their predictions surpassing the 10% MAPE threshold. This consistency underscores their suitability for applications requiring high accuracy and reliability.

Extraction of Variable Contributions

To understand the influence of individual variables on the model's predictions, a detailed analysis is conducted using the CB model, since the previous analysis demonstrated its superiority over the other tested techniques. Table 3.9 presents the feature importance scores, revealing that over 60% of the total importance is attributed to just four variables: *model*, *geocluster*, *vehicle_age*, and *CPI*.

Table 3.9: Feature Importance in CatBoost (CB) Model

Feature	**Importance (%)**	**Cumulative Importance (%)**
Model	23.18	23.18
Geo_cluster	18.85	42.03
Vehicle_age	12.49	54.52
CPI	10.20	64.72
Unemployment	6.43	71.15
Model_age	6.34	77.49
Distance	5.05	82.54
Val_car	4.60	87.14
Opt_ratio	4.56	91.70
IR10	3.72	95.42
Segment	2.29	97.71
Val_opt	1.92	99.63
Cylinders	0.37	100.00

The ranking in Table 3.9 highlights that, aside from *vehicle_age*, the most influential factors are predominantly categorical (such as *model* and *geocluster*) or external economic indicators (*CPI*). The *model* variable captures differences associated with each car model's unique characteristics, including design and consumer appeal, which are not directly measured by other features. The *geocluster* variable accounts for region-specific behaviors and practices, such as local taxation policies or market rotation speeds, which significantly influence residual value (RV) outcomes. External economic factors, namely *CPI* (Consumer Price Index) and *unemployment*, also play crucial roles, reflecting the broader economic environment's impact on residual values. The significant influence of these indicators underscores how economic conditions, such as those experienced during the pandemic, have shaped market behavior and car valuations during the study period. Focusing on continuous features directly related to the vehicle, *vehicle_age* emerges as the most critical factor. Following this, *model_age* is also significant, indicating the demand strength for newer versus older models. *Distance* (vehicle mileage) and *Opt_ratio* (the ratio of optional features to the base price) further contribute, demonstrating how usage intensity and customization levels affect RV. These findings are in line with the expectations considering existing literature and align with industry expert opinions. To gain deeper insights into how each feature affects the model's predictions, SHAP is applied to the CB model. SHAP values allow for the visualization and interpretation of the direction and magnitude of each feature's impact on the RV metric. Figure 3.9 presents a SHAP summary

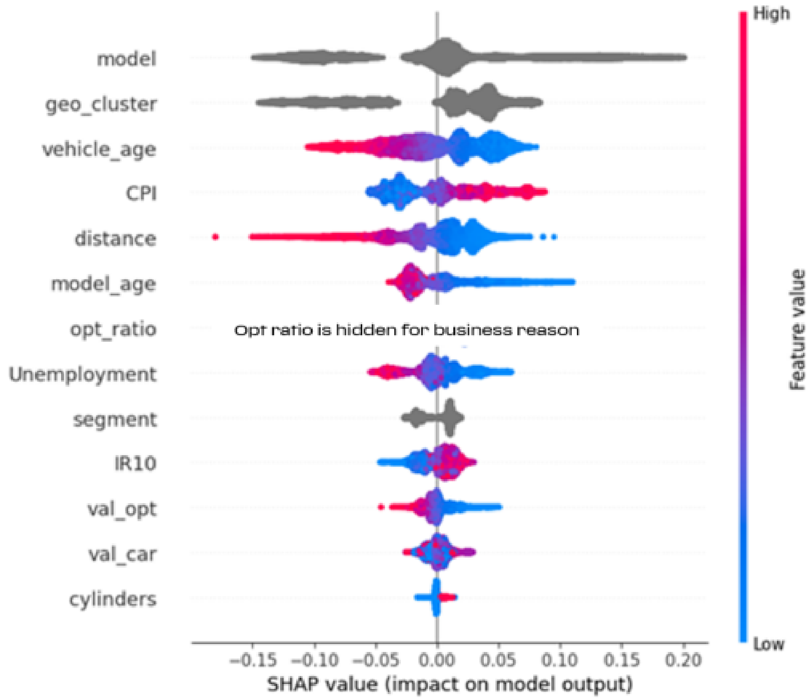

Figure 3.9: SHAP Summary Plot: Feature Impact on Residual Values

plot for the CB model, showing how each feature contributes to the predicted residual value (RV) across all sales records. The features are sorted by mean absolute SHAP value (vertical axis), so “model” and “geo_cluster” appear at the top, followed by “vehicle_age”, “CPI”, “distance”, and so on. Each point in the violin-shaped band represents a single vehicle sale, with its horizontal position indicating the SHAP contribution (negative values push the prediction lower; positive values push it higher). For continuous features, the color gradient (blue =

low, red = high) shows the raw feature value; for categorical features, points are shown in grey (since no numerical "high" or "low" applies). Key insights include:

- **Vehicle Age**: High values of vehicle_age (red) lie almost entirely to the left of zero, meaning that older cars consistently pull the RV prediction downward. Conversely, low vehicle_age (blue) often appears to the right of zero, indicating a positive contribution to RV for younger vehicles.

- **CPI**: When CPI is high (red), SHAP values cluster on the positive side, indicating that higher inflation increases estimated RV. Lower CPI (blue) is nearer zero or slightly negative, meaning mild inflation does not reduce RV as much.

- **Distance (Mileage)**: Higher mileage (red) similarly produces negative SHAP values; as distance increases, points shift left, confirming that more kilometers driven lowers residual value. Lower mileage (blue) appears closer to zero (neutral) or slightly positive.

- **Unemployment Rate**: Although less influential than CPI, higher unemployment (red) nudges SHAP values slightly to the left, a modest negative impact on RV, while lower unemployment (blue) is more neutral or mildly positive.

- **Model and Geo_Cluster (Categorical)**: Plotted in grey because they are nominal categories with huge variation in terms of SHAP values. The model treats "model" as a proxy for intrinsic characteristics (brand reputation, rarity, technical specifications, and standard options) and "geocluster" as a proxy for extrinsic conditions (local economy, regulations, cultural preferences).

These insights are also supported by the correlation relationship shown in the figure 3.6. In addition, it is possible to illustrate the impact of features on individual predictions, an example is presented in Figure 3.10. This SHAP force plot breaks down the RV estimation for a single vehicle, starting from the base value, the average RV expected for a random vehicle in the dataset. Each feature either increases or decreases this base value. In this specific example, *distance* (mileage) and *model_age* contribute negatively, reducing the RV, while *Opt_ratio* and *vehicle_age* have a positive effect, resulting in a final RV of 95%. Additionally,

Figure 3.10: SHAP Force Plot: Example of most important feature contributions for a single Vehicle

Figure 3.11 illustrates the SHAP values for the *distance* feature, focusing on mileage's impact on RV. The plot shows a generally linear, descending trend, indicating that higher mileage typically reduces the RV. However, the rate of depreciation decreases at higher mileage values, suggesting a diminishing impact as mileage increases. Similarly, in Figure 3.12, *vehicle_age* demonstrates a linear declining influence on RV, with a premium effect observed for newer models early in their lifecycle. This premium diminishes as the vehicle ages, reflecting decreasing market demand for older models. Figure 3.13 further refines this perspective by examining the *model_age* feature. The SHAP values reveal a strong positive influence on RV when the model is newly introduced to the market (period T-1), confirming the presence of a novelty or exclusivity premium. This influence sharply declines over time, stabilizing after approximately period T+1 from launch. In all three cases, the intersubsection of the dotted lines marks a critical threshold where the depreciation starts. Observing shifts in this threshold over time can reveal important trends in market behavior, such as changes in consumer preferences or broader economic factors that influence the speed at which vehicles depreciate.

Model Prediction vs. Standard Vehicle Estimation

In Figure 3.14, a quarter-by-quarter comparison of RV estimations derived from our CB model against the actual transaction prices within a specific market area for a particular car model is presented. This figure visually contrasts the average RV predicted by the ML model (blue line) with the mean actual RV observed in

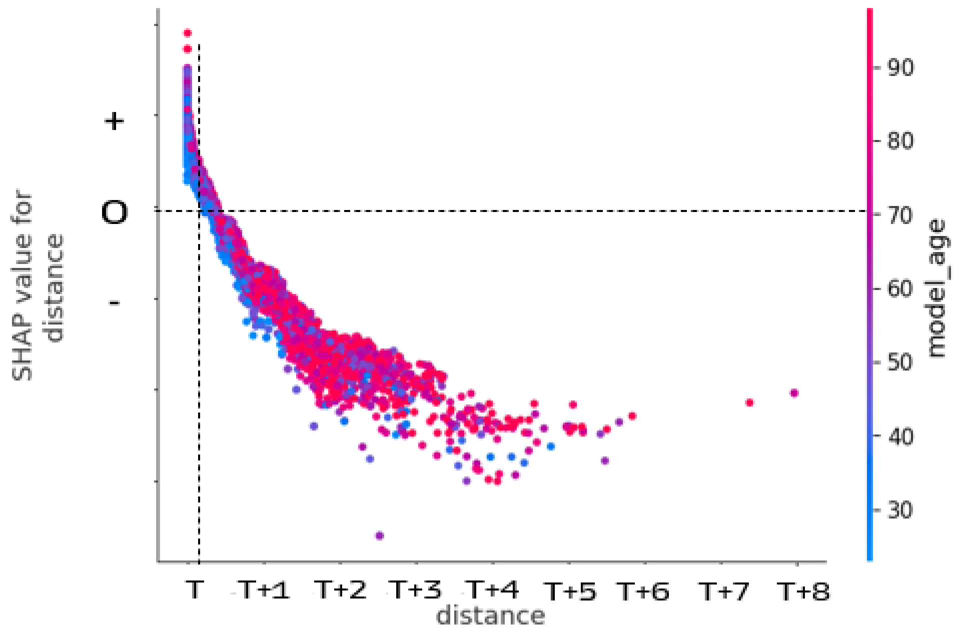

Figure 3.11: SHAP Values for Mileage: Identifying Depreciation Thresholds

transactions each quarter (orange line). The RV distribution within the sample, depicted by the orange line, exhibits notable fluctuations and inconsistencies, especially throughout 2023.

These fluctuations in the actual RV values are primarily due to the varying conditions and characteristics of the vehicles in each quarterly sample, such as differences in mileage, vehicle age, and option rate (the proportion of optional features chosen). While the ML model's estimates (blue line) effectively smooth out some of this inherent noise, they remain partially influenced by sample variability. This influence is particularly evident during periods of economic instability, such as in 2020, when external factors significantly affected the market. To address the bias introduced by varying sample compositions, an alternative estimation method using a "Standard Vehicle" model is employed, as defined in subsection III.D. This approach allows us to track trends in ML-estimated RVs while isolating the impact of sample-specific characteristics. The results of this estimation, focusing on key features such as distance (mileage) and vehicle age, are summarized in Table 3.10. The "Standard Vehicle" model estimates RVs by defining a vehicle with linear variations in key features, such as mileage and age, while maintaining average

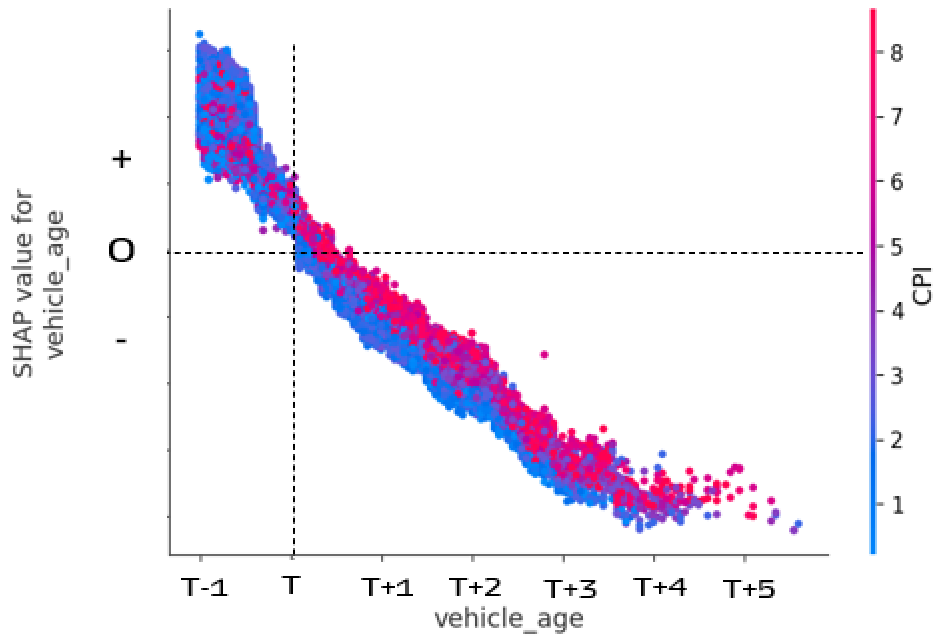

Figure 3.12: SHAP Values for Vehicle Age: Identifying Depreciation Thresholds

Table 3.10: Performance Metrics for Key Features in Standard Vehicle Estimation

Feature	$\mathbf{R}^2$	**RMSE**	**MAE**
Distance	0.23	9433.65	5800.12
Vehicle Age	0.79	10.12	8.30

values for price and option rate to ensure consistency across different samples. Figure 3.15 illustrates the RV estimations using this standardized approach. Comparing Figures 3.14 and 3.15, it is evident that the "Standard Vehicle" model provides a smoother and more consistent RV trend compared to the raw sample-based estimations. Furthermore, Table 3.11 summarizes the number of directional trend changes observed in the RV estimations for each geocluster-market and vehicle model pairing, reporting aggregated statistics (mean, median, maximum, and minimum) computed across all trials. The analysis reveals that the "Standard Vehicle" model significantly reduces the number of fluctuations and directional shifts, thereby enhancing the readability of market trends and decreasing the occurrence of misleading market shifts by approximately 50%.

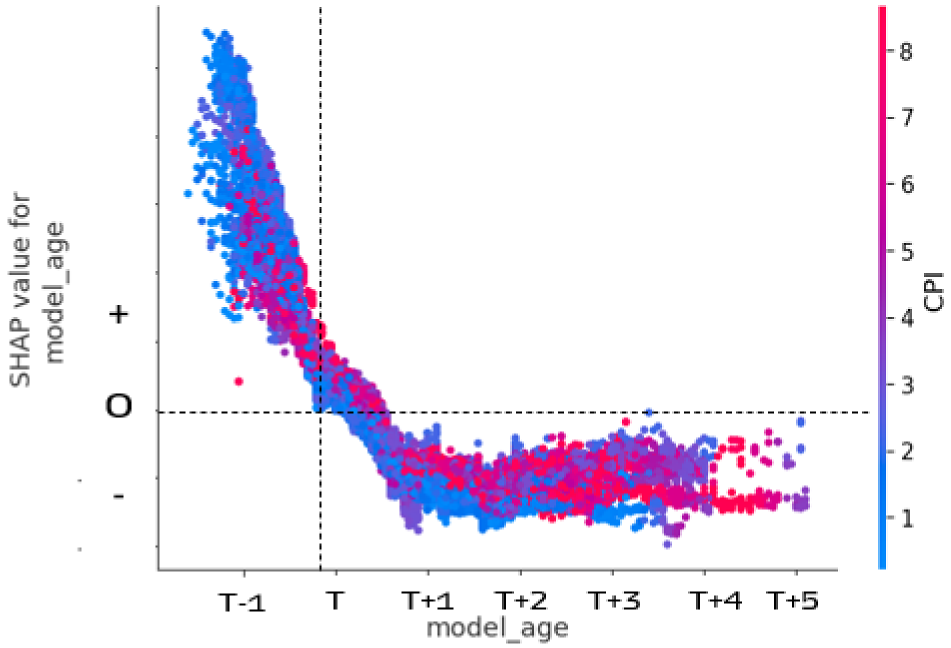

Figure 3.13: SHAP Values for Model Age: Identifying Depreciation Thresholds

Table 3.11: Number of Directional Trend Changes in RV Estimations for each geocluster-market and vehicle model pairing, reporting aggregated statistics

Method	**Mean**	**Median**	**Min**	**Max**	**Std Dev**
Sample	7	6	2	23	5
Standard Vehicle	4	3	0	16	3

3.3.3 Contributions

This study examined the complexities of RV estimation for luxury cars over multiple years and across various markets. The results demonstrate that CatBoost outperforms alternative models, achieving a mean MAPE of 5.58%, a value in line with the current state-of-the-art and below the 10% industry threshold. The analysis (see Tab. 5) reveals that RV is predominantly influenced by the car model and market geocluster, highlighting the need to consider diverse regions and models. Internal discussion with experts in the field has concluded that the model treats “model” as a proxy for intrinsic characteristics (brand reputation, rarity, technical specifications, and standard options) and “geocluster” as a proxy for extrinsic conditions (local economy, regulations, cultural preferences). Additionally,

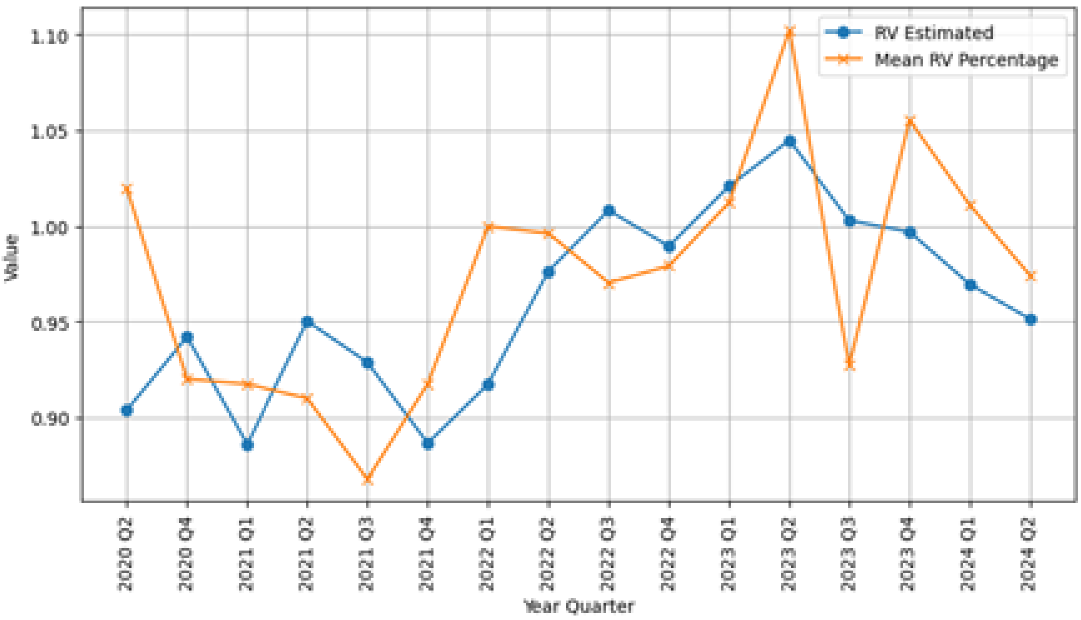

Figure 3.14: Quarter-by-Quarter Residual Value Estimation: ML Model vs. Actual Transaction Prices for a Specific Market Area and Car Model.

car condition, especially vehicle age, and macroeconomic factors such as the Consumer Price Index (CPI) and unemployment rate significantly affect RV. While market region and car model are the primary determinants, economic indicators are nearly as influential as vehicle age. This supports the importance of considering both internal and external factors in estimating RV. By focusing on key metrics like mileage and vehicle age, our approach can determine not only when depreciation begins (for example, in terms of months or mileage) but also quantify the rate at which value declines. This enables rapid, practical assessments; for example, Fig. 9 allows practitioners to quickly correlate specific mileage or age thresholds with corresponding depreciation losses, even without the full ML model. Finally, the introduction of the "standard vehicle" concept enables the isolation of genuine market depreciation trends by controlling for sample variability through standardization. This approach offers a more robust representation of RV changes over time and, given the typically low transaction volumes in the luxury car market, constitutes a significant advancement over existing methodologies for experts in the field. However, this study has several limitations. First, all findings are specific to luxury vehicles; extending these results to non-luxury

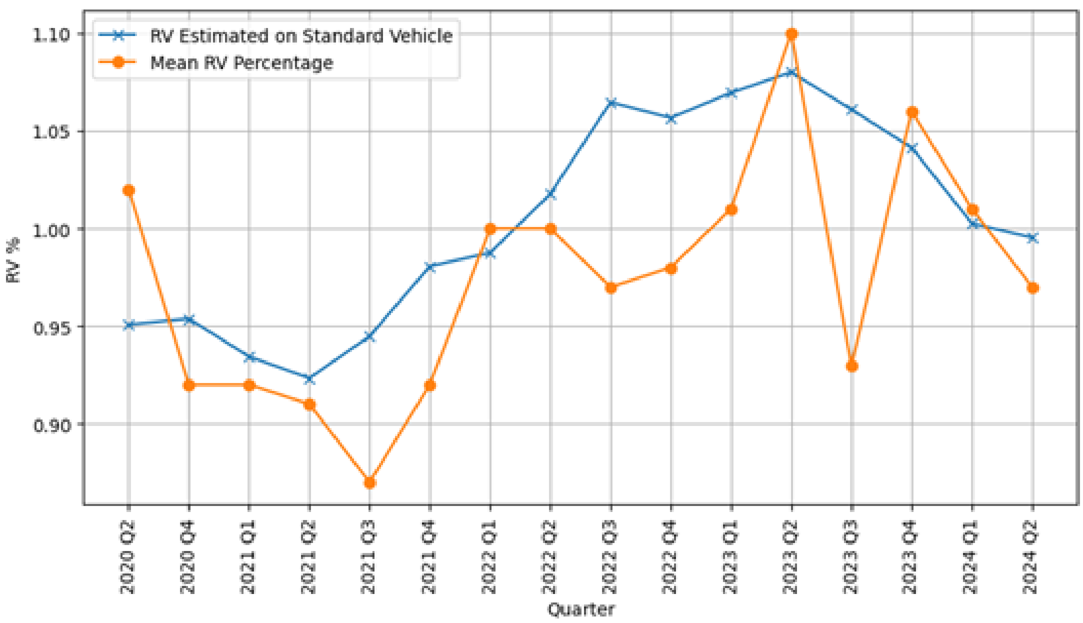

Figure 3.15: Quarter-by-Quarter Residual Value Estimation: Standard Vehicle Model vs. Actual Transaction Prices.

brands would require representative datasets for those segments and external validation to ensure generality. Second, only three macroeconomic variables, CPI, unemployment rate, and 10-year bond yield, are included; broader PESTEL factors (for example, environmental regulations or social trends) are not incorporated and remain avenues for future work. Third, no external market validation (such as public auction data comparison) is performed due to data confidentiality constraints. Future research should integrate additional macroeconomic indicators, such as GDP growth or consumer confidence. It would also be valuable to test model robustness on non-luxury segments and emerging markets using newly acquired datasets. Exploring advanced time-series architectures, such as LSTM, GRU, or transformer models, for long-term RV forecasting represents another promising direction. Additionally, investigating second-order feature interactions (for instance, interactions between model and optional feature rate) could further refine depreciation insights.

Chapter 4

Contributions in Clinic Case Studies

Sports medicine and neurology are two distinct pillars of clinical practice that exemplify how evidence-based strategies and technological innovation are advancing patient care. Sports medicine focuses on acute musculoskeletal injuries and athletic performance, typified by the management of anterior cruciate ligament (ACL) ruptures, one of the most common serious knee injuries in athletes. Neurology, in contrast, tackles chronic diseases of the nervous system, such as Parkinson's disease (PD), a progressive neurodegenerative disorder. Despite their differences, both fields are evolving rapidly: clinicians are drawing on new research insights and integrating emerging technologies (including data analytics and artificial intelligence) to improve outcomes for patients with conditions as disparate as an ACL tear and Parkinson's disease.

In sports medicine, an ACL injury in a young athlete is a pivotal event that can sideline a career and predispose to long-term knee morbidity. Epidemiologically, ACL tears are highly prevalent; recent analyses estimate an incidence of approximately 68–70 cases per 100,000 person-years in the general population of the United States [229]. This translates to a substantial clinical burden, on the order of hundreds of thousands of ACL reconstructions are performed annually, reflecting the frequency of these injuries in high-demand sports. Beyond the immediate instability and pain of an acute tear, the long-term consequences are a major concern. Even with surgical reconstruction and modern rehabilitation, roughly 50% of individuals who suffer an ACL rupture develop radiographic knee osteoarthritis within 10–20 years after the injury [230]. The pathophysiology of this post-traumatic osteoarthritis involves altered knee kinematics (e.g. anterior tibial translation and rotational instability) and cartilage trauma at the time of injury, which accumulate to cause degeneration despite ligament repair. Thus, an ACL tear is not merely a transient ligamentous injury but often a life-altering event with chronic implications for joint health.

Clinical outcomes after ACL reconstruction underscore the need for continual improvements in practice. Meta-analytic evidence indicates that while approximately 80% of patients who undergo ACL reconstruction return to some form of sport, only about 65% successfully regain their pre-injury level of sports participation, and barely 55% resume competitive-level sport [231]. In other words, nearly half of athletes never fully return to their prior competitive capacity after an ACL tear, often due to persistent deficits in knee function, fear of re-injury, or residual instability. Furthermore, those who do return to high-level pivoting sports face a non-trivial risk of sustaining a second ACL injury in either knee. These sobering statistics have driven a shift toward more rigorous rehabilitation protocols and return-to-play decision-making in sports medicine. Contemporary clinical practice emphasizes criterion-based rehabilitation rather than solely time-based milestones: athletes must achieve objective strength and neuromuscular targets (for example, quadriceps strength at least 90% of the uninjured side, and passing sport-specific hop and agility tests) before being cleared for competition. Structured return-to-sport test batteries have been shown to stratify re-injury risk, with those failing to meet cutoffs at significantly higher risk of a repeat ACL tear compared to those who pass all criteria. As such, sports medicine teams now often extend rehabilitation to 9–12 months and incorporate neuromuscular training, proprioceptive exercises, and psychological readiness assessments to optimize safe return-to-play.

In parallel, the field of sports medicine is increasingly integrating technology and data-driven tools into both injury prevention and rehabilitation. Motion analysis systems and wearable sensor devices can quantitatively track knee biomechanics during jumping, landing, or cutting maneuvers, helping clinicians identify abnormal movement patterns that might predispose to injury. Likewise, machine learning models are being developed to assist in diagnosing ACL injuries on imaging and to predict outcomes after reconstruction. Recent reviews highlight the growing role of artificial intelligence (AI) in this domain, from automated MRI analysis that can detect ACL tears with high accuracy, to predictive algorithms that use patient and surgical data to forecast the likelihood of successful return to sport or the risk of osteoarthritis [232]. Although many of these AI-driven tools remain in experimental or early clinical stages, they represent a frontier in sports medicine. The vision is that, in the near future, a combination of personalized rehabilitation programs (tailored via sensor feedback) and AI decision-support systems could further improve functional recovery and reduce complications for athletes with ACL injuries. The ongoing challenge is to validate these technologies and incorporate them in a patient-centered manner, complementing the clinician's expertise without introducing bias or risk.

Turning to neurology, Parkinson's disease is a prime example of a complex condition where clinical practice is being reshaped by both scientific discoveries and technological innovations. PD is among the most prevalent neurodegenerative diseases, second only to Alzheimer's disease in global impact. As of the early 2020s, over 8 million people worldwide are living with Parkinson's disease, and due to aging populations this number is rising steadily. Epidemiological projections indicate that by 2050 the global PD prevalence will more than double (surpassing 25 million cases), making PD an escalating public health challenge [233]. Clinically, PD is characterized by its cardinal motor features, rest tremor, bradykinesia (slowed movement), rigidity, and postural instability, as well as a host of non-motor symptoms (autonomic dysfunction, sleep disorders, cognitive impairment). Standard therapy remains primarily symptomatic, centered on dopamine replacement (levodopa) and neuromodulation (deep brain stimulation in advanced cases). However, there is growing recognition of the heterogeneity in PD's manifestations and progression, which has significant implications for patient management and research into disease-modifying treatments.

A striking example of this heterogeneity is GBA-associated Parkinson's disease. Mutations in the GBA gene (which encodes the lysosomal enzyme glucocerebrosidase) are now understood to be the single most common genetic risk factor for Parkinson's disease, found in approximately 5–15% of PD patients across populations [234]. GBA mutations were initially known for causing Gaucher's disease (a rare lysosomal storage disorder), but heterozygous GBA variants in otherwise healthy individuals substantially increase the risk of developing PD and influence the disease phenotype. Clinically, PD patients who carry GBA mutations tend to have an earlier age of onset, a higher likelihood of cognitive impairment and dementia, and a more rapid disease progression compared to idiopathic PD cases without such mutations [234]. The discovery of this genotype-phenotype correlation has reinforced the concept that PD is not a single monolithic disease, but rather a syndrome with subtypes driven by different molecular pathways (in this case, lysosomal dysfunction and impaired alpha-synuclein degradation). This insight is shaping clinical practice and research: neurologists are increasingly aware of the utility of genetic testing in young-onset or atypical PD patients, and clinical trials are underway for therapies targeting the GBA pathway. For instance, small-molecule chaperones (like ambroxol) that enhance mutant glucocerebrosidase activity, or gene therapies to supplement GBA function, are being explored as potential disease-modifying strategies specifically for GBA-PD. While these investigational treatments are not yet in routine use, they represent a move towards precision medicine in neurology, tailoring interventions to the patient's genetic and molecular profile.

Beyond pharmacogenomics, technology-driven approaches are being adopted in the management of Parkinson's disease to address challenges in monitoring and individualized care. Because PD symptoms fluctuate throughout the day and progress insidiously, continuous and objective monitoring is valuable for optimizing therapy. Wearable sensors and digital applications have emerged as promising tools to capture real-world motor symptoms of PD, such as gait dynamics, tremor frequency, limb rigidity, or bradykinesia, in a way that periodic clinic exams cannot. Accelerometer-based systems, for example, can quantify gait speed, stride length, or tremor amplitude over days and weeks, alerting clinicians to changes that might warrant treatment adjustments. A recent systematic review confirmed the feasibility of using wearable sensors for early PD diagnosis and symptom tracking, noting that these devices can noninvasively monitor motor performance continuously and detect subtle impairments in early-stage disease [235]. However, the evidence also highlights current limitations: many studies had small sample sizes, varied protocols, and lacked standardization, so further validation is needed before such digital biomarkers can be fully integrated into clinical practice [235]. Still, the trajectory is clear. In tandem with wearables, researchers are leveraging machine learning on diverse data streams, from voice recordings and smartphone touchscreen interactions (which can reflect bradykinesia) to facial expression analysis and home-based keyboard tapping tasks, to develop algorithms that could screen for PD or predict symptom exacerbations. These innovations align with the broader trend of telemedicine and remote patient monitoring in neurology, which gained momentum out of necessity and convenience. For patients, this means care that is more personalized and proactive: adjustments to medication can be made based on objective home data, and interventions can be timed before a significant decline occurs.

In summary, the landscapes of both sports medicine and neurology are being transformed by a combination of rigorous research and technological progress. The ACL injury paradigm in sports medicine has shifted toward comprehensive rehabilitation and prevention strategies that acknowledge the long-term joint health risks; concurrently, data-driven tools from motion capture to AI are augmenting clinicians' ability to diagnose, treat, and prevent these injuries. Likewise, in Parkinson's disease, advances in molecular genetics (exemplified by the elucidation of GBA-associated PD) and the advent of digital health devices are ushering in an era of more nuanced and precise care. In both cases, the core aim remains the same: to restore and preserve patients' quality of life. By orchestrating multidisciplinary efforts: surgeons, physiotherapists, neurologists, geneticists, and data scientists, and by applying innovations responsibly, modern clinical practice continually improves its capacity to tackle the challenges of complex injuries and diseases. These parallel evolutions in two very different fields underscore a unifying principle in medicine today: leveraging science and technology to deliver personalized, high-quality care, whether on the playing field or in the neurology clinic.

4.1 Prediction of GBA1 Mutated Status in Parkinson's Disease Patients

Parkinson's Disease (PD), the most common type of parkinsonism, is the second most prevalent neurodegenerative disorder after Alzheimer's Disease [236]. Motor cardinal manifestations of PD include bradykinesia, rest tremor, rigidity and postural and gait impairment [237]. In addition to motor features, most patients with PD also present with several non-motor symptoms [237]. The PD clinical course and the response to treatment vary just like its etiology, which is multifactorial and complex [238, 239] Pathogenic variants associated with a high risk of PD are found in 15% of all cases [239, 240, 241, 242, 243]. Heterozygous variants in the -glucocerebrosidase gene (GBA1) represents the greatest known genetic risk factor for PD [239, 243] and account for 5–30% of cases depending on the population and age at onset of the disease [239]. To date, more than 300 GBA1 variants have been associated with PD, with an overall odds ratio (OR) for developing the disease of approximately 3.5–6 [238]. The clinical features of GBA1-related PD (GBA1-PD) are similar to those of idiopathic PD. However, the average age of onset is often slightly earlier, while clinical progression is generally faster with frequent early motor fluctuations (MF) and levodopa-induced dyskinesias (LID) and shorter survival [244, 245] Non-motor symptoms, namely hyposmia, constipation, orthostatic hypotension, urinary dysautonomia, depression, anxiety, sleep disturbances, and cognitive impairment are more frequent than in non-mutated PD (NM-PD). GBA1 variants can be stratified as mild or severe, based on the type of Gaucher Disease (GD) they are associated with when present in the homozygous state [246, 247]. Mild variants are those that cause nonneuropathic GD (type I), whereas severe mutations are those that cause acute neuropathic GD (type II), and chronic neuropathic GD (type III).[246, 248] Finally, risk variants are linked to an increased risk of PD, but are not causative of GD.13 The clinical phenotype and disease severity of GBA1-PD is influenced by the severity of GBA1 mutation [239]. Indeed, PD patients carrying severe mutations (ie, L444P, IVS2 + 1G>A, c.84dupG, V394L, D409H, RecTL, RecNCil) have an earlier age at disease onset and greater risk of dementia, impulsive–compulsive disorders (ICD), and delusions compared to GBA1-PD carrying mild mutations (ie, N370S) which may resemble

NM-PD [239]. However, this scenario is not completely understood. In fact, PD patients harboring GBA1 risk variants such as E326K have been reported to carry a higher risk of cognitive impairment [239]. The possible phenotypic similarities between GBA1-PD and NM-PD patients undermine the reliable clinical suspicion of GBA1 mutation carriers based solely on individual clinical features [249] Conventional statistical models, which assume linear and additive relationships, often underperform when faced with the high-dimensional, non-linear interactions that characterize PD phenotypes and may therefore miss subtle multivariate signatures of GBA1 status [250] With the increase of computational power and data availability,[251] AI became useful to assist humans in the analysis of large and complex data sets [252]. ML is a large group of methods that has been applied to a broad range of areas, including genetics and genomics [252]. Its objective is to develop algorithms that can learn and adapt without following explicit instructions.18 In addition, ML algorithms excel at capturing complex feature interactions, handling mixed data types, and providing post-hoc explainability through tools such as SHAP.[49, 253] Previous studies have employed machine learning techniques to develop predictive models for PD,[250, 254] as well as multiple hybrid machine learning systems (HMLS) aimed at predicting the mutational status of the LRRK2 and GBA1 genes [255] The aim of this study was to assess whether AI could predict GBA1-mutated genotype in a cohort of deeply phenotyped PD patients. Particularly, the objective was to identify a ML model capable of accurately providing a pre-test estimate of GBA1- mutated status, relying on the clinical and demographic variables with the highest predictive value. Estimating the pre-test probability of carrying pathogenic GBA1 variants may support clinicians in the future in prioritizing patients for genetic testing, thereby enhancing overall cost-effectiveness, especially in resource-limited settings (**C-RQ1**).

4.1.1 Methods

This single center observational study included two consecutive cohorts of GBA1-PD and NM-PD patients recruited at the Movement Disorder Center of AUSL-IRCCS of Reggio Emilia, Italy. The genetic profile was obtained by testing the patients for 11 pathogenic or likely pathogenic LRRK2 variants and GBA1 sequences. If negative, a next-generation sequencing panel targeting 68 genes involved in PD was performed within the ROPAD study. [256] Each consecutive GBA1-PD patient has been matched with a 1:1 pairing method with consecutive NM-PD patients from a pull of 400 consecutive NM-PD subjects. In particular, the variables considered for the 1:1 match were sex, age at evaluation (with a tolerance of 2 years), age at disease onset (with a tolerance of 2 years), the comorbidity burden of the patients quantified through the Charlson Comorbidity Index (CCI) (tolerance of 2 points) and Hoehn Yahr scale (HY). During the pairing method, no other clinical or instrumental data has been considered except the one already mentioned. The inclusion criteria were a diagnosis of PD according to the United Kingdom Parkinson's Disease Society Brain Bank Diagnostic Criteria for Parkinson's Disease [257] and an age of at least 18 years. The exclusion criteria were a diagnosis of atypical or secondary parkinsonism, inability to perform the study protocol, and withdrawal of consent. Written informed consent was obtained from each participant, according to the Declaration of Helsinki.[258] This study was approved by the Ethics Committee of the Area Vasta Emilia Nord (536/2021; 2022/0139218). Demographic and PD characteristics (including sex, age at evaluation, age at disease onset, disease duration, family history of PD) were collected, along with results of genetic testing and motor phenotype at disease onset and at the time of clinical evaluation. In addition, the presence and time of onset (relative to diagnosis) of several non-motor symptoms were recorded, comprising depression, anxiety, apathy, mild cognitive impairment,[259] dementia,[260] hallucinations/delusions, restless leg syndrome, hyposmia, provisional diagnosis of REM behavior disorder [RBD],[261] urinary incontinence, constipation, impulse control disorder-ICD,[262] sialorrhea, dysphagia. Positive family history was de ned as having at least one rst-degree relative diagnosed with PD. The anamnestic motor phenotype at disease onset was classied according to the predominant initial motor presentation (akinetic-rigid, tremor-dominant, axial). The motor phenotype at the time of clinical evaluation was determined using formulas based on MDS-UPDRS scores classifying patients as tremor- dominant, postural instability and gait difficulty (PIGD), or indetermined as proposed by Stebbins et al [263]. Regarding non- motor symptoms, a binary classification (present/absent) was applied. Moreover, the presence and time of onset relative to diagnosis of MF, LID, falls and freezing of gait (FOG) were also considered. Finally, each patient included in the study was evaluated by the following clinical scales: the four parts of the Movement Disorder Society revision of the Unied Parkinson's Disease Rating Scale (MDS-UPDRS) [264] and HY [265] during chronic treatment in the ON-state. The following subscores were also extrapolated from the MDS-UPDRS: tremor, bradykinesia, rigidity, PIGD, gait, dyskinesia, uctuations, hallucinations and psychosis subscores (the items of each subscore are reported in the supplementary materials). Cognitive function was assessed through the Montreal Cognitive Assessment (MoCA) [266]. The total amount of the dopaminergic treatment was converted into the levodopa equivalent daily dose (LEDD) using the updated conversion formulae.

The primary objective of the study was to assess whether AI could predict the GBA1-mutated genotype considering the different impact of clinical features in a homogenous cohort of GBA1-PD and NM-PD extensively phenotyped. In this study, we examined a data set comprising 116 patients, each characterized by 125 features, including target. The missing values were imputed using Multivariate Imputation by Chained Equations (MICE) [267]. Based on genetic analysis, PD patients were dichotomized in two cohorts: GBA1-mutated (GBA1-PD) or NM-PD. GBA1-PD patients were further strati ed into three subgroups: severe, mild, and risk GBA1-PD, depending on whether they carried a severe, mild, or risk variant, [246, 247, 248] respectively, to perform a comparative sub-analysis. For feature selection, we applied the Chi-square test[260] and the Mann–Whitney U test [261] to identify, respectively, nominal and continuous features that significantly differed between the GBA1-PD and NM-PD groups (statistical significance was set at p-value <0.05). These statistical approaches helped in narrowing down the most relevant features for our subsequent modeling. For the predictive modeling, we explored a variety of linear and non-linear algorithms: Support Vector Machines (SVM), [72] Logistic Regression,[268] XGBoost [100] and Random Forest [69] and Stochastic Gradient Descent Classifier [269]. Before model training, we applied a pair-wise correlation filter, and no pair with $|r| > 0.9$ was discovered. Moreover, tree-based algorithms such as XGBoost further mitigate residual multicollinearity by selecting splits on the most informative variable within each correlated cluster. Concerning training, each model was rigorously evaluated through the leave-one-out (LOO) cross-validation [270] to determine its effectiveness in accurately categorizing patients. LOO therefore trains on 115 cases and tests the single hold-out subject at each iteration, maximizing the statistical efficiency of the available data. Exact binomial power calculations (= 0.05) show that with n = 116 predictions the design retains 99% power to detect an accuracy, for example of 0.70, versus the 0.50 expected by chance. The final stage of our analysis focused on feature importance, selecting the best combination of features in order to maximize model accuracy. By examining the contribution of individual features in the predictive models, we were able to identify the most significant factors in influencing genetic categorization. Finally, SHAP was applied to understand the positive or negative impact of each feature in determining the mutated or non-mutated genotype [49]. Statistical analyses were performed using SPSS Statistics for Windows version 20.0 (IBM, Armonk, NY, USA), Python libraries Scikit-Learn version 1.4 (Pedregosa et al, 2011) and SciPy version 1.12.0 (Virtanen et al, 2020).

4.1.2 Results and Discussion

pdflscape booktabs graphicx

Table 4.1: Demographic and clinical variables of the cohorts. Values are expressed as mean [SD]; median {range}.

Variable	Total PD cohort (n=116)	GBA1-PD (n=58)	NM-PD (n=58)	P-value	Effect size
Age at PD onset (years)	56.47 [9.64]; 54 {36–75}	56.62 [9.65]; 53 {36–75}	56.31 [9.70]; 53 {36–74}	0.82	0.02
Age at PD diagnosis (years)	56.90 [9.56]; 56 {36–75}	57.10 [9.57]; 56 {36–75}	56.66 [9.93]; 54.5 {36–74}	0.86	0.02
Age at evaluation (years)	64.45 [9.38]; 63 {43–83}	64.44 [9.48]; 63 {43–83}	64.47 [9.73]; 63 {46–82}	0.99	0.01
CCI	7.07 [2.20]; 7.00 {2–13}	6.88 [1.45]; 7.00 {2–10}	7.33 [2.45]; 7.00 {3–12}	0.40	–0.18
LEDD	2210.19 [1011.96]; 2125 {1–3000}	2157.15 [711.25]; 2140 {1–2200}	2263.12 [1278.20]; 1978 {1–3200}	0.41	–0.13
H&Y	2.41 [0.70]; 2.5 {1–5}	2.47 [0.70]; 2.5 {1–5}	2.35 [0.71]; 2.5 {1–4}	0.27	–0.43
MoCA	22.56 [5.86]; 24 {3–30}	21.75 [5.13]; 23 {5–30}	23.22 [6.32]; 25 {12–30}	0.69	0.15
Disease duration (years)	7.78 [4.87]; 6.5 {1–24}	7.64 [4.97]; 6.5 {1–23}	7.94 [4.81]; 7.5 {2–24}	0.85	0.29
Family history of PD (Yes/No)	42 (36%) / 74 (64%)	28 (48%) / 30 (52%)	14 (24%) / 44 (76%)	0.01	0.25
Early onset PD (Yes/No)	14 (12%) / 102 (88%)	8 (14%) / 50 (86%)	6 (10%) / 52 (90%)	0.57	0.06
Advanced PD therapies (Yes/No)	12 (10%) / 108 (90%)	9 (15%) / 79 (85%)	3 (5%) / 55 (95%)	0.07	0.16
MDS-UPDRS I (ON state)	12.12 [7.44]; 11 {0–34}	13.31 [7.69]; 11.5 {0–29}	10.84 [7.03]; 10 {0–34}	0.06	0.21
MDS-UPDRS II	14.64 [7.98]; 13 {0–40}	13.02 [6.64]; 13 {0–38}	16.25 [7.09]; 15 {0–40}	0.15	0.15
MDS-UPDRS III	32.44 [14.43]; 31 {4–70}	28.13 [9.42]; 29.5 {9–53}	36.75 [16.41]; 32.5 {10–70}	0.02	0.27
MDS-UPDRS IV	4.44 [4.01]; 3 {0–24}	4.18 [3.91]; 3 {0–24}	4.71 [4.13]; 3 {0–24}	0.97	0.01
Tremor subscore	3.55 [3.52]; 1.5 {0–15}	3.41 [3.51]; 1.5 {0–12}	3.70 [3.55]; 1.5 {0–15}	0.93	0.03
Bradykinesia subscore	14.53 [7.11]; 15 {1–30}	14.19 [6.88]; 15 {1–28}	14.88 [7.31]; 15 {0–30}	0.42	0.08
Rigidity subscore	5.29 [4.29]; 5 {0–16}	5.29 [4.29]; 4.5 {0–16}	5.29 [4.29]; 5 {0–16}	0.53	0.17
PIGD subscore	1.62 [1.91]; 1.5 {0–8}	1.71 [1.93]; 1.5 {0–8}	1.51 [1.89]; 1.5 {0–8}	0.05	0.21

One hundred sixteen PD patients were included in the study: 58 GBA1-PD (males: 32; age: 64.64 years; disease duration: 7.6 years; MDS-UPDRS III: 35.98; MoCa: 21.75; CCI: 2.67) and 58 NM-PD (males: 32, age: 64.47 years; disease duration: 7.95 years; MDS-UPDRS III: 28.95; MoCa: 23.22; CCI: 2.48). 32 GBA1-PD carried mild or risk variants, while the remaining 26 GBA1-PD patients carried

Table 4.2: MDS-UPDRS subitems that significantly differed between GBA1-PD and NM-PD. Values are expressed as mean [SD]; median {range}.

Subitem (MDS-UPDRS, ON-med)	NM-PD (n=58)	GBA1-PD (n=58)	P-value	Effect size
MDS-UPDRS 1.1: Cognitive impairment	0.77 [1.061]; 1 {0–3}	1.24 [1.138]; 1 {0–4}	0.001	0.35
MDS-UPDRS 3.3a: Rigidity – Neck	0.81 [0.805]; 1 {0–3}	1.12 [0.774]; 1 {0–3}	0.04	0.23
MDS-UPDRS 3.3a,b: Rigidity – Right/Left Upper Extremity	0.80 [0.750]; 1 {0–3}	1.26 [0.805]; 1 {0–3}	0.02	0.25
MDS-UPDRS 3.4a,b: Finger tapping – Right/Left hand	1.580 [0.3]; 2 {0–3}	1.870 [0.32]; 2 {0–3}	0.04	0.26
MDS-UPDRS 3.5a,b: Hand movements – Right/Left hand	1.22 [0.832]; 1 {0–3}	1.30 [0.754]; 1 {0–3}	0.15	0.12
MDS-UPDRS 3.6a,b: Pronation-supination movements – Right/Left hand	1.45 [0.762]; 1 {0–3}	1.07 [0.741]; 1 {0–3}	0.24	0.14
MDS-UPDRS 3.8a,b: Leg agility – Right/Left leg	1.33 [0.741]; 1 {0–3}	1.43 [0.744]; 1 {0–3}	0.03	0.28
MDS-UPDRS 3.14: Global spontaneity of movement	1.64 [0.672]; 1 {0–3}	1.69 [0.928]; 1 {0–3}	0.03	0.25

severe variants. Table 4.1 shows demographic and clinical variables in the whole cohort, in GBA1-PD and NM-PD subgroups. GBA1-PD and NM-PD subgroups were matched for sex, age at disease onset and age at evaluation, H&Y staging scale and CCI as confirmed by P values reported in Table 4.1. Moreover, these two subgroups did not present significant difference with regard to LEDD and motor phenotypes. Conversely, GBA1-PD patients had more frequently a positive family history for PD. Concerning motor symptoms, MDS-UPDRS part III score was significantly (p= 0.02) higher in GBA1-PD (35.98) com- pared to NM-PD (28.95). Moreover, GBA1-PD showed higher rigidity (p= 0.003), bradykinesia (p= 0.03) and PIGD (p= 0.05) subscores than NM-PD. Table 4.2 shows in details MDS- UPDRS sub-items that significantly differ between GBA1-PD and NM-PD. Twenty-six GBA1-PD patients carried a severe variant (severe GBA1-PD), while 32 patients were carrier of mild or risk variants- ants (mild/risk GBA1-PD). Severe GBA1-PD had an earlier age at PD onset and diagnosis compared to mild/risk GBA1-PD. Table A.1 (supplementary materials) shows details about age at onset and diagnosis in severe GBA1-PD and mild/risk GBA1-PD. Compared to NM-PD, severe GBA1-PD patients presented more severe cognitive impairment according to MDS- UPDRS 1.1. Moreover, concerning motor symptoms, severe GBA1-PD had more severe episodes of FOG according to MDS-UPDRS 3.11. Severe GBA1-PD had a higher total MDS-UPDRS part III score, MDS-UPDRS 3.14, rigidity and PIGD subscores, meaning that they had a global more severe motor phenotype, particularly concerning rigidity and axial symptoms compared to NM-PD. In addition, severe GBA1-PD had significantly higher prevalence of dysphagia. Finally, severe GBA1-PD presented more frequently a positive family history of PD compared to NM-PD. Table A.2 (supplementary materials) shows details about demographic and clinical variables that significantly differed between severe GBA1-PD and NM-PD. In the context of differentiating GBA1-PD from NM-PD, AI methodologies were applied. The classification task leverages features that have been identified as statistically significant, as reported in Table 4.1 and 4.2. To enhance the reliability of our model's decisions, we utilized features that have been identified as significantly distinct between the GBA1-PD and NM-PD groups. Models were tested using a Leave-One-Out approach, so all the patients were tested once. Accuracy is a metric that quantifies the frequency with which a ML model correctly predicts outcomes. It is de ned as the ratio of the number of correct predictions to the total number of predictions made. According to accuracy metric, XGBoost was identified as the most effective machine learning model for this supervised classification task (Table 4.3). The model initially reached an overall accuracy of 71%. Despite the cross-validation safeguards, the modest sample size inevitably raises the risk of overfitting. However, this risk is partially offset by our choice of XGBoost, whose boosted decision-tree ensemble averages the predictions of many weak learners and applies shrinkage as well as column- and row-subsampling, thereby providing an additional guard against variance-driven overfitting. XGBoost model is able to quantify the predictive value of each feature, according to Mean Decrease in Impurity (MDI), utilizing those with the highest predictive value for classification as GBA1- or NM-PD. We selected the combination of these variables that allow to achieve the highest prediction accuracy, reaching 73% accuracy (95% CI 64.6–80.5%), as highlighted in 4.3. Notably, the variables selected by the model exhibited less than 4% missing data and no pair-wise correlation exceeds $|r|= 0.5$. Figure 4.1 illustrates the features importance according to MDI. Cognitive impairment (based on MDS-UPDRS 1.1), family history of PD, MDS-UPDRS 3.8 (mean value between a and b) and rigidity subscore were the most predictive variables in the classification, each exhibiting a relative importance from 0.10 to 0.35. In order to understand the impact of each feature in determining the GBA1-mutated or non-mutated status,[271] we decided to use SHapley Additive exPlanations (SHAP). Figure 4.2 shows SHAP summary plot, which highlights the positive or negative impact of each predictive variable. In SHAP summary table each point refers to a specific patient enrolled, while the color represents the feature's value. Particularly, higher value is represented by the red color, while lower value by the blue one. Finally, a

Table 4.3: Mean leave-one-out accuracy across models.

Model	Accuracy (%)
XGBoost	**71**
Random Forest	64
Logistic Regression	66
Support-Vector Machine	60

Table 4.4: High-confidence subset (absolute SHAP ≥ 0.8) classified by XGBoost.

	Accuracy (%)	n	Cohort proportion (%)
XGBoost	**94**	16	14

positive SHAP value is associated with an increased probability of being GBA1-mutated. Conversely, a negative SHAP value is associated with an increased probability of being non-mutated. Thus, according to this model, the parameters that made more likely GBA1-PD mutated genotype are listed below:

- Positive family history of PD
- High value of MDS-UPDRS 1.1 (cognitive impairment), MDS.UPDRS 3.8 a and b (bilateral legs agility) and rigidity subscore

In addition, SHAP allows for extracting insights on the model's behavior to quantify the confidence of the model on the prediction. Considering only the patients with a SHAP value over 0.8, which means, in cases where the model has a confidence level greater than 80%, we achieved an accuracy of 94% (Table 4). The 80% SHAP threshold was pre-specified in the study protocol as a pragmatic cut-off for"high-confidence."

GBA1 variants are considered as the major genetic risk factor for PD9 and they are found in about 2–30% of PD patients [272]. However, PD patients carrying GBA1 mutations are not easily recognizable because they do not display exclusive features that would clearly distinguish them from NM-PD [249]. Anyway, large population studies highlight common traits in GBA1-PD [273]. According to literature, GBA1-PD patients present an overall earlier age at onset and more pronounced cognitive impairment compared to NM-PD [249]. Moreover, in GBA1-PD the rigid-akinetic phenotype appears to be more common and motor symptoms'

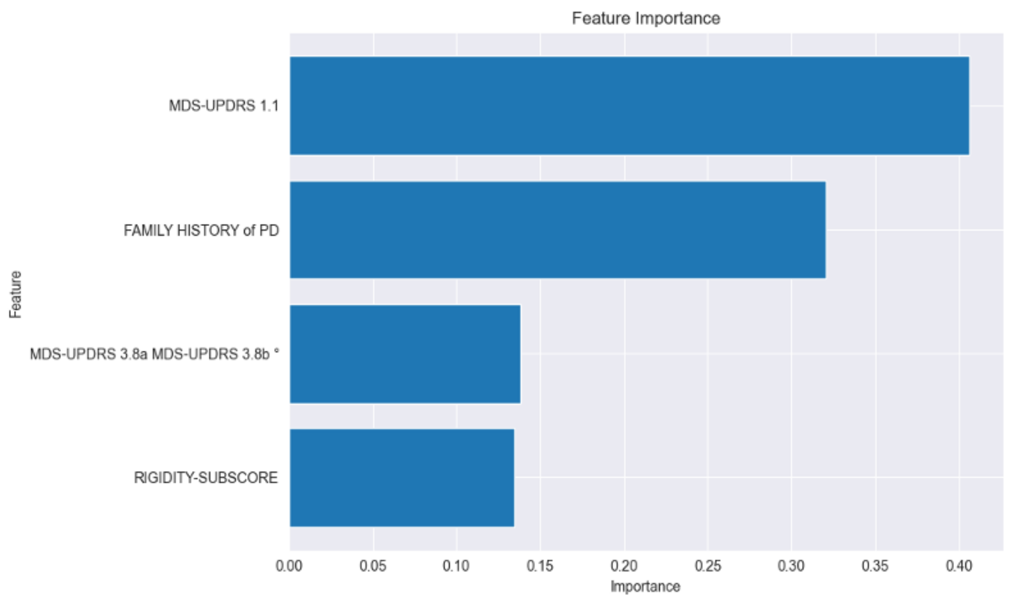

Figure 4.1: Global feature importance ranked by mean decrease in impurity.

progression could be slightly faster [249]. In recent years, ML has been increasingly adopted to develop computer algorithms assisting humans in the analysis of large, complex datasets [252]. In this scenario, the aim of our study was to assess whether AI could help the clinician to suspect a mutated GBA1 genotype, elaborating a predictive model based on demographic and clinical variables from a cohort of 116 PD patients. The 58 GBA1-PD and 58 NM-PD included in the study were matched for sex, age at disease onset and age at evaluation, H&Y scale and CCI. This allowed us to avoid possible con- founding demographic and clinical factors. GBA1 patients were found to have more frequently a positive family history of PD compared to NM-PD. In literature a positive familial history can be identified in 21.5–31% of GBA1-PD, suggesting that more than two-thirds of PD carriers of GBA1 mutations are sporadic [273, 274, 275]. Moreover, according to our results, even if motor phenotype did not significantly differ between GBA1-PD and NM-PD, GBA1-PD showed more severe motor symptoms, particularly in terms of higher rigidity, bradykinesia and PIGD sub- scores. This is not surprising as it has been described that GBA1 mutation carrier display a more aggressive disease course [273, 274, 275, 276, 277] with faster progression of motor disability [251, 273, 278, 279]. Finally, concerning non motor symptoms,

Figure 4.2: SHAP summary plot for the four retained variables.

in this study GBA1-PD showed more severe cognitive impairment compared to NM-PD, according to MDS-UPDRS 1.1 subitem. This result is consistent with literature reporting that GBA1 mutated status is an independent risk factor for cognitive decline in PD and GBA1 genotype severity is linked to the rate of progression of cognitive impairment [274, 280, 281, 282]. Particularly, several previous studies highlighted that GBA1 mutations greatly increase the risk of incident dementia [249, 257, 272, 273, 274, 276, 283, 284, 285] and the risk of cognitive impairment in GBA1 mutation carriers is 2.4 to 3-fold higher than in non-carriers [273, 274, 286]. Furthermore, a particular decline in working memory, visuospatial, and executive function has been previously reported to occur in GBA1-PD patients.11 The prominent cognitive impairment of GBA1-PD is probably due to the severe cortical Lewy pathology documented in these patients. Of note, the above-mentioned motor and non-motor clinical differences between GBA1-PD and NM-PD cohorts are confirmed when we compared NM-PD patients and carriers of GBA1 severe variants. Finally, considering only GBA1-PD subgroups, our results confirmed that patients carrying severe mutation presented an earlier age at onset compared to patients with mild or risk variants [249]. Previous studies highlight that PD onset occurred at a younger age among carriers of severe versus mild GBA1 mutations, ranging from 2 to 13 years [272, 273, 274, 287]. To develop our predictive model, we compared several machine learning models to identify the one with the highest accuracy and XGBoost emerged as the most accurate classifier in our head-to-head comparison. XGBoost is an

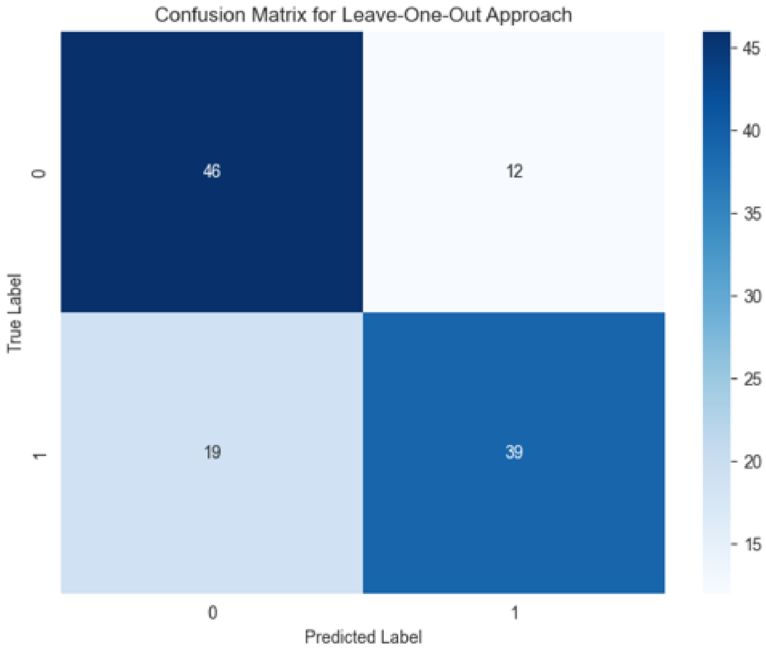

Figure 4.3: Confusion matrix for the leave-one-out evaluation (TN = 46, FP = 12, FN = 19, TP = 39).

open-source algorithm, with its implementation and source code accessible through the official website: xgboost.ai. Its superiority is theoretically well founded: (i) the boosted decision-tree ensemble captures high- order, non-linear interactions among heterogeneous clinical fea- tures; (ii) it accommodates both monotonic and non-monotonic effects without prior transformation; and (iii) it embeds shrinkage, plus column- and row-subsampling, that collectively curb over- tting in modest samples. By contrast, logistic regression and linear-kernel SVMs impose linearity, while Random-Forest and Stochastic-Gradient-Descent classifiers were less effective once highly correlated variables had been pruned.Using artificial intelligence, we developed a model based on four specific features with significant positive or negative impact on the prediction of GBA1-mutated genotype, irrespective of the speci c variant. These four characteristics were firstly selected among the 13 total features that resulted statistically different between GBA1- PD and NM-PD subgroups during the features selection phase. They included family history of PD, rigidity subscore, MDS- UPDRS 1.1 and MDS-UPDRS 3.8 (mean value between a and b). Considering that our GBA1-PD cohort's features are consistent with the literature, we can assume that these results are also applicable to other PD cohorts. Using ML, we were able to identify not only the most impactful features, but also their effects. Particularly, the most significant factors in making the GBA1 mutated genotype more likely respect to NM-PD were positive family history of PD, presence of cognitive impairment (high value MDS- UPDRS 1.1) and marked rigidity (high rigidity subscore). A posi- tive family history, together with higher cognitive and motor bur- den is in line with previously reported findings in GBA1-PD, as outlined above. This model could help clinicians suspect GBA1- PD patients, by calculating the pre-test estimate of GBA1- mutated status, considering the different impact of these pre- dictive features. Indeed, it has the potential to assist clinicians in certain clinical settings in determining which patients should be prioritized for genetic testing. The strength of the ML model lies in its capacity to simultaneously integrate multiple clinical factors within a multivariate framework, including variables that may not be individually emphasized by clinicians (eg, bilateral leg agility subitem), and to assign specific quantitative weight to these factors in the prediction. Concerning limitations of this study, the low sample size may have in uenced negatively the training phase of the model and consequently its accuracy. Although the accuracy resulted in 73%, though SHAP was possible to identify a subset of patient where the model achieved 94%. Particularly, we were able to reach an accu- racy of 94% in patients where the model confidence was above 80%, corresponding to 14% of the dataset. A further limitation of the study, stemming from the limited sample size, is the potential risk of over tting. However, this risk is partially offset by the adoption of XGBoost model, as out- lined above. Finally, as this is a single-center study conducted in Italy, the cohort predominantly consists of individuals of Italian ancestry. Consequently, the algorithm needs to be vali- dated in ethnically diverse populations. In this scenario, future multicenter studies evaluating the predictive accuracy of the proposed model in an independent cohort of patients with PD will be crucial to enhance the model's performance. Our results confirm that the application of AI could represent a promising approach in predicting the GBA1 status in PD patients, underlying the potential role of AI in enhancing targeted genetic testing and personalized medicine in PD. Additionally, future work could explore the integration of other AI methodologies, the inclusion of more comprehensive clinical, instrumental, and genetic data, and the potential for real- world application in clinical settings. The ndings of this study not only contribute to the understanding of GBA1-PD but also highlight the importance of interdisciplinary approaches in tack- ling complex neurological issues. Through addressing these limi- tations and focusing on future advancements, the research community can move closer to achieving more accurate, ef - cient, and personalized care for patients with PD.

4.1.3 Contributions

This work provides proof of concept that a lightweight, explainable classifier can enrich for *GBA1* mutation carriers among routinely assessed PD patients, potentially streamlining genetic testing and accelerating enrolment into genotype-stratified trials. Future studies should validate and extend the model in geographically and ethnically diverse cohorts and assess real-world clinical impact through prospective implementation. By coupling standard machine-learning methodology with transparent feature attribution, the present study advances the goal of precision medicine in PD.

4.2 Biomechanical Phenotypes and Real-time Classification

A critical challenge due to the complex interplay of intrinsic and extrinsic risk factors is the prevention of anterior cruciate ligament (ACL) injuries [50]. Indeed, the risk of ACL trauma is affected by poor neuromuscular control and joint biomechanics [51]. As a consequence of this, most ACL injuries occur without direct knee contact in professional football, but nearly half occur via indirect contact mechanisms [288, 289]. Despite improved knowledge on ACL injuries and injury prevention, the rate of injuries in professional football is not declining [52].

A quick and cost-effective method to identify athletes at high risk could be 2D video-analysis [290]. However, to support orthopedic surgeons and sports physicians, a huge amount of information had to be analysed making the evaluation difficult using conventional methods.

AI techniques, including deep learning (DL) and machine learning (ML), have been used to predict ACL injuries [291, 292], diagnose them through automated video analysis, and identify movement patterns through cluster analysis [293, 294, 295, 296, 297]. Regarding ACL injuries prediction, Tamimi et al. [290] developed a model, using ML, based on morphological features of the knee from MRI scans of adult athletes aged 18 to 40 years. They achieved over 90% accuracy in predicting primary ACL injuries.

In terms of diagnosis through video analysis, Schulc et al. [298] employed AI algorithms to analyze publicly available match recordings, such as YouTube videos involving professional athletes. They reconstructed three-dimensional motion from single-camera views to identify biomechanical patterns associated with ACL injuries. Their AI model achieved an area under the receiver operating characteristic curve (AUC-ROC) of 0.88, demonstrating good diagnostic performance. Moreover, when orthopedic surgeons were provided with the AI-generated 3D poses, their diagnostic accuracy improved by 8%, indicating the potential of such tools to enhance clinical assessments.

Additionally, movement patterns have been identified through cluster analysis of athletes, including youth, collegiate, and professional populations, performing dynamic tasks like cutting maneuvers and jumps, using technologies such as markerless motion capture, high frequency cameras and force plates.

In particular, Sigurðsson and Briem [294] analyzed knee valgus moment patterns in a change of direction, in 213 youth soccer and team handball athletes using 3D motion capture and a force plate. They identified six distinct patterns associated with higher ACL injury risk, particularly among adolescent girls. Similarly, DiCesare et al. [293] and Sigurðsson et al. [292] investigated coordination strategies in 780 primarily female adolescent athletes and 69 collegiate female soccer players, respectively, using 3D motion capture and force plates. They identified distinct coordination strategies during tasks like the drop vertical jump and found that targeted training programs can effectively modify movement patterns associated with ACL injury risk. Additionally, Bird et al. [296] applied unsupervised clustering techniques on data from 668 Marine officer candidates using markerless motion capture and force plates during countermovement jumps. They identified three distinct movement strategy clusters, with the high-risk cluster showing a significantly higher proportion of musculoskeletal injuries.

These applications have shown promise in improving diagnostic accuracy and providing deeper insights into injury mechanisms [298, 299], suggesting that AI techniques can enhance clinical assessments and potentially outperform traditional methods in certain aspects.

Despite the recent scientific progress in the application of AI on biomechanical field, the researchers conducted so far on 2D video-analysis were mainly influenced by a limited sample size [18]. Multiple papers demonstrated that increasing sample size, concerning number of participant and variables acquired, generally improves model accuracy and stability [300, 301]. For this reason, the most reliable applications of ML/DL have come from large-scale studies. Yet the literature includes relatively few investigations with more than 1,000 participants. Access to a cohort exceeding this threshold therefore addresses a clear gap and enables robust training, validation, and subgroup analyses with improved statistical power and external validity.

Building on this evidence, the present work leverages a high-quality dataset providing force plate and high-speed cameras data across three 90° COD trials per limb to deliver an end-to-end pipeline (**S-RQ1**):

1. **Stage 1 — Unsupervised phenotyping**: 45 demographic, force, and kinematic variables from 1,009 footballers are embedded with t-SNE and clustered via agglomerative clustering (Euclidean distance, Ward linkage), yielding four interpretable force–control characteristics, defined as phenotypes.

2. **Stage 2 — Supervised phenotyping**: A Random-Forest classifier reproduces those phenotypes from only 12 high-importance, easy-to-acquire variables, achieving a macro-averaged $F_1 = 0.85$ while the full 45-variable model reaches $F_1 = 0.92$.

3. **Stage 3 — Real-time deployment**: The classifier is embedded in a General Data Protection Regulation (GDPR) compliant application that returns an athlete's phenotype and visualises personalised feedback for clinicians and players.

4.2.1 Methods

The current study is part of a larger study named "CUT THE ACL" [302]. In brief, one-thousand-and-two competitive football (soccer) players (age 16.3 ± 2.8 years, 264 females) were prospectively enrolled. Each player performed three preplanned 90° COD tasks per limb. The 2D evaluation was performed through objective measures (collected through three high-speed cameras) of frontal and sagittal plane joint kinematics at the cut initial foot contact (IC) and maximum knee flexion angle [293]. In addition, scores based on the frontal and sagittal plane joint kinematics, were computed at the maximum knee flexion (MKF) angle frame. Instead, forces were collected through a force platform collecting the first, second and third peak recorded. In addition, the forces were scaled in body weight. A total of 53variables were present in the original database, categorized as follows: 9 demographic variables, 6 score variables, 15 IC variables, 11 force variables, and 12 MKF variables. Demographic and sub-score variables were removed to focus on biomechanical differences between players. In addition, the type of IC contact (heel/flat/toe) and brake strategy, considering they are categorical in contraposition with the rest of the features which are continuous, were removed turning the final value in 48. The appendix Tables A3,A4,A5,A6 described the variables which were considered. Next, for each player, the mean of the six trials were computed for each variable. Moreover, a correlation matrix was built to check multilinearity condition. In the couples of variables with correlation coefficient > 0.9, the one with the larger variance was kept, while the other one was discarded. Only a couple of force variables were highly correlated with each other, therefore only the first peak of medial ground reaction force (mGRF) scaled by body weight was removed turning the amount of variables analyzed in 45. Finally in order to favorite the comparison among all the variables, each variable was divided in terciles and coded to 0, 1, 2.

Stage 1 — Unsupervised Phenotyping

To project the 45 biomechanical characteristics of players onto a 2-dimensional space while preserving relevant information, we adopted t-Distributed Stochastic Neighbor Embedding (t-SNE). t-SNE operates by converting the Euclidean distances between data points in the high-dimensional space into conditional

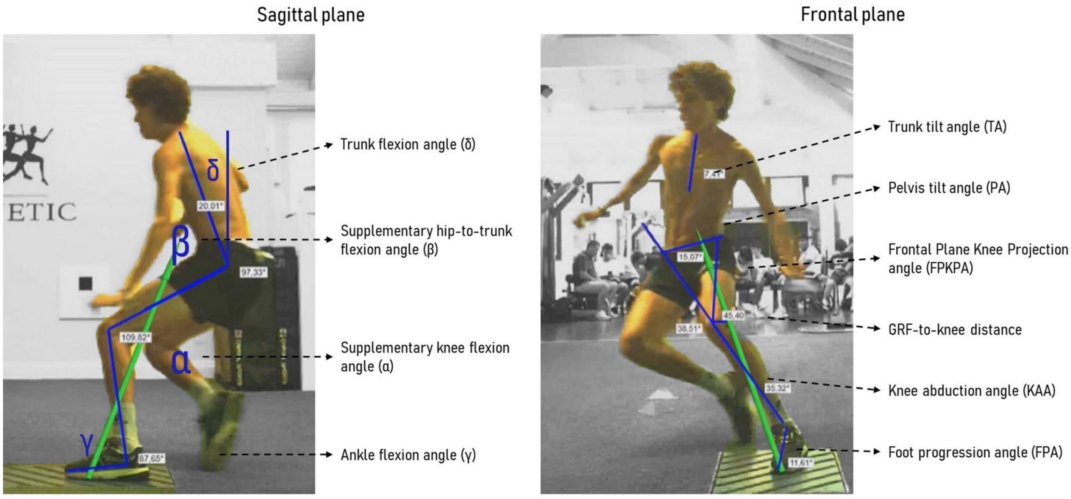

Figure 4.4: Instrumented “Green Room” corridor: three 240 Hz cameras capture 2-D kinematics while a force plate records ground-reaction forces during the 90° cut.

probabilities that represent similarities using a Gaussian kernel. This choice of distribution in the low-dimensional space addresses the "crowding problem" by allowing for moderate distances to be modeled with higher probabilities, effectively spreading out the players [104, 303]. This divergence is minimized using gradient descent, adjusting the positions of the players in the low-dimensional space to best reflect the similarity structure of the high-dimensional data. We applied t-SNE using Python (3.12, Python Software Foundation) and scikit-learn (Version 1.3.1, Pedregosa et al., 2011) a Python library. Next, the clustering analysis was adopted to group the players according to their biomechanical characteristics projection after DR. Due to the unsupervised nature of the clustering process and to the fact that the players were all free from musculoskeletal injuries at the time of the data collection, it can be reasonably assumed that each group presents specific observable traits a players movement biomechanics can be described with in a 90° COD. For this reason, in analogy with genetics, each cluster could be defined as a distinct “biomechanical phenotype” a player belongs to.

Players were grouped in a hierarchical bottom-up approach, where individual start in their own group and are iteratively merged based on similarity [304]. This method, the agglomerative clustering [305], is conceptually simple, non-parametric, and capable of identifying clusters of various shapes [306, 307]. To measure the

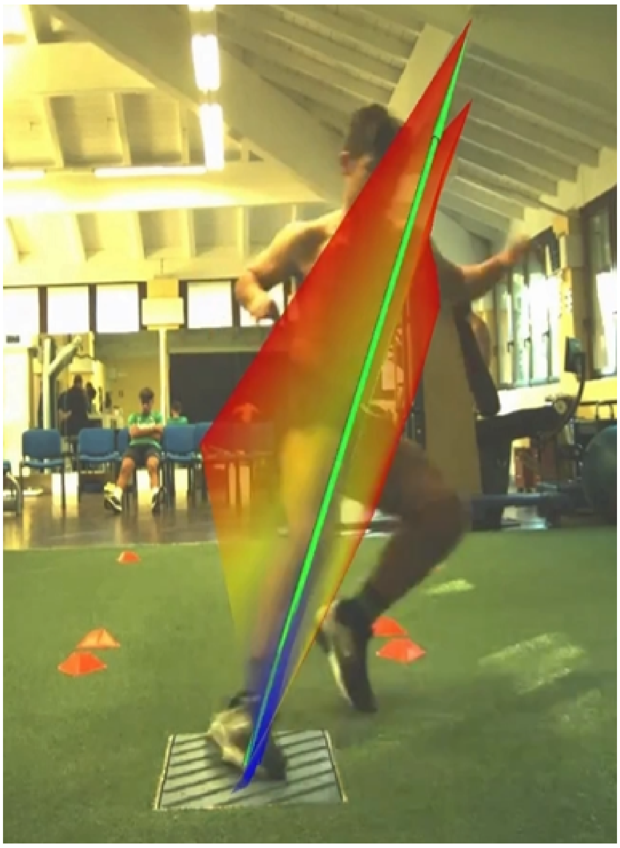

Figure 4.5: Ground-reaction-force vector (blue-green) and impulse surface (red–yellow) on a single COD frame.

similarity between players, the Euclidean distance was used. Additionally, the Ward method was applied to determine how clusters should be merged. The Ward method focuses on minimizing the increase in variance within clusters as they are combined [308]. Next, the silhouette metric [309] was adopted to evaluate the clustering performance. The silhouette measures how similar a player is to its own cluster compared to other clusters. It provides insight into the cohesion, how close players are within their cluster, and separation, how far the clusters are from each other. The silhouette score ranges from -1 to 1, with values closer to 1 indicating well-separated, compact clusters.

We applied Agglomerative Clustering using Python (3.12, Python Software Foundation) and scikit-learn (Version 1.3.1, Pedregosa et al., 2011) a Python library.

During DR, hyperparameters, configuration variables that manage t-SNE implementation, were involved. In particular, t-SNE is affected by perplexity, a parameter which balance local and global structures; lower values highlight small clusters, while higher values reveal broader patterns.

Moreover, since t-SNE involves random processes, it is also affected by a random state, a seed value used to initialize the algorithm.

The representation of the players biomechanical characteristics at the end of DR influences the clustering phase leading to an increase or decrease in silhouette score.

As a consequence of this, RBFOT [310], a non linear optimizer, was selected in order to tune hyperparameters of t-SNE with the objective to maximize the Silhouette.

The search of the correct set of these parameters allows to reach a better final result, from a silhouette of 0.33 using standard values to 0.42 which indicate a sub-optimal result. Continuous data were described with the mean and standard deviation, while categorical data were described with frequency and percentage over the total. The Kruskal-Wallis test, was used to identify statistically significant differences between the identified clusters. Differences were considered statistically significant if $p < 0.05$. Partial Eta-square was computed for each variable to measure the effect size. All the statistical analyses were performed in Python (3.12, Python Software Foundation) using SciPy library (Version 1.10.1; Virtanen et al., 2020).Continuous data were described with the mean and standard deviation, while categorical data were described with frequency and percentage over the total. The Kruskal-Wallis test, was used to identify statistically significant differences between the identified clusters. Differences were considered statistically significant if $p < 0.05$. Partial Eta-square was computed for each variable to measure the effect size. All the statistical analyses were performed in Python (3.12, Python Software Foundation) using SciPy library (Version 1.10.1; Virtanen et al., 2020).

Stage 2 — Supervised phenotyping

A Random-Forest classifier (Scikit-learn 1.3.1, Python 3.12) [**?**] was trained on the complete 45-feature set to predict the phenotype label from stage 1. Model development followed a stratified 3-fold cross-validation scheme. The number of estimators, tested with different random states, was tuned by the optimizer

RBFOpt [**?**]. Random Forest was preferred after brief internal benchmarking for its predictive power and low inference latency on a standard workstation, properties that make the model suitable for real-time deployment.

Stage 3 — Real-time deployment

Real-time deployment requires a lean input layer. Twelve variables were therefore selected by intersecting (i) the highest-ranked features in the Random-Forest importance list, computed as the mean and standard deviation of accumulation of the impurity decrease within each tree, with (ii) variables that clinicians can easily capture in the Green-Room. This compromise maintains predictive power while limiting data entry time. The graphical interface was built with Streamlit 1.30 running on Python 3.12 inside a lightweight Docker image [**?**]. The container bundles both the Streamlit front-end (‘app.py‘) and the serialised Random-Forest model (‘model.pkl‘), eliminating any need for external calls. Deployment occurs on an on-premise Ubuntu 22.04 server equipped with **2**× Intel Xeon E5-2680 v4 CPUs @ 2.40 GHz and 4 GB RAM, situated behind the clinic’s virtual private network (VPN). Because all computation and storage stay on this machine, raw biomechanical data and predictions remain entirely within the clinic’s infrastructure, ensuring full GDPR compliance without cloud reliance.

4.2.2 Results and Discussion

Phenotype Characteristics

The 1002 footballers were successfully clustered into 4 groups. Each group entail specific biomechanical characteristics, the phenotype 4.6. Out of the 36 variables used to represent and cluster the players, each phenotype was summarized as follow to maximize the clinical interpretability and usability of the findings : low force (LF) the lowest peak vertical and posterior GRF. Low control (LC) lowest total score, poor frontal plane control, poor sagittal plane control. High control (HC) highest total score, good frontal plane control, good sagittal plane control. High forces (HF) the highest peak vertical and posterior GRF. In appendix table A.7 were described the results of the statistical analysis and reported mean and variance for continuous values and frequencies for categorical variables. Significant differences were found among the clusters in terms of demographic characteristics, initial contact, and both sagittal and frontal plane kinetics and kinematics. From these results, a description was made for each cluster:

- **Cluster 0: Low-Force / Low-Control (LF–LC).** It is characterized by a very low total score, indicative of a generally suboptimal movement quality. There is poor control in the frontal plane, with a tendency toward knee valgus (inward collapse of the knee) and pelvic tilt. In the sagittal plane, control is also lacking, with limited knee and hip flexion, which leads to increased joint loading. Additionally, this cluster is marked by low ground reaction forces during impact and suboptimal propulsive force during push-off.

- **Cluster 1: Low-Force / High-Control (LF–HC).** It demonstrates a medium-high total score, reflecting a sufficient quality of movement. The frontal plane is well-controlled, with minimal tendency towards knee valgus and good core stability. In the sagittal plane, knee and hip flexion are average, contributing to a more balanced movement. However, similar to Cluster 0, the ground reaction forces during impact are low, and the propulsive force during push-off is also not optimal.

- **Cluster 2: High-Force / Low-Control (HF–LC).** It is defined by a low total score, indicating globally suboptimal movement quality. There is poor control in the frontal plane, with a tendency towards knee valgus and pelvic tilt. In the sagittal plane, knee and hip flexion are average, but there is excessive trunk flexion, which may compromise movement efficiency. Unlike the previous clusters, Cluster 2 is characterized by high ground reaction forces during impact and optimal propulsive force during push-off.

- **Cluster 3: High-Force / High-Control (HF–HC).** Cluster 3, HC and HF. It shows a medium-high total score, indicating sufficient movement quality. The frontal plane is well-controlled, with limited knee valgus and good core stability. In the sagittal plane, knee and hip flexion are average, which supports a balanced movement pattern. This cluster is also marked by high ground reaction forces during impact and optimal propulsive force during push-off.

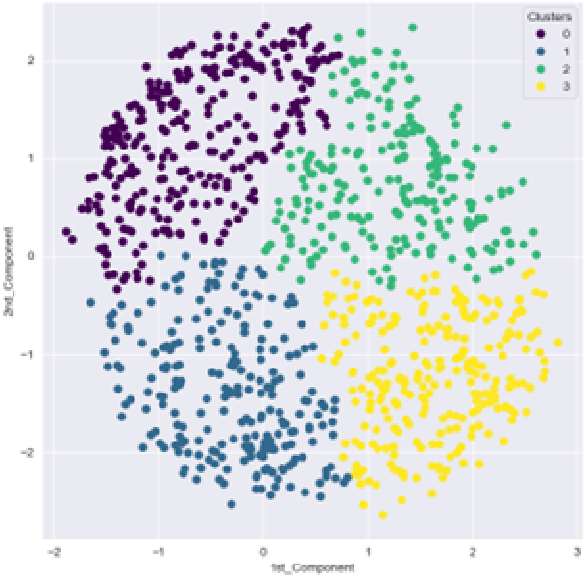

Figure 4.6: t-SNE embedding coloured by the four biomechanical phenotypes. Each dot represents a player.

Classifier Performance

Twelve variables, those in Figure 4.7, were retained for deployment because they sit near the top of the Random-Forest importance ranking *and* can be captured in a low amount of time during the test in Green Room. Table 4.5 reports the 3-fold stratified cross-validated average scores of a Random Forest ($n\_estimators = 560, random\_state = 831$) trained on the full 45-variable set. Separate experiments across random_state values 0–1000 confirmed robustness, with a standard deviation of 0.01 in F1. In contrast, Table 4.6 shows performance when only the twelve deployable variables are supplied. Using all 45 predictors

Table 4.5: Random-Forest average scores (45 features,3-fold stratified CV).

Phenotype	Precision	Recall	F1	Support
LF–LC (0)	0.95	0.94	0.94	95
LF–HC (1)	0.91	0.93	0.92	80
HF–LC (2)	0.91	0.92	0.92	76
HF–HC (3)	0.91	0.90	0.91	80

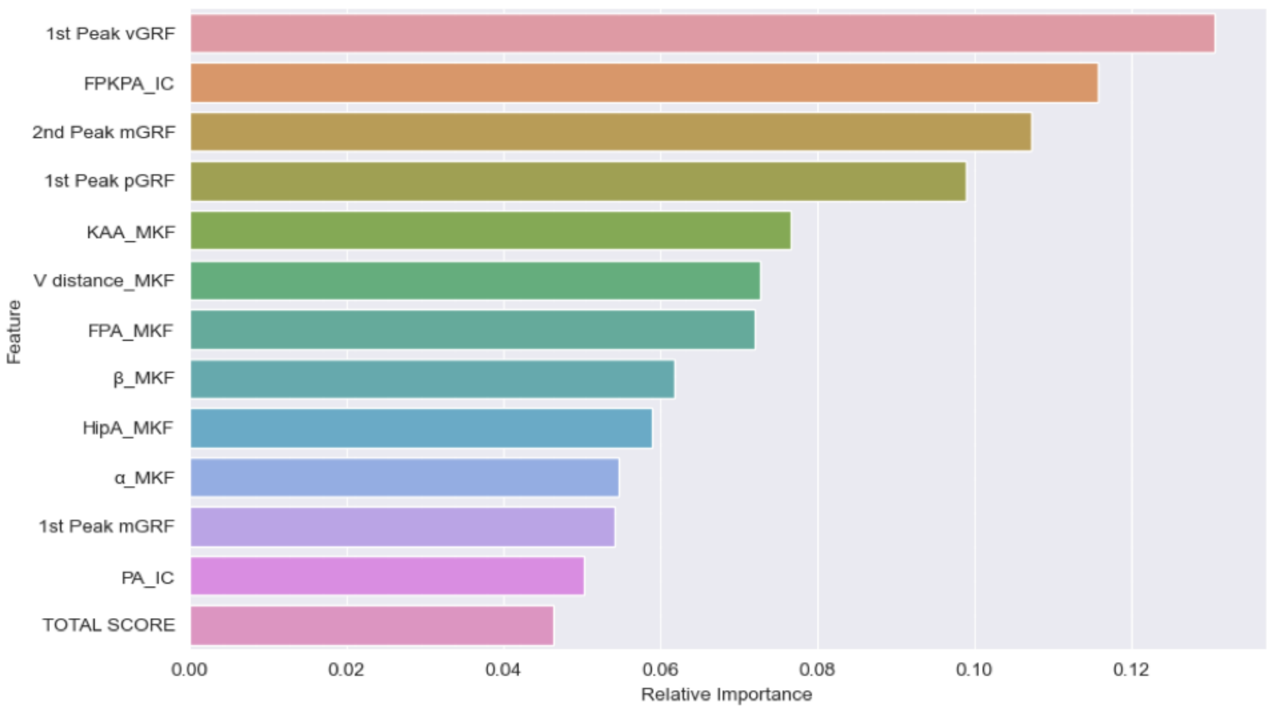

Figure 4.7: Relative importance of the twelve features selected for real-time use.Abbreviations: FPKPA = Frontal-Plane Knee Projection Angle, IC = initial contact, PA = pelvis angle, MKF = maximum-knee-flexion frame, vGRF/pGRF/mGRF = vertical/anterior–posterior / medio-lateral ground-reaction force, KAA = Knee-Abduction Angle, *V distance* = knee–GRF-vector distance. Total Score = Sum of the evaluations given by the doctor during the test.

raises macro-averaged F1 from 0.85 to 0.92, a 7-point gain that nevertheless demands more data entry time. The twelve-feature model, therefore, represents a pragmatic compromise: its accuracy remains within acceptable clinical bounds while cutting input burden to a level feasible for routine Green-Room COD assessments. The corresponding confusion matrix for the twelve-feature model is provided in Figure 4.8.

Clinical Workflow

The application supports the execution of a compact seven-step sequence:

1. **Check-in & warm-up** the athlete registers and completes a 5-min dynamic warm-up while demographics (age, sex) are logged.

Table 4.6: Random-Forest average scores (12 features,3-fold stratified CV).

Phenotype	Precision	Recall	F1	Support
LF–LC (0)	0.87	0.88	0.87	95
LF–HC (1)	0.87	0.85	0.86	80
HF–LC (2)	0.83	0.79	0.81	76
HF–HC (3)	0.83	0.88	0.85	80

2. **COD trials** — six maximal-speed 90° cuts (three per limb) are executed over the embedded force plate while cameras record kinematics.
3. **Manual feature extraction** force and video files are opened in Gpem Screen Editor (Gpem s.r.l., Genova, Italy); the clinician reads the twelve required values (e.g. $\mathrm{vGRF}_{\mathrm{max}}$, $\mathrm{pGRF}_{\mathrm{max}}$, $\mathrm{FPKPA}_{\mathrm{IC}}$) and enters them into the form.
4. **Real-time classification** pressing *Submit* invokes the Random-Forest, which returns the phenotype (LF, LC, HF, HC).
5. **Immediate feedback** an avatar with ground-reaction force arrow, and a 2×2 force–control grid appear (Figure **??**b); phenotype-specific recommendations are displayed at the bottom for clinician-athlete discussion.
6. **Data archiving** an anonymised record (hashed ID, features, label) is appended to the research database, enabling periodic model retraining.

The most important finding of the present study was that we were able, through and unsupervised clustering algorithm, to distinguish specific biomechanical traits of the 90° COD task in our cohort of >1000 young football players. Each cluster was therefore defined as a biomechanical phenotype for the 90° COD task. All the players belonging to one phenotypes has common biomechanical traits that could be used in the assessment of their movement quality and ACL injury risk (Figure 2).

Previous studies have utilized smaller datasets and primarily linear algorithms to analyze movement patterns in athletes. Di Cesare et al. [293] used principal component analysis (PCA) combined with hierarchical clustering to classify coordination patterns during the Drop Vertical Jump (DVJ) task in a sample of

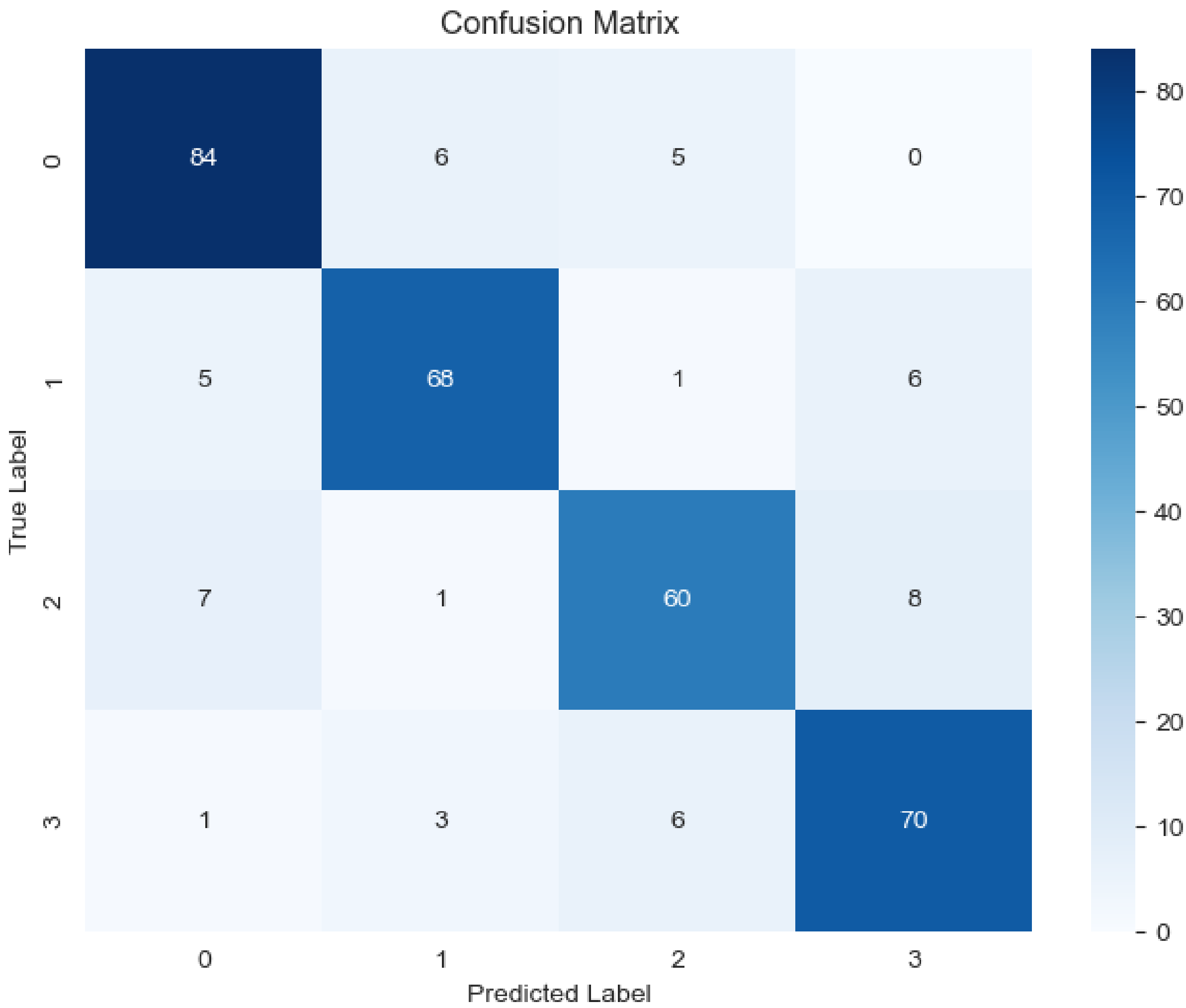

Figure 4.8: Confusion matrix for the twelve-feature classifier.

780 football, and volleyball adolescent athlete, of whom 680 were female, finding three clusters. Similarly to our cluster 4, their Cluster C1-High Neuromuscular Recruitment group exhibited significantly higher power production compared to the other two groups, despite no significant differences in anthropometric measures. Sigurðsson and Briem et al. [292] performed cluster analysis on knee valgus moments during COD tasks with a cohort of 213 youth football and handball athletes, 39% of which were male, using Ward.D2 (Charrad et al., 2014; Murtagh Legendre, 2014) method. They identified 39 clusters with a C-index of 0.049, indicating an optimal separation. Similarly to our study, where female participants were

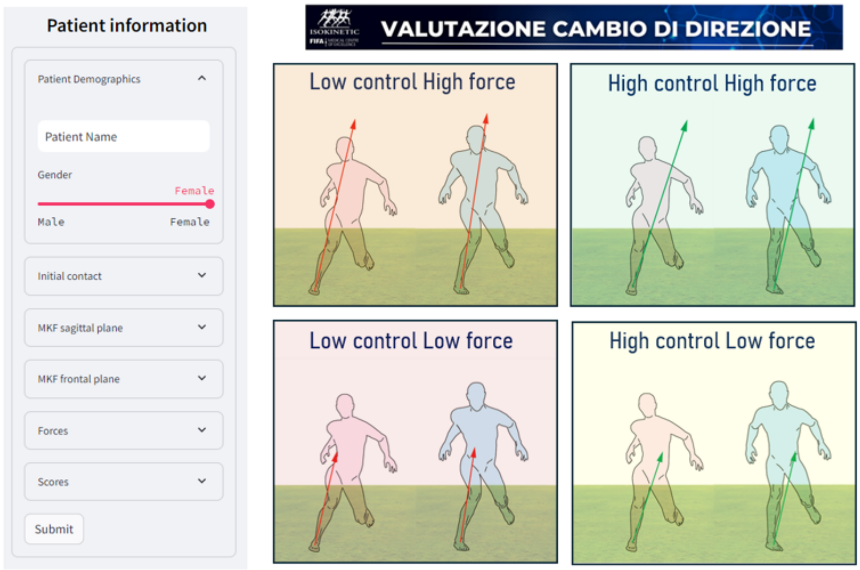

Figure 4.9: Application interface before submitting.

more prevalent in clusters associated with low control (Clusters 1 and 2), their study also discovered a high occurrence of female athletes associated with "Early Peaks" a factor which can lead to an increased risk of ACL injuries, a finding also reported by Sigurðsson et al.[11], in a population from 14 to 17 years old. Bird et al. [296] applied k-means (MacQueen, 1967; Hartigan Wong, 1979; Lloyd, 1982) clustering on a cohort of 668, 559 male and 129 female, Marine Officer Candidates (MOCs) performing countermovement jumps (CMJ). Three distinct clusters of countermovement jump (CMJ) movement strategies were identified, with the high-risk cluster having the highest proportion of musculoskeletal injuries (MSKIs) during military training. The low-risk cluster had higher kinetic variables (e.g., braking rate of force development, braking/propulsive net impulse) and shorter, less flexed joint kinematics compared to the high-risk cluster. Moreover, like in our study, Female MOCs were primarily distributed in the high-risk cluster, while male MOCs were more evenly distributed across clusters. Instead, Rauch et al. [297] examined professional NBA players' CMJ. They used k-means to identify clusters based on joint contributions to movement strategies in a study involving 178 male professional athletes free from any musculoskeletal injuries at the time of data collection. They demonstrated that elite basketball players tend to fall into one of three downward phase movement strategies during the CMJ. Despite

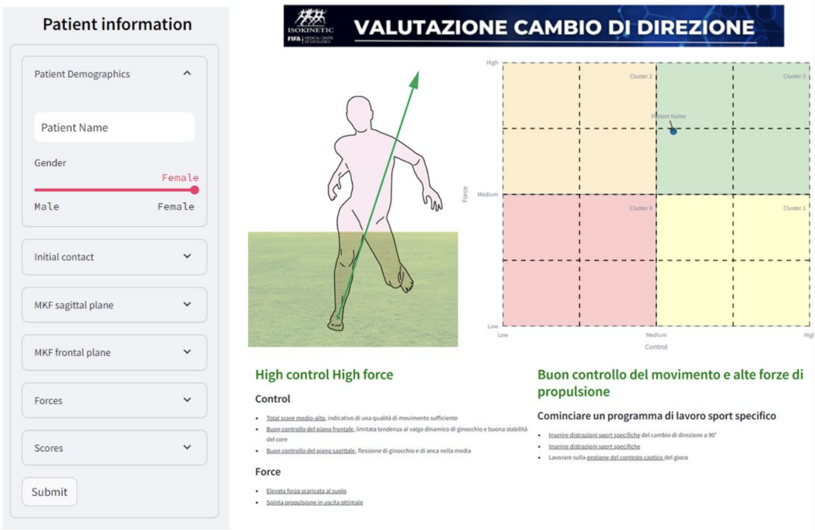

Figure 4.10: Application interface after submitting.

having significantly different downward phase strategies, athletes from each cluster can produce similar jump heights, highlighting the existence of different strategies that allow athletes to achieve excellent results in terms of performance.

In contrast to other studies, our study employed t-SNE, a non-linear dimensionality reduction technique, in combined with agglomerative clustering on a much larger dataset of over 1000 young football players both female (n = 264) and male (n = 738) players. This approach enabled us to capture complex, non-linear relationships inherent in COD movements that linear algorithms with smaller cohort might overlook [104, 303].

Due to the prospective nature of the present study, the players were all free from musculoskeletal injuries at the time of the data collection. Therefore, inferences about the risk of ACL injury should be drawn with caution, for this reason, robust statistics for each variable including were provided. Association to HF-LC o LF-LC cluster have to be considered as a warning when assessing the results of the change of direction test: healthy players exhibiting kinematics associable to HF-LC o LF-LC cluster might deserve attention and corrective intervention to mitigate the risk of knee overloading [50, 301].

The distribution of professional and amateur players varied among the clusters. The LC–HF cluster, despite exhibiting lower movement control, had the highest proportion of professional players (37.99%), along with greater height (174.2 ± 8.6cm) and body weight (660.3 ± 99.2N) ($p < 0.001$) (Table E). This suggests that while these players can produce higher forces, they may sacrifice movement control, potentially increasing injury risk.

In contrast, the HC–HF cluster, which combined high force production with excellent movement control, had a higher proportion of amateur players (73.97%) and was composed of lighter (623.3 ± 101 N) and slightly shorter players (170.7 ,± 9.5cm) (Table A7). They generated higher normalized vertical ground reaction forces (1st Peak vGRF_BW: 3 ± 0.6, N) compared to the heavier LC–HF players (2.8 ± 0.6 N) (p<0.001). The HC–HF groups ability to produce high forces despite being lighter suggests that improved control can aid in force generation.

Regarding gender distribution, the LC–HF and HC–HF clusters had a higher prevalence of male players (85.59% and 83.06%, respectively), whereas the LC–LF and HC–LF clusters included a greater proportion of female players (37.02% and 34.30%, respectively) (Table A7). In particular female players result to be more frequent in the LC-LF cluster demonstrated both low force production and lower movement control w.r.t. male players. Therefore, female players seem to approach the cut with a worse strategy and male players seem more prone to injury-performance conflict as described by Dos'Santos et al. [31]. Focusing on scoring system, total score, combining limb, pelvis, and trunk stability, shock absorption, and movement strategy, was significantly lower in the LC–LF and LC–HF clusters, with 97.58% and 86.89% of participants, respectively, scoring between 0–4 out of 10 (Table E). In contrast, the HC–LF and HC–HF clusters had higher percentages of participants scoring in the 5–7 range, indicating better overall movement quality (p<0.001). Della Villa et al. [18] reported that players with a total score of <4/10 in one or more trials should, therefore, LC–HF and HC–HF clusters deserve greater attention and could be potentially targeted for additional preventative training.

Despite these insights, while the LC–HF group may have a higher risk of joint overloading, this does not necessarily translate to an immediate injury risk without further analysis. Future studies should integrate these biomechanical profiles with injury data to establish more definitive links between movement strategies and ACL injury risk.

The present study has some limitations. First, the task evaluated was a preplanned COD maneuver, which may not fully capture the variability of in-game scenarios. The assessment of unplanned changes of direction has been proposed to better resemble game situations, although this introduces greater variability that could affect the repeatability of the protocol. Additionally, the dimensionality reduction combined with clustering, while valuable for identifying patterns, may oversimplify the complexity of individual movement strategies. Future analysis of these clusters may provide more definitive insights into injury risk prediction, particularly when combined with ACL injury data.

The clinical relevance of this study lies in its potential to inform targeted interventions. By identifying players in the LC–HF and LC–LF clusters, coaches and clinicians can implement specific training programs aimed at improving movement control and reducing excessive joint loading. This tailored approach could mitigate ACL injury risk while enhancing performance. Moreover, the large cohort and comprehensive biomechanical analysis provide a substantial benchmark for future assessments and contribute to the growing body of normative data in this field.

4.2.3 Contributions

This study develops a complete framework for the identification and deployment of biomechanical phenotypes in football players, translating laboratory-grade biomechanics into a clinically actionable tool. At the unsupervised level, forty-five demographic, force, and kinematic variables from a cohort of 1,009 athletes were embedded via t-SNE and clustered with agglomerative linkage, yielding four distinct and interpretable force–control phenotypes that condense complex multidimensional data into clinically meaningful movement strategies. Building on this representation, a Random Forest classifier was trained to reproduce phenotype assignment using only twelve high-importance variables that can be rapidly collected in routine settings, reaching a macro-averaged $F_1 = 0.85$ compared to $F_1 = 0.92$ with the full feature set, thereby achieving a pragmatic balance between predictive accuracy and clinical feasibility. Finally, the model was operationalized through a GDPR-compliant application that enables real-time classification during change-of-direction assessments, providing immediate phenotype output and visual feedback tailored to both clinicians and athletes. By integrating unsupervised discovery, supervised simplification, and compliant deployment, this work contributes a scalable and interpretable decision-support pipeline that bridges biomechanics research with real-world sports medicine practice, supporting both injury-risk management and personalized performance interventions.

Chapter 5

Conclusion

This thesis set out to test whether interpretable, scalable machine learning can meaningfully support three high-leverage decisions along the luxury-automotive journey: personalization, retention, and residual value (RV), and whether the same design patterns transfer to adjacent high-stakes domains.

In response to the personalization research questions (P-RQ1–P-RQ2), we demonstrated that recommender systems can move beyond media and retail and operate effectively in the highly configurable, low-data regime of luxury vehicles. For P-RQ1, viability and accuracy under sparse, long-tailed interactions, the proposed hybrid model that couples collaborative filtering with content signals achieved high predictive performance, with Precision@5 = 0.65, NDCG@5 = 0.66, and a Hit Rate 1.00, while maintaining strong personalization (Mean Inter-List Diversity 0.81). These results confirm that accurate and non-trivial option suggestions can be generated even when user–item data are scarce. For P-RQ2, bias control with preserved personalization and scalability, the model reduced popularity amplification (Popularity Ratio = 1.25 $\rightarrow$ 0.86 from top to long list cutoffs) without eroding relevance, and generalized across new model generations in temporal tests. This contributes a concrete industrial case where recommendations are both technically valid and operationally useful under scarcity.

Addressing the waiting-list research questions (WL-RQ1–WL-RQ3), we presented an interpretable, end-to-end framework for predicting cancellations and translating model outputs into action. For WL-RQ1 (predictive performance under interpretability and governance constraints), a CatBoost classifier distilled to 15 variables via RFE and tuned with Optuna achieved (F1=0.85) in cross-validation and (F1=0.93) on a temporally separated hold-out. For WL-RQ2 (understanding drivers of risk), SHAP delivered transparent global and local attributions. For WL-RQ3 (operationalization), a Streamlit interface and an on-premises, retrieval-augmented LLM enabled concise, rule-grounded recommendations suitable for daily dealership operations. Although still in an experimental phase, early validation indicates faster triage, timelier outreach, and a smoother customer experience during extended waiting-list periods.

For RV, we investigated three questions. In response to RQ1 (accuracy), CatBoost delivered state-of-the-art performance for the luxury segment (mean MAPE (=5.58%), below the 10% industry threshold). For RQ2 (drivers), model family and market geocluster dominated RV variation: "model" behaves as a proxy for intrinsic attributes (reputation, rarity, technical specification, standard options), while "geocluster" captures extrinsic conditions (local economy, regulation, culture). Vehicle age and macroeconomic indicators (CPI, unemployment, bond yields) are nearly as influential as age alone, underscoring the need to account for both internal and external drivers. For RQ3 (trend identification under sparse transactions), the "standard vehicle" construct isolates genuine market depreciation by controlling for sample composition.

A cross-domain analysis reinforced these choices. The same methodological principles developed for luxury automotive: learning under scarcity, constraining complexity for interpretability, and deploying models within compliant, human-in-the-loop systems, were applied to two clinical and sports contexts characterized by even more pronounced data limitations. In response to the clinical research question (C-RQ1), which examined whether interpretable ML could detect genotype-specific subtypes of Parkinson's disease from small, heterogeneous clinical datasets, a lightweight, explainable XGBoost–SHAP pipeline enriched for *GBA1* mutation carriers (73 % accuracy). This demonstrated that, even when patient samples are limited and costly to acquire, transparent ML can prioritize candidates for genetic testing and accelerate genotype-stratified trials. By coupling routinely collected clinical measures with model-driven feature attribution, the approach illustrated how interpretable AI can fill critical gaps in precision medicine workflows, offering a cost-efficient and ethically sound path for integrating predictive tools into clinical decision-making.

In sports medicine, addressing the biomechanical research question (S-RQ1) on whether unsupervised clustering and supervised classification could operationalize force-control phenotypes linked to ACL-injury risk, we applied the same interpretability-first logic to high-dimensional yet sample-limited motion-capture data. Unsupervised phenotyping of change-of-direction mechanics revealed four consistent movement strategies that condensed complex joint and force-based information into clinically interpretable profiles. A distilled Random-Forest classifier trained on twelve high-importance variables reproduced phenotype assignments with macro-F1=0.85 (vs 0.92 with the full feature set), striking a pragmatic balance between predictive accuracy and clinical feasibility. The resulting pipeline was embedded in a GDPR-compliant real-time application capable of classifying new athletes during standard assessment sessions and providing immediate visual feedback to practitioners. This transformation, from laboratory-grade biomechanics to an actionable, interpretable decision-support tool, illustrates how models designed for scarcity and interpretability can translate across domains, maintaining rigor while meeting the operational constraints of real-world practice

Beyond point results, the three automotive levers interact. Better personalization strengthens engagement; stronger engagement reduces cancellations; stable demand and lower churn sustain RV; healthy RV, in turn, reinforces brand desirability and customer confidence, which feeds back into configuration uptake. Framing the lifecycle as this coupled system clarifies why interpretable, operational ML at each node can compound value rather than produce isolated wins.

Limitations remain. Findings are specific to luxury brands; external validation on non-luxury and emerging markets is needed. Macroeconomic coverage was intentionally parsimonious (CPI, unemployment, bond yields); broader PESTEL factors and market microstructure data could refine RV dynamics. For waiting-list management, prospective field studies should quantify causal impact on outreach timing and customer outcomes. For recommendations, additional fairness and preference-diversity audits are warranted to ensure long-term satisfaction rather than short-term acceptance.

Future work will extend market coverage and stress-test transferability beyond luxury; integrate richer macro and behavioral signals (e.g., consumer confidence, policy shocks, secondary-market liquidity); explore compact transformer variants and time-series architectures where they add measurable lift without eroding interpretability; formalize model governance aligned with GDPR and the EU AI Act (monitoring drift, bias, and rationale stability); and continue the cross-domain agenda by evaluating clinical and sports deployments prospectively and at scale.

In sum, the thesis offers a practical blueprint for bringing interpretable ML into day-to-day decisions across the luxury-automotive lifecycle and shows that the same blueprint travels to other high-stakes, low-data environments. By combining accuracy with attribution, and models with interfaces that respect operational realities, the work advances a measured, auditable path to data-driven personalization, retention, and value management, within automotive and beyond.

Appendix A

Supplementary Tables

	severe GBA1-PD (n=26)		mild/risk GBA1-PD (n=32)		P-value	Effect size
Age at PD onset (years)	53.19 [7.95]	{42–71}	59.41 [10.14]	{38–75}	0.01	–0.38
Age at diagnosis (years)	53.65 [7.84]	{42–71}	59.91 [10.03]	{38–75}	0.01	–0.36

Table A.1: Age at onset and diagnosis in severe and mild/risk GBA1-PD. Values are mean [SD]; median {range}.

Table A.2: Demographic and clinical variables that significantly differed between severe GBA1-PD and NM-PD. Values are expressed as mean [SD]; median {range}.

Variable	**severe GBA1-PD**	**NM-PD**	**P-value**	**Effect size**
Family history of PD (Yes/No)	54 % (14/26) / 46 % (12/26)	24 % (14/58) / 76 % (44/58)	0.01	0.27
MDS-UPDRS 3.16a,b: Kinetic tremor R/L hand	0.27 [0.44]; 0 {0–1}	0.07 [0.25]; 0 {0–1}	0.01	0.29
MDS-UPDRS 1.1: Cognitive impairment	1.23 [1.19]; 1.00 {0–3}	0.78 [0.81]; 1.00 {0–3}	0.01	0.36
MDS-UPDRS 3.11: Freezing of gait	1.12 [1.15]; 1 {0–3}	0.40 [0.83]; 0 {0–3}	0.02	0.34
Advanced therapies (Yes/No)	19 % (5/26) / 81 % (21/26)	5 % (3/58) / 95 % (55/58)	0.04	0.23
Dysphagia (Yes/No)	50 % (13/26) / 50 % (13/26)	28 % (16/58) / 72 % (42/58)	0.05	0.24
MDS-UPDRS 3.14: Global spontaneity of movement	1.73 [0.94]; 2.00 {0–3}	1.64 [0.66]; 2.00 {0–3}	0.05	0.31
MDS-UPDRS III ON Total Score	36.8 [13.38]; 37.5 {11–61}	28.95 [13.42]; 29.5 {2–59}	0.04	0.27
Rigidity Subscore	5.62 [3.02]; 5.5 {0–12}	3.97 [3.14]; 3.0 {0–12}	0.03	0.30
PIGD Subscore	7.77 [5.02]; 7.5 {1–19}	5.17 [4.19]; 4.0 {0–16}	0.03	–0.30

Table A.3: Demographic variables

Feature Name	**Definition**
Team (Professionals/Amateurs)	Team type
Age	Age in years
Height	Height of the player
BMI	Body mass index
Familiarità	Yes/no, a member of the family had an ACL injury
Role	Role of the player
Aggressivity	Level of aggressivity (self-assessed)
Gender	Male/Female
BW	Body weight

Table A.4: Variables at initial contact (IC) and MKF

Feature Name	**Definition**
FPA	Foot projection angle
KAA	Knee abduction angle
FPKPA	Frontal plane knee projection angle
V distance	Ground reaction force (GRF) vector–knee distance
PA	Pelvis drop angle
TA	Trunk tilt angle
γ	Ankle dorsiflexion (gamma)
α	Supplementary knee flexion (alpha)
β	$180°-$ (hip flexion + trunk flexion) (beta)
δ	Trunk flexion (delta)
HipA	Hip flexion angle
(IC) Speed	Speed reached at initial contact (IC)
(heel/flat/toe)	Foot support type

Table A.5: Medial, vertical and posterior GRFs and w.r.t. BW

Feature Name	**Definition**
1st, 2nd, and 3rd Peaks mGRF	Medial ground reaction forces
1st and 2nd Peak vGRF	Vertical ground reaction forces
1st and 2nd Peak pGRF	Posterior ground reaction forces

Table A.6: Score variables

Feature Name	**Definition**
LS	Limb stability score
PS	Pelvis stability score
TS	Trunk stability score
SA	Shock absorption score
SM	Movement strategy score
TOTAL SCORE	Total score obtained

Variable	LC LF	HC LF	LC HF	HC HF	Effect Size	p-value
Demographics						
Team (Professionals/Amateurs)	amat: 207 (71.63%), pro: 82 (28.37%)	amat: 188 (77.69%), pro: 54 (22.31%)	amat: 142 (62.01%), pro: 87 (37.99%)	amat: 179 (73.97%), pro: 63 (26.03%)	0.01	0.002
Age	16.1 ± 2.7	16.7 ± 3.0	16.3 ± 2.7	16.1 ± 2.6	0.00	0.145
Height	168 ± 9.6	167.2 ± 9.4	174.2 ± 8.6	170.7 ± 9.5	0.08	<0.001
BMI	21.4 ± 2.9	21.7 ± 2.5	22.2 ± 3.0	21.8 ± 2.7	0.01	0.008
Familiarity	no: 250 (86.51%), yes: 39 (13.49%)	no: 184 (76.03%), yes: 58 (23.97%)	no: 187 (81.66%), yes: 42 (18.34%)	no: 202 (83.47%), yes: 40 (16.53%)	0.01	0.017
BW	594.1 ± 96.6	596.1 ± 86.7	660.3 ± 99.2	623.3 ± 101	0.08	<0.001
Aggressivity	3.3 ± 0.9	3.3 ± 0.8	3.5 ± 0.9	3.3 ± 0.9	0.00	0.084
Role	GK: 33 (11%), DF: 97 (34%), MID: 96 (33%), FW: 63 (22%)	GK: 19 (8%), DF: 86 (36%), MID: 83 (34%), FW: 54 (22%)	GK: 33 (14%), DF: 80 (35%), MID: 46 (20%), FW: 70 (31%)	GK: 21 (9%), DF: 80 (33%), MID: 71 (30%), FW: 70 (29%)	0.00	0.456
Gender	male: 182 (63%), female: 107 (37%)	male: 159 (65.7%), female: 83 (34.3%)	male: 196 (85.6%), female: 33 (14.4%)	male: 201 (83.1%), female: 41 (16.9%)	0.05	<0.001
Scores						
LS	0: 286 (99%), 1: 2 (0.7%), 2: 1 (0.3%)	0: 217 (89.7%), 1: 19 (7.9%), 2: 6 (2.5%)	0: 229 (100%)	0: 218 (90.1%), 1: 19 (7.9%), 2: 5 (2.1%)	0.04	<0.001
PS	0: 277 (95.9%), 1: 11 (3.8%), 2: 1 (0.3%)	0: 193 (79.8%), 1: 38 (15.7%), 2: 11 (4.6%)	0: 216 (94.3%), 1: 10 (4.4%), 2: 3 (1.3%)	0: 187 (77.3%), 1: 50 (20.7%), 2: 5 (2.1%)	0.06	<0.001
TS	0: 142 (49.1%), 1: 104 (36.0%), 2: 43 (14.9%)	1: 101 (41.7%), 0: 77 (31.8%), 2: 64 (26.5%)	0: 109 (47.6%), 1: 78 (34.1%), 2: 42 (18.3%)	1: 107 (44.2%), 2: 83 (34.3%), 0: 52 (21.5%)	0.06	<0.001
SA	1: 161 (55.7%), 0: 125 (43.3%), 2: 3 (1.0%)	1: 190 (78.5%), 2: 48 (19.8%), 0: 4 (1.7%)	1: 182 (79.5%), 0: 35 (15.3%), 2: 12 (5.2%)	1: 177 (73.1%), 2: 58 (24.0%), 0: 7 (2.9%)	0.24	<0.001
SM	0: 199 (68.9%), 1: 72 (24.9%), 2: 18 (6.2%)	2: 133 (55.0%), 1: 78 (32.2%), 0: 31 (12.8%)	2: 99 (43.2%), 1: 82 (35.8%), 0: 48 (21.0%)	2: 131 (54.1%), 1: 84 (34.7%), 0: 27 (11.2%)	0.30	<0.001
Total Score	0–4: 282 (97.6%), 5–7: 7 (2.4%), 8–10: 0 (0%)	0–4: 148 (61.2%), 5–7: 89 (36.8%), 8–10: 5 (2.1%)	0–4: 199 (86.9%), 5–7: 30 (13.1%), 8–10: 0 (0%)	0–4: 138 (57.0%), 5–7: 102 (42.2%), 8–10: 2 (0.8%)	0.36	<0.001
Initial Contact						
(heel/flat/toe)	0.6 ± 0.7	0.1 ± 0.4	0.3 ± 0.5	0.2 ± 0.4	0.11	<0.001
FPA_IC	-44.2 ± 15	-22.7 ± 18.8	-38.0 ± 15.6	-15.9 ± 16.7	0.35	<0.001
KAA_IC	23.4 ± 3.7	16.2 ± 3.5	21.2 ± 3.5	17.1 ± 3.7	0.44	<0.001
FPKPA_IC	30.3 ± 5.8	19.1 ± 6.7	28.3 ± 5.9	18.6 ± 6.0	0.46	<0.001
V distance_IC	19.3 ± 6.3	10.9 ± 5.1	17.2 ± 5.2	10.0 ± 4.1	0.38	<0.001
PA_IC	-18.2 ± 5.4	-11.0 ± 4.9	-15.5 ± 4.8	-10.5 ± 4.2	0.33	<0.001
TA_IC	0.0 ± 7.6	1.1 ± 4.9	-0.2 ± 6.4	1.7 ± 4.5	0.00	0.052
γ_IC	-16.4 ± 11.3	-4.8 ± 8.3	-14.4 ± 9.2	-9.5 ± 9.3	0.19	<0.001
α_IC	151.9 ± 6.9	146.2 ± 6.3	147.9 ± 7.4	146.3 ± 7.3	0.11	<0.001
β_IC	108.9 ± 8.7	103.1 ± 11.5	98.2 ± 9.5	102.4 ± 11.0	0.14	<0.001
δ_IC	19.9 ± 6.5	21.1 ± 7.7	25.1 ± 6.8	21.0 ± 8.2	0.07	<0.001
HipA_IC	50.6 ± 4.9	55.1 ± 5.5	55.9 ± 5.5	55.6 ± 5.4	0.16	<0.001
Strike time	0.3 ± 0.1	0.4 ± 0.1	0.4 ± 0.1	0.4 ± 0.1	0.04	<0.001
Approaching speed	4.9 ± 1.6	5.0 ± 1.8	4.8 ± 1.7	5.2 ± 1.2	0.00	0.123
IC Speed	3.1 ± 1.0	3.1 ± 0.9	3.5 ± 1.1	3.7 ± 0.9	0.07	<0.001
Maximum Knee Flexion (MKF)						
FPA_MKF	-44.9 ± 13.0	-19.9 ± 16.2	-36.1 ± 13.4	-15.6 ± 15.8	0.43	<0.001
KAA_MKF	43.9 ± 5.0	33.6 ± 6.3	40.2 ± 5.0	33.2 ± 5.5	0.46	<0.001
FPKPA_MKF	51.7 ± 8.8	37.9 ± 11.6	48.6 ± 7.8	34.8 ± 9.9	0.39	<0.001
V>K_MKF	0.0 ± 0.1	0.4 ± 0.6	0.0 ± 0.2	0.4 ± 0.6	0.17	<0.001
V distance_MKF	8.6 ± 2.2	4.5 ± 2.8	7.2 ± 2.5	3.8 ± 2.2	0.43	<0.001
PA_MKF	-20.1 ± 5.6	-13.8 ± 5.3	-17.9 ± 5.3	-13.5 ± 4.8	0.23	<0.001
TA_MKF	-0.9 ± 11.2	0.5 ± 8.5	-1.6 ± 11.0	0.3 ± 7.8	0.00	0.312
γ_MKF	1.2 ± 5.7	3.8 ± 6.9	-0.1 ± 4.8	1.0 ± 4.5	0.07	<0.001
α_MKF	127.3 ± 8.8	113.5 ± 6.6	120.1 ± 8.0	113.6 ± 7.0	0.38	<0.001
β_MKF	109.4 ± 10.5	95.9 ± 12.5	95.8 ± 10.7	95.4 ± 12.2	0.24	<0.001
δ_MKF	23.0 ± 7.5	28.1 ± 8.9	29.1 ± 7.6	27.1 ± 9.0	0.08	<0.001
HipA_MKF	46.8 ± 5.6	55.0 ± 5.9	54.2 ± 5.4	56.7 ± 5.8	0.33	<0.001
Forces (GRFs)						
2nd Peak mGRF	484 ± 134.9	347.1 ± 85.4	656.6 ± 149.6	557.7 ± 137.3	0.46	<0.001
3rd Peak mGRF	503.4 ± 106.5	462.1 ± 93.2	577.3 ± 108.5	516.2 ± 109.2	0.13	<0.001
1st Peak vGRF	1218.6 ± 270.2	1235.6 ± 227.1	1797.7 ± 386.7	1870.7 ± 403.6	0.54	<0.001
2nd Peak vGRF	1085.9 ± 190.0	1014.4 ± 171.0	1250.4 ± 211.7	1153.9 ± 219.2	0.15	<0.001
1st Peak pGRF	597.9 ± 162.1	557.1 ± 126.7	865.4 ± 202.2	919.1 ± 212.6	0.51	<0.001
2nd Peak mGRF_BW	0.8 ± 0.2	0.6 ± 0.1	1.0 ± 0.2	0.9 ± 0.2	0.39	<0.001
3rd Peak mGRF_BW	0.9 ± 0.1	0.8 ± 0.1	0.9 ± 0.1	0.8 ± 0.1	0.09	<0.001
1st Peak vGRF_BW	2.1 ± 0.4	2.1 ± 0.4	2.8 ± 0.6	3.0 ± 0.6	0.45	<0.001
2nd Peak vGRF_BW	1.8 ± 0.2	1.7 ± 0.2	1.9 ± 0.2	1.9 ± 0.2	0.13	<0.001
1st Peak pGRF_BW	1.0 ± 0.3	0.9 ± 0.2	1.3 ± 0.3	1.5 ± 0.4	0.40	<0.001
1st Peak mGRF	103.7 ± 60.5	43.2 ± 32.4	104.8 ± 71.3	48.5 ± 46.3	0.26	<0.001

Table A.7: Biomechanical differences at initial contact (IC), maximum knee flexion (MKF), and ground reaction forces (GRFs).

Bibliography

[1] Bain & Company. *Luxury in Transition: Securing Future Growth*. https://www.bain.com/insights/luxury-in-transition-securing-future-growth/. Accessed August 2025. 2025.

[2] McKinsey & Company. *Shifting Gears: What Buyers Are Saying About the Luxury Car Experience*. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/shifting-gears-what-buyers-are-saying-about-the-luxury-car-experience. Accessed August 2025. 2025.

[3] Bain & Company. *Luxury in Transition: Securing Future Growth*. https://www.bain.com/insights/luxury-in-transition-securing-future-growth/. Accessed August 2025. 2025.

[4] Ferrari N.V. *ANOTHER YEAR OF STRONG PERFORMANCE, ALL 2024 TARGETS EXCEEDED. FURTHER GROWTH IN 2025*. https://www.ferrari.com/en-EN/corporate/articles/2024-full-year-and-fourth-quarter-financial-results. Accessed August 30, 2025. 2025.

[5] Reuters. *Ferrari lifts forecasts on pricier models and personal touches*. https://www.reuters.com/business/autos-transportation/ferrari-ups-fy-forecasts-after-strong-q2-larger-offer-personalisations-2024-08-01/. Accessed August 30, 2025. 2024.

[6] KPMG. *The Market of Luxury Goods*. https://assets.kpmg.com/content/dam/kpmg/gr/pdf/2024/02/gr-kpmg-future-of-consumer-goods-the-market-of-luxury-goods.pdf. Accessed 30 Aug 2025. 2024.

[7] Branding Strategy Insider. *Luxury Brand Strategy: Managing Exclusivity and Availability*. https://brandingstrategyinsider.com/luxury-brand-strategy-managing-exclusivity-and-availability/. Accessed 30 Aug 2025. 2024.

[8] Konvi Magazine. *How Exclusivity Shapes the Luxury Market: Scarcity and Demand*. https://konvi.app/en_GB/magazine/article/how-exclusivity-shapes-the-luxury-market-scarcity-demand/. Accessed 30 Aug 2025. 2025.

[9] Akshay Rao and Kent B. Monroe. "An Empirical Analysis of Scarcity Strategies in the Automobile Industry". In: *Journal of Marketing Research* 62.3 (2025). Accessed 30 Aug 2025, pp. 415–432. DOI: 10.1177/00222437211O535117.

[10] S&P Global Mobility. *Driving Success with Customer Loyalty in the Automotive Industry*. https://www.spglobal.com/automotive-insights/en/blogs/2025/02/strategies-to-grow-customer-loyalty-automotive-industry. Accessed 30 Aug 2025. 2025.

[11] CBT News. *Innovative Strategies to Enhance Your Dealership's Customer Retention*. https://www.cbtnews.com/innovative-strategies-to-enhance-your-dealerships-customer-retention/. Accessed 30 Aug 2025. 2025.

[12] Deloitte. *2024 Global Automotive Consumer Study*. https://www2.deloitte.com/us/en/pages/manufacturing/articles/global-automotive-consumer-study.html. Accessed 30 Aug 2025. 2024.

[13] McKinsey & Company. *Five trends shaping tomorrow's luxury car market*. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/five-trends-shaping-tomorrows-luxury-car-market. Accessed 30 Aug 2025. 2024.

[14] Boston Consulting Group. *The Future of Automotive: Value Chains, Mobility, and Profit Pools*. https://www.bcg.com/publications/2023/future-of-automotive-value-chains-mobility-profit-pools. Accessed 30 Aug 2025. 2023.

[15] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th. Pearson, 2021.

[16] McKinsey & Company. *Artificial Intelligence in Mobility: How AI Will Reshape the Auto Industry*. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/artificial-intelligence-in-mobility. Accessed 30 Aug 2025. 2023.

[17] Deloitte. *AI in the Automotive Industry: Driving Change Across the Value Chain*. https://www2.deloitte.com/global/en/pages/manufacturing/articles/ai-in-the-automotive-industry.html. Accessed 30 Aug 2025. 2024.

[18] Boston Consulting Group. *The Future of AI in Automotive and Mobility*. https://www.bcg.com/publications/2024/ai-in-automotive-and-mobility. Accessed 30 Aug 2025. 2024.

[19] Tom M. Mitchell. *Machine Learning*. Classic definition of ML. McGraw-Hill, 1997.

[20] Dietmar Jannach and Lukas Lerche. "Recommendation Technologies for Configurable Products". In: *AI Magazine* 32.3 (2011), pp. 99–108. DOI: 10.1609/aimag.v32i3.2369.

[21] Francesco Ricci, Lior Rokach, and Bracha Shapira. "Introduction to Recommender Systems Handbook". In: *Recommender Systems Handbook*. 2nd. Springer, 2015, pp. 1–35. DOI: 10.1007/978-0-387-85820-3_1.

[22] Gediminas Adomavicius and Alexander Tuzhilin. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions". In: *IEEE Transactions on Knowledge and Data Engineering* 17.6 (2005), pp. 734–749. DOI: 10.1109/TKDE.2005.99.

[23] Alexander Felfernig, Stefan Reiterer, Michael Stettinger, et al. "Knowledge-Based Recommender Systems for Configuration Tasks". In: *AI EDAM* 28.1 (2014), pp. 5–21. DOI: 10.1017/S0890060413000564.

[24] Maurizio Ferrari Dacrema et al. "Design and Evaluation of Cross-Domain Recommender Systems". In: *Recommender Systems Handbook*. 4th. Springer, 2022, pp. 485–516. DOI: 10.1007/978-1-0716-2197-4_13.

[25] Harald Steck et al. “Deep Learning for Recommender Systems: A Netflix Case Study”. In: *AI Magazine* 42.3 (2021), pp. 59–70. DOI: 10.1609/aimag.v42i3.18140.

[26] Andrew I. Schein et al. “Methods and Metrics for Cold-Start Recommendations”. In: *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* 2002, pp. 253–260. DOI: 10.1145/564376.564421.

[27] Yiu-Kai Ng et al. “Addressing the Cold-Start Problem in Recommender Systems”. In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshop.* 2008, pp. 89–97. DOI: 10.1145/1352793.1352837.

[28] Robin Burke. “Hybrid Recommender Systems: Survey and Experiments”. In: *User Modeling and User-Adapted Interaction* 12.4 (2002), pp. 331–370. DOI: 10.1023/A:1021240730564.

[29] E. Caṇo and M. Morisio. “Hybrid Recommender Systems: A Systematic Literature Review”. In: *Intelligent Data Analysis* 21.6 (2017), pp. 1487–1524. DOI: 10.3233/IDA-163209.

[30] Xiangnan He et al. “Neural Collaborative Filtering”. In: *Proceedings of the 26th International Conference on World Wide Web.* 2017, pp. 173–182. DOI: 10.1145/3038912.3052569.

[31] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. “Collaborative Deep Learning for Recommender Systems”. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2015, pp. 1235–1244. DOI: 10.1145/2783258.2783273.

[32] A. Ampountolas. “Predicting hotel booking cancellations: a comprehensive machine learning approach”. In: *Journal of Revenue and Pricing Management* (2025), pp. 1–12.

[33] S. E. Schaeffer and S. V. R. Sanchez. “Forecasting client retention—a machine-learning approach”. In: *Journal of Retailing and Consumer Services* 52 (2020), p. 101918.

[34] M. Tekin and M. Gök. “Performance comparison of classification algorithms in hotel booking cancellation prediction”. In: *Artificial Intelligence Theory and Applications* 1.1 (2021), pp. 8–19.

[35] M. A. Jishan et al. “Hotel booking cancellation prediction using applied Bayesian models”. In: *Proc. International Conference on Decision Aid Sciences and Applications (DASA)*. IEEE, 2024, pp. 1–5.

[36] A. Essien and G. Chukwukelu. “Deep learning in hospitality and tourism: a research framework agenda for future research”. In: *International Journal of Contemporary Hospitality Management* 34.12 (2022), pp. 4480–4515.

[37] G. K. Dziugaite, S. Ben-David, and D. M. Roy. “Enforcing interpretability and its statistical impacts: Trade-offs between accuracy and interpretability”. In: *arXiv preprint arXiv:2010.13764* (2020).

[38] S. E. Schaeffer and S. V. R. Sanchez. “Forecasting client retention—a machine-learning approach”. In: *Journal of Retailing and Consumer Services* 52 (2020), p. 101918.

[39] K. Cowling and J. Cubbin. “Hedonic price indexes for United Kingdom cars”. In: *Economic Journal* 82.327 (1972), p. 963. DOI: 10.2307/2230261.

[40] M. Ohta and Z. Griliches. “Automobile prices revisited: Extensions of the hedonic hypothesis”. In: *NBER*. 1998, pp. 415–480. URL: https://www.nber.org/system/files/chapters/c3966/c3966.pdf.

[41] J. Du, L. Xie, and S. Schroeder. “Practice Prize Paper—PIN optimal distribution of auction vehicles system: Applying price forecasting, elasticity estimation, and genetic algorithms to used-vehicle distribution”. In: *Marketing Science* 28.4 (2009), pp. 637–644. DOI: 10.1287/mksc.1080.0470.

[42] J.-D. Wu, C.-C. Hsu, and H.-C. Chen. “An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference”. In: *Expert Systems with Applications* 36.4 (2009), pp. 7809–7817. DOI: 10.1016/j.eswa.2008.11.019.

[43] S. Lessmann and S. Voß. “Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy”. In: *International Journal of Forecasting* 33.4 (2017), pp. 864–877. DOI: 10.1016/j.ijforecast.2017.04.003.

[44] C. Gleue et al. “Decision support for the automotive industry: Forecasting residual values using artificial neural networks”. In: *Business and Information Systems Engineering* 61.4 (2019), pp. 385–397. DOI: 10.1007/s12599-018-0527-3.

[45] E. Liu et al. “Research on the prediction model of the used car price in view of the PSO-GRA-BP neural network”. In: *Sustainability* 14.15 (2022), p. 8993. DOI: 10.3390/su14158993.

[46] J. Sharma and S. K. Mitra. “Developing a used car pricing model applying multivariate adaptive regression splines approach”. In: *Expert Systems with Applications* 236 (2024), p. 121277. DOI: 10.1016/j.eswa.2023.121277.

[47] K. Ranjith. “Prediction of resale value of pre-owned luxury cars in the Indian market employing machine learning techniques”. In: *International Journal of Research in Applied Science and Engineering Technology* 12.7 (2024), pp. 1026–1037. DOI: 10.22214/ijraset.2024.63709.

[48] S. Bergmann and S. Feuerriegel. “Machine learning for predicting used car resale prices using granular vehicle equipment information”. In: *Expert Systems with Applications* 263 (2025), p. 125640. DOI: 10.1016/j.eswa.2024.125640.

[49] Scott M. Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions”. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Vol. 30. Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

[50] Timothy E. Hewett et al. “Mechanisms, prediction, and prevention of ACL injuries: Cut risk with three sharpened and validated tools”. In: *Journal of Orthopaedic Research* 34.11 (2016), pp. 1843–1855. DOI: 10.1002/jor.23414.

[51] Trent Nessler, Linda Denney, and Justin Sampley. “ACL Injury Prevention: What Does Research Tell Us?” In: *Current Reviews in Musculoskeletal Medicine* 10.3 (2017), pp. 281–288. DOI: 10.1007/s12178-017-9416-5.

[52] Markus Waldén et al. “ACL injuries in men’s professional football: a 15-year prospective study on time trends and return-to-play rates reveals only 65% of players still play at the top level 3 years after ACL rupture”. In: *British Journal of Sports Medicine* 50.12 (2016), pp. 744–750. DOI: 10.1136/bjsports-2015-095952.

[53] E. Sidransky and et al. “Multicenter analysis of glucocerebrosidase mutations in Parkinson’s disease”. In: *New England Journal of Medicine* 361.17 (Oct. 2009), pp. 1651–1661. DOI: 10.1056/NEJMoa0901281.

[54] T. B. Stoker and R. A. Barker. “Recent developments in the treatment of Parkinson’s Disease”. In: *F1000Research* 9 (July 2020), F1000 Faculty Rev–862. DOI: 10.12688/f1000research.25634.1.

[55] Wikipedia contributors. *Machine Learning*. https://en.wikipedia.org/wiki/Machine_learning. Accessed: 2025-10-05.

[56] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th ed. Pearson, 2021.

[57] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning”. In: *Nature* 521.7553 (2015), pp. 436–444.

[58] Alan M Turing. “Computing machinery and intelligence”. In: *Mind* 59.236 (1950), pp. 433–460.

[59] Wikipedia contributors. *Turing Test*. https://en.wikipedia.org/wiki/Turing_test. Accessed: 2025-10-05.

[60] Arthur L Samuel. “Some studies in machine learning using the game of checkers”. In: *IBM Journal of Research and Development* 3.3 (1959), pp. 210–229.

[61] Frank Rosenblatt. “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain”. In: *Psychological Review* 65.6 (1958), pp. 386–408. DOI: 10.1037/h0042519.

[62] Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969.

[63] John McCarthy et al. “A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955”. In: *AI Magazine*. Vol. 27. 4. Reprint of the original 1955 proposal. 2006, pp. 12–14.

[64] James Lighthill. *Artificial intelligence: a general survey*. Tech. rep. Science Research Council, UK, 1973.

[65] Edward Hance Shortliffe. *Computer-based medical consultations: MYCIN*. Elsevier, 1976.

[66] Daniel Crevier. *AI: The tumultuous history of the search for artificial intelligence*. Basic Books, 1993.

[67] Leo Breiman et al. *Classification and Regression Trees*. Wadsworth, 1984.

[68] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.

[69] Leo Breiman. “Random forests”. In: *Machine Learning* 45.1 (2001), pp. 5–32.

[70] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[71] Lawrence R Rabiner. “A tutorial on hidden Markov models and selected applications in speech recognition”. In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286.

[72] Corinna Cortes and Vladimir Vapnik. “Support-vector networks”. In: *Machine Learning*. Vol. 20. 3. Springer, 1995, pp. 273–297.

[73] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. “Reducing the dimensionality of data with neural networks”. In: *Science* 313.5786 (2006), pp. 504–507.

[74] Rajat Raina, Anand Madhavan, and Andrew Y Ng. “Large-scale deep unsupervised learning using graphics processors”. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM. 2009, pp. 873–880.

[75] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet classification with deep convolutional neural networks". In: *Advances in Neural Information Processing Systems*. Vol. 25. 2012, pp. 1097–1105.

[76] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780.

[77] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks". In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE. 2013, pp. 6645–6649.

[78] Ashish Vaswani et al. "Attention is all you need". In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017.

[79] Andre Esteva et al. "Dermatologist-level classification of skin cancer with deep neural networks". In: *Nature* 542.7639 (2017), pp. 115–118. DOI: `10.1038/nature21056`.

[80] Eric J Topol. "High-performance medicine: the convergence of human and artificial intelligence". In: *Nature Medicine* 25.1 (2019), pp. 44–56.

[81] Stefan Lessmann et al. "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research". In: *European Journal of Operational Research* 247.1 (2015), pp. 124–136.

[82] Shihao Gu, Bryan Kelly, and Dacheng Xiu. "Empirical asset pricing via machine learning". In: *The Review of Financial Studies* 33.5 (2020), pp. 2223–2273.

[83] Susanne Böhm and Patrick Riederer. "Autonomous driving: A comparative perspective through case studies in Germany, the USA, and Japan". In: *AI Perspectives* 3.1 (2021), pp. 1–12.

[84] Haitao Yuan and Guoliang Li. "A Survey of Traffic Prediction: From Spatio-Temporal Data to Intelligent Transportation". In: *Data Science and Engineering* 6.1 (2021), pp. 63–85. DOI: `10.1007/s41019-020-00151-z`.

[85] Andreas Kamilaris and Francesc X Prenafeta-Boldú. "Deep learning in agriculture: A survey". In: *Computers and Electronics in Agriculture* 147 (2018), pp. 70–90. DOI: `10.1016/j.compag.2018.02.016`.

[86] David Rolnick et al. "Tackling climate change with machine learning". In: *arXiv preprint arXiv:1906.05433* (2019). DOI: 10.48550/arXiv.1906.05433.

[87] Sebastian Raschka. "Model evaluation, model selection, and algorithm selection in machine learning". In: *arXiv preprint arXiv:1811.12808* (2018). URL: https://arxiv.org/abs/1811.12808.

[88] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. ISBN: 978-0-387-31073-2.

[89] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. 2nd ed. MIT Press, 2018. ISBN: 978-0-262-03924-6.

[90] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009. DOI: 10.1007/978-0-387-84858-7.

[91] Sandeep Singh Sengar et al. "Generative artificial intelligence: a systematic review and applications". In: *Multimedia Tools and Applications* 83.9 (2024), pp. 6019–6052. DOI: 10.1007/s11042-024-20016-1.

[92] Pedro Domingos. "A few useful things to know about machine learning". In: *Communications of the ACM* 55.10 (2012), pp. 78–87. DOI: 10.1145/2347736.2347755.

[93] Leonardo Brigato and Luca Iocchi. "A Close Look at Deep Learning with Small Data". In: *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021, pp. 2490–2497. DOI: 10.1109/ICPR48806.2021.9412492.

[94] Ben Cottier et al. "The rising costs of training frontier AI models". In: *arXiv preprint arXiv:2405.21015* (2024). DOI: 10.48550/arXiv.2405.21015.

[95] IBM. *CPU vs. GPU for Machine Learning*. *IBM Think* article. Discusses when GPUs are justified vs CPU/ML models in cost and infrastructure. 2025. URL: https://www.ibm.com/think/topics/cpu-vs-gpu-machine-learning.

[96] Zachary C. Lipton. "The mythos of model interpretability". In: *Queue* 16.3 (2018), pp. 31–57. DOI: 10.1145/3236386.3241340.

[97] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. “"Why should I trust you?": Explaining the predictions of any classifier”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1135–1144. DOI: 10.1145/2939672.2939778.

[98] Cynthia Rudin. “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”. In: *Nature Machine Intelligence* 1.5 (2019), pp. 206–215. DOI: 10.1038/s42256-019-0048-x.

[99] Christoph Molnar. *Interpretable Machine Learning*. 2nd ed. Lulu.com, 2022. URL: https://christophm.github.io/interpretable-ml-book/.

[100] Tianqi Chen and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’16)*. San Francisco, CA, USA: ACM, 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.

[101] Liudmila Prokhorenkova et al. “CatBoost: unbiased boosting with categorical features”. In: *Advances in Neural Information Processing Systems*. Vol. 31. 2018.

[102] Haoyi Huang et al. “Towards a comprehensive evaluation of dimension reduction methods”. In: *Communications Biology* 5.1 (2022), pp. 1–16. DOI: 10.1038/s42003-022-03628-x.

[103] Harold Hotelling. “Analysis of a complex of statistical variables into principal components”. In: *Journal of Educational Psychology* 24.6 (1933), pp. 417–441. DOI: 10.1037/h0071325.

[104] Laurens van der Maaten and Geoffrey Hinton. “Visualizing data using t-SNE”. In: *Journal of Machine Learning Research*. Vol. 9. 2008, pp. 2579–2605.

[105] Leland McInnes, John Healy, and James Melville. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. In: *arXiv preprint arXiv:1802.03426* (2018).

[106] Beyza Eken et al. "A Multivocal Review of MLOps Practices, Challenges and Open Issues". In: vol. 58. 2. Peer-reviewed. New York, NY, USA: Association for Computing Machinery, 2025. DOI: 10.1145/3747346.

[107] Sarah Azzabi, Zakiya Alfughi, and Abdelkader Ouda. "Data Lakes: A Survey of Concepts and Architectures". In: *Computers* 13.7 (2024), p. 183. DOI: 10.3390/computers13070183.

[108] Sang Gyu Kwak and Jong Hae Kim. "Statistical data preparation: management of missing values and outliers". In: *Korean Journal of Anesthesiology* 70.4 (2017), pp. 407–415.

[109] S.S. Sahoo et al. "Ontology-based feature engineering in machine learning improves performance with heterogeneous clinical data". In: *Scientific Reports* 12 (2022), p. 23101. DOI: 10.1038/s41598-022-23101-3.

[110] J. M. Nápoles-Duarte et al. "Stmol: A component for building interactive molecular visualizations within streamlit web-applications". In: *Frontiers in Molecular Biosciences* 9 (2022), p. 990846. DOI: 10.3389/fmolb.2022.990846.

[111] Noyan Aendikov and Aeila Azayeva. "Integration of GIS and machine learning analytics into Streamlit application". In: *Procedia Computer Science* 231 (2024), pp. 691–696. DOI: 10.1016/j.procs.2023.12.160.

[112] Răzvan Buga et al. "Streamlit Application and Deep Learning Model for Brain Metastasis Monitoring After Gamma Knife Treatment". In: *Biomedicines* 13.2 (2025), p. 423. DOI: 10.3390/biomedicines13020423.

[113] Moses Openja, Niklas Koehler, and Christian Peukert. "Studying the Practices of Deploying Machine Learning Projects on Docker". In: *arXiv preprint arXiv:2206.00699* (2022). URL: https://arxiv.org/abs/2206.00699.

[114] Ben Vaillancourt et al. "Reproducible and Portable Workflows for Scientific Computing and HPC in the Cloud". In: *arXiv preprint arXiv:2006.05016* (2020). URL: https://arxiv.org/abs/2006.05016.

[115] Mihir Muzumdar et al. “Navigating the Docker Ecosystem: A Comprehensive Taxonomy and Survey”. In: *arXiv preprint arXiv:2403.17940* (2024). URL: `https://arxiv.org/abs/2403.17940`.

[116] Franck Vigneron and Lester W. Johnson. “Measuring perceptions of brand luxury”. In: *Journal of Brand Management* 11.6 (2004), pp. 484–506.

[117] Klaus-Peter Wiedmann, Nadine Hennigs, and Astrid Siebels. “Value-based segmentation of luxury consumption behavior”. In: *Psychology & Marketing* 26.7 (2009), pp. 625–651. DOI: `10.1002/mar.20292`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/mar.20292`.

[118] Young Jee Han, Joseph C. Nunes, and Xavier Drèze. “Signaling Status with Luxury Goods: The Role of Brand Prominence”. In: *Journal of Marketing* 74.4 (2010), pp. 15–30. URL: `https://msbfile03.usc.edu/digitalmeasures/jnunes/intellcont/Brand%20Prominence%201-12-10-1.pdf`.

[119] KPMG. *Influences on luxury consumer trends*. 2022. URL: `https://kpmg.com/ph/en/home/insights/2022/03/influences-on-luxury-consumer-trends.html`.

[120] Subramanian Balachander, Yan Liu, and Axel Stock. “An Empirical Analysis of Scarcity Strategies in the Automobile Industry”. In: *Management Science* 55.10 (2009), pp. 1623–1637. DOI: `10.1287/mnsc.1090.1056`. URL: `https://pubsonline.informs.org/doi/10.1287/mnsc.1090.1056`.

[121] Florian Zettelmeyer et al. *How Inventory Matters: Empirical Evidence on the Effects of Inventory Levels on Automobile Prices*. Working Paper 12177. National Bureau of Economic Research, 2006. URL: `https://www.nber.org/papers/w12177`.

[122] Porsche AG. *The new Porsche 911 S/T: special-edition model marks 60th anniversary of the 911*. 2023. URL: `https://newsroom.porsche.com/en/press-kits/911-s-t/highlights.html`.

[123] Ferrari N.V. *Ferrari N.V. Q4 and Full-Year 2024 Results Conference Call Transcript*. Management commentary indicating personalizations ~20% of cars & spare parts revenue. 2025. URL: `https://seekingalpha.com/`

article/4718251-ferrari-n-v-race-q4-2024-earnings-call-transcript.

[124] Ferrari N.V. *Ferrari 2024 Full-Year Results*. Corporate results news and presentation. 2025. URL: https://www.ferrari.com/en-EN/corporate/articles/2024-full-year-results.

[125] Ferrari S.p.A. *Ferrari Personalisation: Atelier and Tailor Made*. 2025. URL: https://www.ferrari.com/en-EN/auto/personalization.

[126] Automobili Lamborghini. *Customization: Ad Personam*. 2025. URL: https://www.lamborghini.com/en-en/customization.

[127] Bentley Motors. *Coachbuilt by Mulliner*. 2025. URL: https://www.bentleymotors.com/en/models/mulliner/coachbuilt.html.

[128] Porsche AG. *Porsche dealerships receive worldwide new corporate architecture ("Destination Porsche")*. 2019. URL: https://newsroom.porsche.com/en/2019/company/porsche-centres-worldwide-corporate-architecture-prototype-17258.html.

[129] Porsche AG. *Porsche Track Experience*. 2025. URL: https://experience.porsche.com/en/track/track-experience/about-track-experience.

[130] Ferrari S.p.A. *Ferrari XX Programme | Corse Clienti*. 2025. URL: https://www.ferrari.com/en-EN/corse-clienti/xx-programmes.

[131] Ferrari S.p.A. *Ferrari F1 Clienti | Corse Clienti*. 2025. URL: https://www.ferrari.com/en-EN/corse-clienti/f1-clienti.

[132] Ferrari S.p.A. *MyFerrari, Your World*. 2025. URL: https://www.ferrari.com/en-EN/auto/myferrari.

[133] McKinsey & Company. *Car leasing in Europe: Managing residual value for a 12 billion opportunity*. 2023. URL: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/car-leasing-in-europe-managing-residual-value-for-a-12-billion-euro-opportunity.

[134] iSeeCars. *The Top 25 Vehicles That Hold Their Value Best*. 2025. URL: https://www.iseecars.com/cars-that-hold-their-value-study.

[135] McKinsey & Company. *Five trends shaping tomorrow's luxury-car market.* 2022. URL: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/five-trends-shaping-tomorrows-luxury-car-market.

[136] Porsche Cars North America. *Porsche Approved: Certified Pre-Owned.* 2025. URL: https://www.porsche.com/usa/approved-used/.

[137] J.D. Power. *2024 U.S. ALG Residual Value Awards.* 2023. URL: https://www.jdpower.com/business/press-releases/2024-us-alg-residual-value-awards.

[138] McKinsey & Company. *Collectible cars: From niche market to growth and innovation engine.* 2025. URL: https://www.mckinsey.com.br/our-insights/collectible-cars-from-niche-market-to-growth-and-innovation-engine.

[139] Francesco Ricci, Lior Rokach, and Bracha Shapira. "Introduction to Recommender Systems Handbook". In: *Recommender Systems Handbook.* Springer, 2011, pp. 1–35. DOI: 10.1007/978-0-387-85820-3_1.

[140] Badrul Sarwar et al. "Item-based Collaborative Filtering Recommendation Algorithms". In: *Proceedings of the 10th International Conference on World Wide Web.* ACM, 2001, pp. 285–295. DOI: 10.1145/371920.372071.

[141] David Goldberg et al. "Using Collaborative Filtering to Weave an Information Tapestry". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 1992, pp. 61–70. DOI: 10.1145/142750.142754.

[142] Paul Resnick et al. "GroupLens: An Open Architecture for Collaborative Filtering of Netnews". In: *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work.* ACM, 1994, pp. 175–186. DOI: 10.1145/192844.192905.

[143] Greg Linden, Brent Smith, and Jeremy York. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering". In: *IEEE Internet Computing.* Vol. 7. 1. IEEE, 2003, pp. 76–80. DOI: 10.1109/MIC.2003.1167344.

[144] Yehuda Koren, Robert Bell, and Chris Volinsky. “Matrix Factorization Techniques for Recommender Systems”. In: *Computer* 42.8 (2009), pp. 30–37. DOI: 10.1109/MC.2009.263.

[145] Jonathan L. Herlocker et al. “Evaluating collaborative filtering recommender systems”. In: *ACM Transactions on Information Systems* 22.1 (2004), pp. 5–53. DOI: 10.1145/963770.963772.

[146] Sean M. McNee, John Riedl, and Joseph A. Konstan. “Being accurate is not enough: How accuracy metrics have hurt recommender systems”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI EA)*. ACM, 2006, pp. 1097–1101. DOI: 10.1145/1125451.1125659.

[147] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. “Performance of recommender algorithms on top-N recommendation tasks”. In: *Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys)*. ACM, 2010, pp. 39–46. DOI: 10.1145/1864708.1864721.

[148] Saul Vargas and Pablo Castells. “Rank and relevance in novelty and diversity metrics for recommender systems”. In: *Proceedings of the Fifth ACM Conference on Recommender Systems (RecSys)*. ACM, 2011, pp. 109–116. DOI: 10.1145/2043932.2043955.

[149] Gediminas Adomavicius and Alexander Tuzhilin. “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”. In: *IEEE Transactions on Knowledge and Data Engineering* 17.6 (2005), pp. 734–749. DOI: 10.1109/TKDE.2005.99.

[150] Robin Burke. “Hybrid Recommender Systems: Survey and Experiments”. In: *User Modeling and User-Adapted Interaction* 12.4 (2002), pp. 331–370. DOI: 10.1023/A:1021240730564.

[151] Steffen Rendle. “Factorization Machines”. In: *Proceedings of the 2010 IEEE International Conference on Data Mining*. IEEE, 2010, pp. 995–1000. DOI: 10.1109/ICDM.2010.127.

[152] Xiangnan He et al. “Neural Collaborative Filtering”. In: *Proceedings of the 26th International Conference on World Wide Web*. ACM, 2017, pp. 173–182. DOI: 10.1145/3038912.3052569.

[153] Balázs Hidasi et al. "Session-based Recommendations with Recurrent Neural Networks". In: *Proceedings of the 4th International Conference on Learning Representations*. 2016.

[154] Paul Covington, Jay Adams, and Emre Sargin. "Deep Neural Networks for YouTube Recommendations". In: *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 2016, pp. 191–198. DOI: 10.1145/2959100.2959190.

[155] Shuai Zhang et al. "Large Language Models for Recommender Systems: A Survey and Roadmap". In: *arXiv preprint arXiv:2305.19860* (2023).

[156] Wei Wei et al. "Generative Recommendation: Towards Next-generation Recommender Paradigm". In: *Proceedings of the 29th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2023, pp. 5463–5464. DOI: 10.1145/3580305.3599576.

[157] Chen Gao, Xiangnan He, and Yongfeng Zhang. "Recommendation with Generative Retrieval and Large Language Models". In: *arXiv preprint arXiv:2306.15292* (2023).

[158] Jiabin Mao et al. "Guardrails for Large Language Models in Recommendation". In: *Proceedings of the 30th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2024.

[159] Thorsten Schmidt, Anastasia Marbach, and Frank Mantwill. "Recommender Systems for Variant Management in the Automotive Industry". In: *Proceedings of the Design Society: International Conference on Engineering Design*. Cambridge University Press, 2022. DOI: 10.1017/pds.2022.123.

[160] Anastasia Mavridou, Dimitris Assimakopoulos, and Pericles Mitkas. "Mining Affective Needs of Automotive Industry Customers for Building a Mass-Customization Recommender System". In: *International Journal of Product Development* 13.2 (2011), pp. 108–126. DOI: 10.1504/IJPD.2011.042620.

[161] Wei Cai et al. "An Inspiration Recommendation System for Automotive Scenarios with Sparse Item Preference Data". In: *Systems* 12.11 (2024), p. 491. DOI: 10.3390/systems12110491.

[162] Ian Ang. *Understanding Customer Data with AI Recommender Systems in the Automotive Industry.* https://strathprints.strath.ac.uk/89349/. University of Strathclyde, Working Paper. 2023.

[163] Statworx. *Increasing In-Car Service Sales through a Personalized Recommendation System.* https://www.statworx.com/en/case-studies/increasing-in-car-service-sales-through-a-personalized-recommendation-system/. Accessed September 7, 2025. 2025.

[164] Porsche AG. *Porsche Adds a Recommendation System to Its Car Configurator.* https://recommender-systems.com/news/2021/02/04/porsche-adds-a-recommendation-system-to-its-car-configurator/. Accessed September 7, 2025. 2021.

[165] Renascence Consulting. *Customer Experience (CX) Design in Automotive Industry.* https://www.renascence.io/journal/customer-experience-cx-design-in-automotive-industry-examples-case-studies. Accessed September 7, 2025. 2024.

[166] Amazon Web Services. *Ferrari Advances Generative AI for Customer Experience.* https://aws.amazon.com/solutions/case-studies/ferrari-generative-ai-case-study/. Accessed September 7, 2025. 2025.

[167] Guibing Guo. "Resolving Data Sparsity by Similarity Fusion in Collaborative Filtering". In: *Proceedings of the 26th AAAI Conference on Artificial Intelligence*. AAAI, 2012, pp. 1719–1720. URL: https://www.ijcai.org/Proceedings/13/Papers/495.pdf.

[168] Erion Çano and Maurizio Morisio. "Hybrid Recommender Systems: A Systematic Literature Review". In: *Intelligent Data Analysis* 23.2 (2019), pp. 422–438. DOI: 10.3233/IDA-173957.

[169] Bihi Sabiri et al. "Hybrid Quality-Based Recommender Systems: A Systematic Literature Review". In: *Journal of Imaging* 11.1 (2025), p. 12. DOI: 10.3390/jimaging11010012. URL: https://www.mdpi.com/2313-433X/11/1/12.

[170] Steffen Rendle et al. "Revisiting the Performance of iALS on Item Recommendation Benchmarks". In: *Proceedings of the 16th ACM Conference on Recommender Systems*. 2022, pp. 427–435. DOI: 10.1145/3523227.3548486.

[171] Steffen Rendle et al. "Revisiting the Performance of iALS on Item Recommendation Benchmarks". In: *Proceedings of the 16th ACM Conference on Recommender Systems (RecSys '22)*. Seattle, WA, USA: ACM, 2022, pp. 427–435. DOI: 10.1145/3523227.3548486.

[172] Yifan Hu, Yehuda Koren, and Chris Volinsky. "Collaborative Filtering for Implicit Feedback Datasets". In: *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM '08)*. Pisa, Italy: IEEE, 2008, pp. 263–272. DOI: 10.1109/ICDM.2008.22.

[173] Takuya Akiba et al. "Optuna: A Next-Generation Hyperparameter Optimization Framework". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '19)*. Anchorage, AK, USA: ACM, 2019, pp. 2623–2631. DOI: 10.1145/3292500.3330701.

[174] Maurizio Ferrari Dacrema. "Demonstrating the Equivalence of List-Based and Aggregate Metrics to Measure the Diversity of Recommendations (Student Abstract)". In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI '21)*. Virtual Event: AAAI Press, 2021, pp. 15779–15780.

[175] Frederick F. Reichheld. *The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*. Harvard Business School Press, 1996.

[176] Sunil Gupta, Donald R. Lehmann, and Jennifer A. Stuart. "Valuing Customers". In: *Journal of Marketing Research* 41.1 (2004), pp. 7–18. DOI: 10.1509/jmkr.41.1.7.25084.

[177] Frederick F. Reichheld and W. Earl Sasser. "Zero Defections: Quality Comes to Services". In: *Harvard Business Review* 68.5 (1990), pp. 105–111.

[178] Claes Fornell. "A National Customer Satisfaction Barometer: The Swedish Experience". In: *Journal of Marketing* 56.1 (1992), pp. 6–21. DOI: 10.1177/002224299205600103.

[179] Robert M. Morgan and Shelby D. Hunt. “The Commitment-Trust Theory of Relationship Marketing”. In: *Journal of Marketing* 58.3 (1994), pp. 20–38. DOI: 10.1177/002224299405800302.

[180] Ruth N. Bolton. “A Dynamic Model of the Duration of the Customer’s Relationship with a Continuous Service Provider: The Role of Satisfaction”. In: *Marketing Science* 17.1 (1998), pp. 45–65. DOI: 10.1287/mksc.17.1.45.

[181] Aurélie Lemmens and Christophe Croux. “Bagging and Boosting Classification Trees to Predict Churn”. In: *Journal of Marketing Research* 43.2 (2006), pp. 276–286. DOI: 10.1509/jmkr.43.2.276.

[182] Scott A. Neslin et al. “Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models”. In: *Journal of Marketing Research* 43.2 (2006), pp. 204–211. DOI: 10.1509/jmkr.43.2.204.

[183] Kristof Coussement and Dirk Van den Poel. “Churn Prediction in Subscription Services: An Application of Support Vector Machines While Comparing Two Parameter-Selection Techniques”. In: *Expert Systems with Applications* 34.1 (2008), pp. 313–327. DOI: 10.1016/j.eswa.2006.09.038.

[184] Wouter Verbeke et al. “Building comprehensible customer churn prediction models with advanced rule induction techniques”. In: *Expert Systems with Applications* 39.17 (2012), pp. 12247–12263. DOI: 10.1016/j.eswa.2012.04.072.

[185] Youngjung Suh. “Machine learning based customer churn prediction in home appliance rental business”. In: *Journal of Big Data* 10.1 (2023), p. 41. DOI: 10.1186/s40537-023-00721-8.

[186] U. G. Joy et al. “A Big Data-Driven Hybrid Model for Enhancing Streaming Service Customer Retention Through Churn Prediction Integrated With Explainable AI”. In: *IEEE Access* 12 (2024). Open Access, includes SHAP/EBM explanations, pp. 69130–69150. DOI: 10.1109/ACCESS.2024.xxxxxxx.

[187] Meryem Chajia and El Habib Nfaoui. “Customer Churn Prediction Approach Based on LLM Embeddings and Logistic Regression”. In: *Future Internet* 16.12 (2024), p. 453. DOI: 10.3390/ïňĄ16120453.

[188] "Taco Bell and KFC's Owner Says AI-Driven Marketing Is Boosting Purchases". In: *The Wall Street Journal* (2024). URL: https://www.wsj.com/articles/taco-bell-and-kfcs-owner-says-ai-driven-marketing-is-boosting-purchases-ab3a5f36.

[189] "LVMH Bets on AI to Navigate Luxury Goods Slowdown". In: *The Wall Street Journal* (2024). URL: https://www.wsj.com/articles/lvmh-bets-on-ai-to-navigate-luxury-goods-slowdown-0438e328.

[190] "The trust recession: why customers don't trust AI (and how to fix it)". In: *TechRadar Pro* (2024). URL: https://www.techradar.com/pro/the-trust-recession-why-customers-dont-trust-ai-and-how-to-fix-it.

[191] Sobot AI. *AI Customer Service Case Studies: 2025 Support, Satisfaction, Cost.* Case studies citing retention, satisfaction, and efficiency improvements. 2025. URL: https://www.sobot.io/article/ai-customer-service-case-studies-2025-support-satisfaction-cost/.

[192] Jean-Noël Kapferer and Vincent Bastien. *The Luxury Strategy: Break the Rules of Marketing to Build Luxury Brands.* 2nd ed. Kogan Page, 2015.

[193] Deloitte. *Global Powers of Luxury Goods 2023: Engaging the Connected Luxury Consumer.* Deloitte Insights, 2023. URL: https://www2.deloitte.com/global/en/pages/consumer-business/articles/global-powers-of-luxury-goods.html.

[194] Pegah Barakati et al. "Luxury Car Data Analysis: A Literature Review". In: *Data* 9.4 (2024), p. 48. DOI: 10.3390/data9040048.

[195] Agustín J Sánchez-Medina, C Eleazar, et al. "Using machine learning and big data for efficient forecasting of hotel booking cancellations". In: *International Journal of Hospitality Management* 89 (2020), p. 102546.

[196] Muhammed Tekin and Murat Gök. "Performance comparison of classification algorithms in hotel booking cancellation prediction". In: *Artificial Intelligence Theory and Applications* 1.1 (2021), pp. 8–19.

[197] Yiying Chen et al. "Comparison and analysis of machine learning models to predict hotel booking cancellation". In: *2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*. Atlantis Press. 2022, pp. 1363–1370.

[198] Gintare Karolina Dziugaite, Shai Ben-David, and Daniel M Roy. "Enforcing interpretability and its statistical impacts: Trade-offs between accuracy and interpretability". In: *arXiv preprint arXiv:2010.13764* (2020).

[199] Bloomberg L.P. *Bloomberg - Financial News and Data*. `https://www.bloomberg.com`. Accessed: 2025-04-15. 2025.

[200] Xue-wen Chen and Jong Cheol Jeong. "Enhanced recursive feature elimination". In: *Sixth international conference on machine learning and applications (ICMLA 2007)*. IEEE. 2007, pp. 429–435.

[201] Streamlit Inc. *Streamlit: The fastest way to build data apps*. `https://streamlit.io`. Accessed: 2025-04-15. 2019.

[202] Aaron Grattafiori et al. "The llama 3 herd of models". In: *arXiv preprint arXiv:2407.21783* (2024).

[203] European union. "General data protection regulation". In: (2016).

[204] European Commission. *Regulation (EU) 2024/... of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. 2024.

[205] Apostolos Ampountolas. "Predicting hotel booking cancellations: a comprehensive machine learning approach". In: *Journal of Revenue and Pricing Management* (2025), pp. 1–12.

[206] B. Frieske and S. Stieler. "The "semiconductor crisis" as a result of the COVID-19 pandemic and its impacts on the automotive industry and its supply chains". In: *World Electric Vehicle Journal* 13.10 (2022). DOI: `10.3390/wevj13100189`.

[207] *Car leasing in Europe: Managing residual value for a €12 billion opportunity.* Accessed: 2025-09-07. 2025. URL: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/car-leasing-in-europe-managing-residual-value-for-a-12-billion-euro-opportunity.

[208] G. A. Akerlof. "The market for "lemons": Quality uncertainty and the market mechanism". In: *The Quarterly Journal of Economics* 84.3 (1970), pp. 488–500. DOI: 10.2307/1879431.

[209] I. Bauer, L. Zavolokina, and G. Schwabe. "Is there a market for trusted car data?" In: *Electronic Markets* 30.2 (2020), pp. 211–225. DOI: 10.1007/s12525-019-00368-5.

[210] A. Jerenz. *Revenue management and survival analysis in the automobile industry.* Springer, 2008. DOI: 10.1007/978-3-8349-9840-8.

[211] F. J. Fabozzi. "The fundamentals of equipment leasing". In: *Handbook of Finance.* Wiley, 2008, pp. 815–823. DOI: 10.1002/9780470404324.hof002079.

[212] *A virtuous cycle for top-line growth.* Accessed: 2025-09-07. 2015. URL: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/a-virtuous-cycle-for-top-line-growth.

[213] D. C. Rode, P. S. Fischbeck, and S. R. Dean. "Residual risk and the valuation of leases under uncertainty and limited information". In: *The Journal of Structured Finance* 7.4 (2002), pp. 37–49. DOI: 10.3905/jsf.2002.320265.

[214] S. R. Grenadier. "Leasing and credit risk". In: *Journal of Financial Economics* 42 (1996), pp. 333–364. DOI: 10.1016/0304-405X(96)00882-3.

[215] H. Pirotte and C. Vaessen. "Residual value risk in the leasing industry: A European case". In: *European Journal of Finance* 14.2 (2008), pp. 157–177. DOI: 10.1080/13518470701705637.

[216] T. Cerquitelli et al. "Data-driven estimation of heavy-truck residual value at the buy-back". In: *IEEE Access* 8 (2020), pp. 102409–102418. DOI: 10.1109/ACCESS.2020.2998940.

[217] S. M. Prado. “The European used-car market at a glance: Hedonic resale price valuation in the automotive leasing industry”. In: *Economics Bulletin* 29.3 (2009), pp. 2086–2099. URL: https://ideas.repec.org/a/ebl/ecbull/eb-09-00355.html.

[218] J. Du, L. Xie, and S. Schroeder. “PIN optimal distribution of auction vehicles system: Applying price forecasting, elasticity estimation, and genetic algorithms to used-vehicle distribution”. In: *Marketing Science* 28.4 (2009), pp. 637–644. DOI: 10.1287/mksc.1080.0470.

[219] J. D. Wu, C. C. Hsu, and H. C. Chen. “An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference”. In: *Expert Systems with Applications* 36.4 (2009), pp. 7809–7817. DOI: 10.1016/j.eswa.2008.11.019.

[220] J. Sharma and S. K. Mitra. “Developing a Used Car Pricing Model Applying Multivariate Adaptive Regression Splines Approach”. In: *Expert Systems with Applications* 236 (2024), p. 121277. DOI: 10.1016/j.eswa.2023.121277.

[221] A. Ranjith et al. “Residual value prediction in the Indian luxury pre-owned market using neural networks”. In: *International Journal of Automotive Technology* 236 (2024), p. 121277. DOI: 10.22214/ijraset.2024.63709.

[222] H. Huayi et al. “Residual Value Predictions”. In: *2022 IEEE International Conference on Computing (ICOCO)*. 2022. DOI: 10.1109/ICOCO56118.2022.10031995.

[223] H. Huayi et al. “Residual Value of Used Car Analysis and Prediction”. In: *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. 2023. DOI: 10.1109/ICECET58911.2023.10389355.

[224] M. Kim et al. “Trustworthy residual vehicle value prediction for auto finance”. In: *AI Magazine* 44.4 (2023), pp. 394–405. DOI: 10.1002/aaai.12136.

[225] L. J. Tashman. “Out-of-sample tests of forecasting accuracy: An analysis and review”. In: *International Journal of Forecasting* 16.4 (2000), pp. 437–450. DOI: 10.1016/S0169-2070(00)00065-0.

[226] S. M. Lundberg and S.-I. Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems* (2017). DOI: 10.48550/arXiv.1705.07874.

[227] S. Lessmann and S. Voß. "Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy". In: *International Journal of Forecasting* 33 (2017), pp. 864–877. DOI: 10.1016/j.ijforecast.2017.04.003.

[228] S. Bergmann and S. Feuerriegel. "Machine learning for predicting used car resale prices using granular vehicle equipment information". In: *Expert Systems with Applications* 263 (2025), p. 125640. DOI: 10.1016/j.eswa.2024.125640.

[229] Mehdi Motififard et al. "Demographic and Injury Characteristics as Potential Risk Factors for Anterior Cruciate Ligament Injuries: A Multicentric Cross-Sectional Study". In: *J. Clin. Med.* 13.17 (2024), p. 5063. DOI: 10.3390/jcm13175063.

[230] E. Carlos Rodriguez-Merchan and Carlos A. Encinas-Ullan. "Knee Osteoarthritis Following Anterior Cruciate Ligament Reconstruction: Frequency, Contributory Elements, and Recent Interventions to Modify the Route of Degeneration". In: *Arch. Bone Jt. Surg.* 10.11 (2022), pp. 951–958.

[231] Clare L. Ardern et al. "Fifty-five per cent return to competitive sport following anterior cruciate ligament reconstruction surgery: an updated systematic review and meta-analysis including aspects of physical functioning and contextual factors". In: *Br. J. Sports Med.* 48.21 (2014), pp. 1543–1552. DOI: 10.1136/bjsports-2013-093398.

[232] Luca Andriollo et al. "The Role of Artificial Intelligence in Anterior Cruciate Ligament Injuries: Current Concepts and Future Perspectives". In: *Healthcare (Basel)* 12.3 (2024), p. 300. DOI: 10.3390/healthcare12030300.

[233] Dongning Su et al. "Projections for prevalence of Parkinson's disease and its driving factors in 195 countries and territories to 2050: a modelling study of Global Burden of Disease Study 2021". In: *BMJ* 388 (2025), e080952. DOI: 10.1136/bmj-2024-080952.

[234] Laura Smith and Anthony H. V. Schapira. "GBA Variants and Parkinson Disease: Mechanisms and Treatments". In: *Cells* 11.8 (2022), p. 1261. DOI: 10.3390/cells11081261.

[235] Hai Li, Massimiliano Zecca, and Jiajun Huang. "Evaluating the Utility of Wearable Sensors for the Early Diagnosis of Parkinson Disease: Systematic Review". In: *J. Med. Internet Res.* 27 (2025), e69422. DOI: 10.2196/69422.

[236] *Diagnosis and Treatment of Parkinson Disease: A Review*. https://pubmed.ncbi.nlm.nih.gov/32044947/. PubMed; accessed 15 Dec 2023. 2023.

[237] Lorraine V. Kalia and Anthony E. Lang. "Parkinson's disease". In: *The Lancet* 386.9996 (2015), pp. 896–912. DOI: 10.1016/S0140-6736(14)61393-3.

[238] Chee Young Kim and Roy N. Alcalay. "Genetic forms of Parkinson's disease". In: *Seminars in Neurology* 37.2 (2017), pp. 135–146. DOI: 10.1055/s-0037-1601567.

[239] Francesco Cavallieri, Rodrigo G. Cury, Tainá Guimarães, et al. "Recent advances in the treatment of genetic forms of Parkinson's disease: hype or hope?" In: *Cells* 12.5 (2023), p. 764. DOI: 10.3390/cells12050764.

[240] Ajesh Cherian and Karthik P. Divya. "Genetics of Parkinson's disease". In: *Acta Neurologica Belgica* 120.6 (2020), pp. 1297–1305. DOI: 10.1007/s13760-020-01473-5.

[241] Joanna O. Day and Sonia Mullin. "The genetics of Parkinson's disease and implications for clinical practice". In: *Genes* 12.7 (2021), p. 1006. DOI: 10.3390/genes12071006.

[242] Cornelis Blauwendraat, Mike A. Nalls, and Andrew B. Singleton. "The genetic architecture of Parkinson's disease". In: *The Lancet Neurology* 19.2 (2020), pp. 170–178. DOI: 10.1016/S1474-4422(19)30287-X.

[243] Anna Westenberger, Veronika Skrahina, Tetiana Usnich, et al. "Relevance of genetic testing in the gene-targeted trial era: the Rostock Parkinson's disease study". In: *Brain* 147.8 (2024), pp. 2652–2667. DOI: 10.1093/brain/awae188.

[244] Matthew E. Gegg, Emily Menozzi, and Anthony H. V. Schapira. "Glucocerebrosidase-associated Parkinson disease: pathogenic mechanisms and potential drug treatments". In: *Neurobiology of Disease* 166 (2022), p. 105663. DOI: 10.1016/j.nbd.2022.105663.

[245] Valeria Yahya, Alessio Di Fonzo, and Edoardo Monfrini. "Genetic evidence for endolysosomal dysfunction in Parkinson's disease: a critical overview". In: *International Journal of Molecular Sciences* 24.7 (2023), p. 6338. DOI: 10.3390/ijms24076338.

[246] Zarazatz Gan-Or, Itay Amshalom, Lukasz L. Kilarski, et al. "Differential effects of severe vs mild GBA mutations on Parkinson disease". In: *Neurology* 84.9 (2015), pp. 880–887. DOI: 10.1212/WNL.0000000000001315.

[247] Selma C. Parlar et al. "Classification of GBA1 variants in Parkinson's disease: the GBA1-PD browser". In: *Movement Disorders* 38.3 (2023), pp. 489–495. DOI: 10.1002/mds.29314.

[248] Ernest Beutler, Tal Gelbart, and Charles R. Scott. "Hematologically important mutations: Gaucher disease". In: *Blood Cells, Molecules, and Diseases* 35.3 (2005), pp. 355–364. DOI: 10.1016/j.bcmd.2005.07.005.

[249] Giacomina M. Riboldi and Alberto B. Di Fonzo. "GBA, Gaucher disease, and Parkinson's disease: from genetic to clinic to new therapeutic approaches". In: *Cells* 8.4 (2019), p. 364. DOI: 10.3390/cells8040364.

[250] J. Zhang et al. "Prediction of Parkinson's disease using machine learning methods". In: *Biomolecules* 13.12 (2023), p. 1761. DOI: 10.3390/biom13121761.

[251] Yanqing Duan, John S. Edwards, and Yogesh K. Dwivedi. "Artificial intelligence for decision making in the era of big data–evolution, challenges and research agenda". In: *International Journal of Information Management* 48 (2019), pp. 63–71. DOI: 10.1016/j.ijinfomgt.2019.01.021.

[252] Maxwell W. Libbrecht and William S. Noble. "Machine learning applications in genetics and genomics". In: *Nature Reviews Genetics* 16.6 (2015), pp. 321–332. DOI: 10.1038/nrg3920.

[253] Abdullahi Tunde Aborode et al. "The role of machine learning in discovering biomarkers and predicting treatment strategies for neurodegenerative diseases: A narrative review". In: *NeuroMarkers* 2.1 (2025), p. 100034. ISSN: 2950-5887. DOI: https://doi.org/10.1016/j.neumar.2024.100034. URL: https://www.sciencedirect.com/science/article/pii/S295058872400034X.

[254] Mazen B. Makarious, Hannah L. Leonard, Daniele Vitale, et al. "Multimodality machine learning predicting Parkinson's disease". In: *NPJ Parkinson's Disease* 8.1 (2022). https://www.nature.com/articles/s41531-022-00288-w, p. 35. DOI: 10.1038/s41531-022-00288-w.

[255] Golnaz Hajianfar, Samira Kalayinia, and Mohammads Hosseinzadeh. "Prediction of Parkinson's disease pathogenic variants using hybrid machine learning systems and radiomic features". In: *Physica Medica* 113 (2023), p. 102647. DOI: 10.1016/j.ejmp.2023.102647.

[256] Viktoria Skrahina, Hannah Gaber, Eike J. Vollstedt, et al. "The Rostock international Parkinson's disease (ROPAD) study: protocol and initial findings". In: *Movement Disorders* 36.4 (2021), pp. 1005–1010. DOI: 10.1002/mds.28416.

[257] A. J. Hughes et al. "Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases". In: *Journal of Neurology, Neurosurgery*
Psychiatry 55.3 (1992), pp. 181–184. DOI: 10.1136/jnnp.55.3.181.

[258] World Medical Association. "World medical association declaration of Helsinki: ethical principles for medical research involving human subjects". In: *Journal of the American Medical Association* 310.20 (2013), pp. 2191–2194. DOI: 10.1001/jama.2013.281053.

[259] Irene Litvan, Jennifer G. Goldman, Alexander I. Tröster, et al. "Diagnostic criteria for mild cognitive impairment in Parkinson's disease: Movement Disorder Society task force guidelines". In: *Movement Disorders* 27.3 (2012), pp. 349–356. DOI: 10.1002/mds.24893.

[260] Murat Emre, Dag Aarsland, Richard Brown, et al. "Clinical diagnostic criteria for dementia associated with Parkinson's disease". In: *Movement Disorders* 22.12 (2007), pp. 1689–1707. DOI: 10.1002/mds.21507.

[261] Michael J. Sateia. "International classification of sleep disorders–third edition". In: *Chest* 146.5 (2014), pp. 1387–1394. DOI: 10.1378/chest.14-0970.

[262] Roberto Ceravolo et al. "Impulse control disorders in Parkinson's disease: definition, epidemiology, risk factors, neurobiology and management". In: *Parkinsonism and Related Disorders* 15 (2009), S111–S115. DOI: 10.1016/S1353-8020(09)70847-8.

[263] G. T. Stebbins et al. "How to identify tremor dominant and postural instability/gait difficulty groups with the movement disorder society unified Parkinson's disease rating scale: comparison with the unified Parkinson's disease rating scale". In: *Movement Disorders* 28.5 (2013), pp. 668–670. DOI: 10.1002/mds.25383.

[264] C. G. Goetz, B. C. Tilley, S. R. Shaftman, et al. "Movement Disorder Society–sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results". In: *Movement Disorders* 23.15 (2008), pp. 2129–2170. DOI: 10.1002/mds.22340.

[265] C. G. Goetz, W. Poewe, O. Rascol, et al. "Movement Disorder Society task force report on the Hoehn and Yahr staging scale: status and recommendations". In: *Movement Disorders* 19.9 (2004), pp. 1020–1028. DOI: 10.1002/mds.20213.

[266] Ziad S. Nasreddine, Natalie A. Phillips, Valerie Bédirian, et al. "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment". In: *Journal of the American Geriatrics Society* 53.4 (2005), pp. 695–699. DOI: 10.1111/j.1532-5415.2005.53221.x.

[267] Stef Van Buuren and Karin Groothuis-Oudshoorn. "MICE: Multivariate imputation by chained equations in R". In: *Journal of Statistical Software* 45 (2011), pp. 1–67. DOI: 10.18637/jss.v045.i03.

[268] Michael P. LaValley. “Logistic regression”. In: *Circulation* 117.18 (2008), pp. 2395–2399. DOI: 10.1161/CIRCULATIONAHA.106.682658.

[269] Shun-Ichi Amari. “Backpropagation and stochastic gradient descent method”. In: *Neurocomputing* 5.4 (1993), pp. 185–196. DOI: 10.1016/0925-2312(93)90006-O.

[270] Thomas T. Wong. “Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation”. In: *Pattern Recognition* 48.9 (2015), pp. 2839–2846. DOI: 10.1016/j.patcog.2015.03.009.

[271] Solveig T. Jost, Maria A. Kaldenbach, Angelo Antonini, et al. “Levodopa dose equivalency in Parkinson’s disease: updated systematic review and proposals”. In: *Movement Disorders* (2023). DOI: 10.1002/mds.29410.

[272] Jan Neumann, Johanna Bras, Emma Deas, et al. “Glucocerebrosidase mutations in clinical and pathologically proven Parkinson’s disease”. In: *Brain* 132.7 (2009), pp. 1783–1794. DOI: 10.1093/brain/awp044.

[273] Fabrizio Blandini et al. “Glucocerebrosidase mutations and synucleinopathies: toward a model of precision medicine”. In: *Movement Disorders* 34.1 (2019), pp. 9–21. DOI: 10.1002/mds.27583.

[274] Raffaele Cilia, Sara Tunesi, Giuseppe Marotta, et al. “Survival and dementia in GBA-associated Parkinson’s disease: the mutation matters”. In: *Annals of Neurology* 80.5 (2016), pp. 662–673. DOI: 10.1002/ana.24777.

[275] Ying Li, Takahito Sekine, Masaki Funayama, et al. “Clinicogenetic study of GBA mutations in patients with familial Parkinson’s disease”. In: *Neurobiology of Aging* 35.4 (2014), 935.e3–935.e8. DOI: 10.1016/j.neurobiolaging.2013.09.019.

[276] Samuel E. Winder-Rhodes, Jonathan R. Evans, Maria Ban, et al. “Glucocerebrosidase mutations influence the natural history of Parkinson’s disease in a community-based incident cohort”. In: *Brain* 136.2 (2013), pp. 392–399. DOI: 10.1093/brain/aws318.

[277] Katharina Brockmann, Kari Srulijes, Sandra Pflederer, et al. "GBA-associated Parkinson's disease: reduced survival and more rapid progression in a prospective longitudinal study". In: *Movement Disorders* 30.3 (2015), pp. 407–411. DOI: 10.1002/mds.26071.

[278] Ahsan Munappy et al. "Data management challenges for deep learning". In: *45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. 2019, pp. 140–147. DOI: 10.1109/SEAA.2019.00030.

[279] Anat Thaler, Tom Gurevich, Avi Bar Shira, et al. "A "dose" effect of mutations in the GBA gene on Parkinson's disease phenotype". In: *Parkinsonism and Related Disorders* 36 (2017), pp. 47–51. DOI: 10.1016/j.parkreldis.2016.12.014.

[280] Jing Ren, Guolin Zhou, Yong Wang, et al. "Association of GBA genotype with motor and cognitive decline in Chinese Parkinson's disease patients". In: *Frontiers in Aging Neuroscience* 15 (2023), p. 1091919. DOI: 10.3389/fnagi.2023.1091919.

[281] Anna A. Szwedo, Ingar Dalen, Kirsten F. Pedersen, et al. "GBA and APOE impact cognitive decline in Parkinson's disease: a 10-year population-based study". In: *Movement Disorders* 37.5 (2022), pp. 1016–1027. DOI: 10.1002/mds.28932.

[282] Thomas B. Stoker, Megan Camacho, Samuel Winder-Rhodes, et al. "Impact of GBA1 variants on long-term clinical progression and mortality in incident Parkinson's disease". In: *Journal of Neurology, Neurosurgery and Psychiatry* 91.7 (2020), pp. 695–702. DOI: 10.1136/jnnp-2020-322857.

[283] Sergió Jesús, Iria Huertas, Ignacio Bernal-Bernal, et al. "GBA variants influence motor and non-motor features of Parkinson's disease". In: *PLoS ONE* 11.12 (2016), e0167749. DOI: 10.1371/journal.pone.0167749.

[284] Tom Shiner, A. Mirelman, Mona Gana Weisz, et al. "High frequency of GBA gene mutations in dementia with Lewy bodies among Ashkenazi Jews". In: *JAMA Neurology* 73.12 (2016), pp. 1448–1453. DOI: 10.1001/jamaneurol.2016.1593.

[285] Roy N. Alcalay, Elizabeth Caccappolo, Hobson Mejia-Santana, et al. “Cognitive performance of GBA mutation carriers with early-onset PD: the CORE-PD study”. In: *Neurology* 78.18 (2012), pp. 1434–1440. DOI: 10.1212/WNL.0b013e318253d54b.

[286] Billy Creese et al. “Glucocerebrosidase mutations and neuropsychiatric phenotypes in Parkinson’s disease and Lewy body dementias: review and meta-analyses”. In: *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 177.2 (2018), pp. 232–241. DOI: 10.1002/ajmg.b.32549.

[287] Zarazatz Gan-Or, Nir Giladi, Uri Rozovski, et al. “Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset”. In: *Neurology* 70.24 (2008), pp. 2277–2283. DOI: 10.1212/01.wnl.0000304039.11891.29.

[288] Simona Lucarno et al. “Systematic Video Analysis of Anterior Cruciate Ligament Injuries in Professional Female Soccer Players”. In: *The American Journal of Sports Medicine* 49.7 (2021), pp. 1794–1802. DOI: 10.1177/03635465211008169.

[289] Francesco Della Villa et al. “Systematic video analysis of ACL injuries in professional male football (soccer): injury mechanisms, situational patterns and biomechanics study on 134 consecutive cases”. In: *British Journal of Sports Medicine* 54.23 (2020), pp. 1423–1432. DOI: 10.1136/bjsports-2019-101247.

[290] Ibrahim Tamimi et al. “A Prediction Model for Primary Anterior Cruciate Ligament Injury Using Artificial Intelligence”. In: *Orthopaedic Journal of Sports Medicine* 9.9 (2021), p. 23259671211027543. DOI: 10.1177/23259671211027543.

[291] Satu Jauhiainen et al. “Predicting ACL Injury Using Machine Learning on Data From an Extensive Screening Test Battery of 880 Female Elite Athletes”. In: *The American Journal of Sports Medicine* 50.11 (2022), pp. 2917–2924. DOI: 10.1177/03635465221112095.

[292] Haraldur B. Sigurðsson and Kristin Briem. “Cluster analysis successfully identifies clinically meaningful knee valgus moment patterns: frequency of early peaks reflects sex-specific ACL injury incidence”. In: *Journal of*

Experimental Orthopaedics 6.1 (2019), p. 37. DOI: 10.1186/s40634-019-0205-5.

[293] Christopher A. DiCesare et al. "Distinct Coordination Strategies Associated with the Drop Vertical Jump Task". In: *Medicine & Science in Sports & Exercise* 52.5 (2020), pp. 1088–1098. DOI: 10.1249/MSS.0000000000002235.

[294] Haraldur B. Sigurðsson et al. "Don't Peak Too Early: Evidence for an ACL Injury Prevention Mechanism of the 11+ Program". In: *International Journal of Sports Physical Therapy* 17.5 (2022), pp. 823–831. DOI: 10.26603/001c.36524.

[295] Byron Donaldson, Neil Bezodis, and Helen Bayne. "Characterising coordination strategies during initial acceleration in sprinters ranging from highly trained to world class". In: *Journal of Sports Sciences* 41.19 (2023), pp. 1768–1778. DOI: 10.1080/02640414.2023.2298100.

[296] Matthew B. Bird et al. "Unsupervised Clustering Techniques Identify Movement Strategies in the Countermovement Jump Associated With Musculoskeletal Injury Risk During US Marine Corps Officer Candidates School". In: *Frontiers in Physiology* 13 (2022), p. 868002. DOI: 10.3389/fphys.2022.868002.

[297] Jacob Rauch et al. "Different Movement Strategies in the Countermovement Jump Amongst a Large Cohort of NBA Players". In: *International Journal of Environmental Research and Public Health* 17.17 (2020), p. 6394. DOI: 10.3390/ijerph17176394.

[298] Attila Schulc et al. "Identifying Anterior Cruciate Ligament Injuries Through Automated Video Analysis of In-Game Motion Patterns". In: *Orthopaedic Journal of Sports Medicine* 12.3 (2024), p. 23259671231221579. DOI: 10.1177/23259671231221579.

[299] Jason Corban et al. "Artificial Intelligence in the Management of Anterior Cruciate Ligament Injuries". In: *Orthopaedic Journal of Sports Medicine* 9.7 (2021), p. 23259671211014206. DOI: 10.1177/23259671211014206.

[300] Zaixu Cui and Gaolang Gong. "The effect of machine learning regression algorithms and sample size on individualized behavioral prediction with functional connectivity features". In: *NeuroImage* 178 (2018), pp. 622–637. DOI: 10.1016/j.neuroimage.2018.06.001.

[301] Daniyal Rajput, Wei-Jen Wang, and Chun-Chuan Chen. "Evaluation of a decided sample size in machine learning applications". In: *BMC Bioinformatics* 24.1 (2023), p. 48. DOI: 10.1186/s12859-023-05156-9.

[302] Francesco Della Villa et al. "Kinematics of 90 change of direction in young football players: Insights for ACL injury prevention from the CUTtheACL study on 6008 trials". In: *Knee Surgery, Sports Traumatology, Arthroscopy* 32.10 (2024), pp. 2666–2678. DOI: 10.1002/ksa.12230.

[303] Martin Wattenberg, Fernanda Viégas, and Ian Johnson. "How to Use t-SNE Effectively". In: *Distill* (Oct. 2016). DOI: 10.23915/distill.00002. URL: https://distill.pub/2016/misread-tsne/.

[304] K. Sasirekha and P. Baby. "Agglomerative Hierarchical Clustering Algorithm-A Review". In: 2013. URL: https://api.semanticscholar.org/CorpusID:1359166.

[305] Marcel R. Ackermann et al. "Analysis of Agglomerative Clustering". In: *Algorithmica* 69.1 (May 2014), pp. 184–215. DOI: 10.1007/s00453-012-9717-4.

[306] K. Chidananda Gowda and G. Krishna. "Agglomerative clustering using the concept of mutual nearest neighbourhood". In: *Pattern Recognition* 10.2 (Jan. 1978), pp. 105–112. DOI: 10.1016/0031-3203(78)90018-3.

[307] Manoranjan Dash and Huan Liu. "Efficient hierarchical clustering algorithms using partially overlapping partitions: 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2001". In: *Advances in Knowledge Discovery and Data Mining, 5th Pacific-Asia Conference, PAKDD 2001, Proceedings*. Springer, 2001, pp. 495–506. DOI: 10.1007/3-540-45357-1_52.

[308] Fionn Murtagh and Pierre Legendre. “Ward’s Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm”. In: *Journal of Classification* 31.3 (Oct. 2014), pp. 274–295. DOI: 10.1007/s00357-014-9161-z.

[309] Peter J. Rousseeuw. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of Computational and Applied Mathematics* 20 (Nov. 1987), pp. 53–65. DOI: 10.1016/0377-0427(87)90125-7.

[310] Alberto Costa and Giacomo Nannicini. “RBFOpt: An open-source library for black-box optimization with costly function evaluations”. In: *Mathematical Programming Computation* 10.4 (2018), pp. 597–629. DOI: 10.1007/s12532-017-0131-5.