https://nmrj.ui.ac.ir/
New Marketing Research Journal
E-ISSN: 2228- 7744
Vol. 16, Issue 2, No.61, 2026
Document Type: Research Paper
Received: 11/02/2025 Accepted: 14/06/2026

Predicting Customer Satisfaction in Online Business-to-Business (B2B) Wholesaler: An Ensemble with Genetic Algorithm (GA) Tuning and SHAP Explainability

Amir Hosein Esmaielpour
Ph.D. student, Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran
Amir.esmaielpour@khu.ac.ir

Maryam Ameli *
Assistant professor, Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran
m.ameli@khu.ac.ir

Ehsan Golchin
M.Sc. student, Department of Business Management, Faculty of Management, Kharazmi University, Tehran, Iran
Ehsangolchin@khu.ac.ir

Abstract

The rapid growth of Business-to-Business (B2B) e-commerce platforms necessitates robust, data-driven frameworks for anticipating and managing customer experiences. This study addressed the challenge of predicting customer satisfaction within an online industrial B2B wholesaler in Iran by proposing an optimized machine learning ensemble approach. Using an operational dataset comprising 349 distinct orders, we formulated the prediction task as a multi-class ordinal classification problem with 3 categories: *Satisfied*, *Neutral*, and *Dissatisfied*. A majority-voting ensemble architecture was developed, integrating 3 diverse base learners—CatBoost, Random Forest (RF), and a Multi-Layer Perceptron (MLP) neural network. To maximize predictive accuracy, the hyperparameters of each base algorithm were independently tuned by using a Genetic Algorithm (GA). Empirical evaluation on a held-out validation set showed that the proposed GA-tuned majority-voting ensemble achieved outstanding performance with an Accuracy of 0.9575, Recall of 0.9476, F1-score of 0.9379, and Matthews Correlation Coefficient (MCC) of 0.8966. Moreover, an equal-weight TOPSIS evaluation across multiple metrics ranked the ensemble model first among all alternative configurations. To address the "black-box" nature of the ensemble, a SHAP explainability analysis was conducted, revealing that operational parameters—such as *Days Since Last Purchase* and indicators of basket volume and weight—served as the

*Corresponding author
Esmaielpour, A. H. , Ameli, M. and Golchin, E. (2026). Predicting Customer Satisfaction in Online B2B wholesaler : An Ensemble with Genetic Algorithm Tuning and SHAP Explainability. *New Marketing Research Journal*, 16 (2), 95 - 124 .

10.22108/nmrj.2026.144327.3159

primary drivers of satisfaction, while demographic attributes and purely financial variables exhibited negligible predictive power. This scalable pipeline provides actionable strategic insights for wholesale managers to implement proactive customer retention interventions.

Keywords*:* Customer Satisfaction Prediction, Machine Learning, Genetic Algorithm (GA), Consumer Behavior Analysis, E-Commerce.

Introduction

In the contemporary digital commerce paradigm, understanding and anticipating customer satisfaction is widely recognized as a cornerstone of corporate longevity, market sustainability, and effective customer relationship management. Within the Business-to-Business (B2B) e-commerce sector, proliferation of digital storefronts and wholesale platforms has dramatically intensified market competition. Unlike Business-to-Consumer (B2C) dynamics, B2B e-commerce involves highly complex operational pipelines characterized by large order volumes, high financial stakes, specialized transaction conditions, and extended procurement cycles. Consequently, the quality of the customer experience delivered by a wholesaler extends far beyond commodity pricing, encompassing multifaceted operational dimensions, such as delivery timeliness, packaging durability, shipping accuracy, logistical transparency, and after-sales service responsiveness.

Despite the critical importance of these factors, traditional B2B operations often rely on retrospective feedback mechanisms—such as periodic phone surveys or post-delivery online ratings—to gauge client sentiment. These reactive approaches inherently limit the ability of an organization to implement timely, mitigating interventions before client dissatisfaction escalates into churn. To address this limitation, contemporary academic literature has shifted from theoretical, post-hoc explanations of customer loyalty toward proactive, data-driven predictive modeling enabled by machine learning and deep learning algorithms. By leveraging real-time transactional, behavioral, and demographic features, predictive pipelines can identify vulnerable client accounts before formal grievances are recorded.

However, deploying predictive models within emerging markets, such as Iran, introduces distinct macro-environmental complexities. Wholesale enterprises operating in Iran contend with macroeconomic volatility, supply chain disruptions, fluctuating regulatory frameworks, and unique localized purchasing behaviors. In this context, predicting customer sentiment requires robust analytical models capable of extracting subtle behavioral patterns from relatively small, heterogeneous datasets without succumbing to overfitting. Although sophisticated deep learning architectures typically demand large datasets for effective generalization, machine learning ensembles and intelligent optimization algorithms—such as Genetic Algorithms (GAs)—offer a mathematically rigorous and computationally efficient alternative for modeling localized operational realities.

This research directly addressed a notable gap in the existing literature: the absence of tailored, explainable machine learning frameworks for predicting multi-class customer satisfaction in small-sample B2B online wholesale environments operating under economic constraints. The primary objective of this study was to construct, optimize, and evaluate a comprehensive machine learning architecture capable of predicting customer satisfaction prior to the receipt of formal feedback. By validating this model on operational data from an industrial B2B wholesaler in Iran, this study established a portable framework that balanced high predictive accuracy with clinical interpretability. In doing so, it would equip corporate decision-makers with the analytical infrastructure needed to optimize resource allocation, enhance logistical workflows, and systematically foster long-term B2B client loyalty.

Materials & Methods

The methodological workflow of this study followed a highly structured and scientifically rigorous pipeline comprising data acquisition, preprocessing, architectural design of the ensemble model, evolutionary hyperparameter optimization, and multi-criteria evaluation. The empirical foundation of this research was built upon an operational dataset of 349 unique customer orders extracted from the Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) databases of Arman Roshan Fartak, a prominent industrial wholesaler specializing in industrial adhesives in Iran. The target variable representing customer

satisfaction was meticulously collected through a synchronized dual-channel feedback mechanism that integrated systematic online post-purchase questionnaires with follow-up telephonic surveys conducted by specialized account managers. To align with the internal evaluation metrics of organizations, the target variable was encoded as a multi-class ordinal variable with 3 distinct levels: Class 0 ("Satisfied" clients), Class 1 ("Neutral" clients), and Class 2 ("Dissatisfied" clients).

The predictive feature space encompassed 3 primary dimensions: procurement demographics, purchase volume and value indicators, and relational interaction signals. Specifically, the features included: the gender of the corporate purchasing officer, the age of the purchasing officer, the geographical location (categorically binned into metropolitan hubs versus regional municipalities), the cumulative financial expenditure incurred by the client up to the point of the current transaction ("Total Spend"), the aggregate physical weight of all products purchased ("KG"), the commission structure applied to the sales agent ("Commission"), the exact number of days elapsed since the client's immediately preceding transaction ("Days Since Last Purchase"), and the corporate legal status of the client enterprise (private sector versus public or governmental entity).

The data preprocessing phase was executed with utmost care to ensure data integrity and prevent any form of data leakage. Of the initial 349 records, a small subset of 3 records (approximately 0.86%) exhibiting systematic and unrecoverable omissions in critical operational fields was entirely removed from the sample, yielding a final analytical dataset of 346 records. Categorical variables were transformed by using standard label encoding and one-hot encoding frameworks. To accommodate the specific structural requirements of the neural network component, continuous numerical variables were standardized by using a StandardScaler protocol, mapping the data to a mean of zero and a standard deviation of 1, thereby eliminating scaling biases. To establish a rigorous defense against overfitting and ensure robust generalization given the limited sample size, all subsequent training and optimization steps were embedded within a stratified 5-fold cross-validation framework, maintaining exact class proportions across all validation folds.

The core predictive engine consisted of a heterogeneous majority-voting ensemble architecture designed to leverage the complementary mathematical strengths of 3 distinct base learners: CatBoost, Random Forest (RF), and a Multi-Layer Perceptron (MLP). CatBoost, a symmetric gradient-boosting tree framework, was selected for its inherent mathematical superiority in processing heterogeneous tabular data and its algorithmic defenses against target leakage. Random Forest was incorporated as a robust, non-linear bagging baseline known for its low variance and resilience to localized noise. The MLP neural network was integrated to capture high-dimensional, non-linear interactions within the feature space. Rather than relying on computationally blind hyperparameter selection methods, such as Grid Search or Random Search, this study deployed an evolutionary GA to independently discover the optimal hyperparameter configurations for each base learner.

The GA optimization employed a population size of 20 chromosomes over a maximum of 40 generations with a tournament selection size of 3, a crossover probability of 0.85, a mutation probability of 0.10, and an early stopping criterion set to 10 consecutive generations without fitness improvement. The fitness function guiding the evolutionary search was defined as the stratified cross-validated weighted F1-score. To aggregate the individual base model outputs into a single final prediction, a hard voting (majority voting) protocol was chosen over soft voting following a rigorous preliminary sign test, which confirmed that hard voting offered superior mathematical stability and lower susceptibility to uncalibrated probability estimates given the sample size of this study.

Research Findings

The empirical results derived from the cross-validated training and testing phases demonstrated the substantial performance advantages achieved by integrating evolutionary optimization with an ensemble architecture. During the baseline evaluation of individual uncombined learners, CatBoost emerged as the most powerful standalone algorithm, exhibiting superior performance across nearly all primary evaluation metrics compared to RF and the MLP network. However, following the implementation of the hard-voting aggregation protocol, the finalized majority-voting ensemble delivered the most balanced, robust, and highly accurate performance, establishing its statistical dominance. On the held-out validation datasets, the proposed ensemble achieved an overall accuracy of 0.9575, a precision of 0.9396, a recall of 0.9476, and a weighted F1-score of 0.9379. To validate classification performance beyond simple accuracy, we computed advanced metrics, yielding a

Matthews Correlation Coefficient (MCC) of 0.8966 and a Cohen's Kappa (CK) coefficient of 0.8957. These values confirmed that the model's predictive power was highly stable and entirely independent of any underlying class imbalances within the operational dataset.

To mathematically validate the superiority of the ensemble approach without relying on subjective, single-metric interpretations, a Multi-Criteria Decision-Making (MCDM) evaluation was performed using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The evaluation matrix integrated 6 core performance dimensions—Accuracy, Precision, Recall, F1-score, MCC, and CK—calculated across all models. To ensure completely unbiased results, an equal-weighting scheme was applied across all criteria. The calculated TOPSIS closeness coefficients (Ci*Ci*), which represented each model's Euclidean proximity to the mathematically ideal solution and distance from the anti-ideal solution, explicitly yielded the following ranking: Ensemble (C=0.9695*C*=0.9695, Rank 1) > CatBoost (C=0.9187*C*=0.9187, Rank 2) > Random Forest (C=0.3437*C*=0.3437, Rank 3) > Multi-Layer Perceptron (C=0.1037*C*=0.1037, Rank 4). This rigorous multi-criteria analysis confirmed that the ensemble architecture systematically minimized classification errors when evaluated across all operational performance indicators simultaneously.

An in-depth analysis of the classification mechanics was conducted by inspecting the cross-validated confusion matrices for all four model variants. The ensemble architecture demonstrated an extraordinary capacity to identify vulnerable client segments, achieving a true positive count of 25 for Class 2 ("Dissatisfied")—meaning it attained a perfect 100% detection rate for dissatisfied accounts with zero false negatives in this critically important operational category. For Class 1 ("Neutral"), the ensemble maintained high precision, misclassifying only 2 out of 36 instances as "Satisfied". For Class 0 ("Satisfied"), the model exhibited a minimal margin of error with only a single instance incorrectly flagged as "Neutral". In contrast, the standalone MLP showed severe confusion, frequently failing to differentiate the subtle boundaries separating Class 1 and Class 2 instances.

To unlock the "black-box" architecture of the optimized ensemble and extract actionable strategic insights, we applied SHAP (SHapley Additive exPlanations) values, which were grounded in cooperative game theory. The global feature importance hierarchy was established by calculating the mean absolute SHAP values averaged comprehensively across all three classification classes. The resulting SHAP summary plot revealed an extremely pronounced skew in predictive influence. The behavioral variable of "Days Since Last Purchase" emerged as the dominant driver of customer satisfaction predictions, achieving a mean |SHAP||*SHAP*| value exceeding 1.6—a magnitude that mathematically quadrupled the influence of any other feature in the space. The volume and value metrics, namely "Total Spend" and the total physical weight of purchased goods ("KG"), exhibited moderate predictive influence acting as secondary drivers. Crucially, purely financial variables—such as the agent's "Commission"—along with all consumer demographic attributes—including the purchasing officer's age, gender, geographic city classification, and organizational company type—displayed negligible SHAP values, exerting almost no discernible impact on the model's ultimate classification decisions.

Discussion of Results & Conclusion

The empirical insights generated by this research carry both profound theoretical value for the domain of digital B2B marketing and immediate, actionable utility for operational management within wholesale enterprises. By demonstrating that a heterogeneous ensemble engine, optimized via evolutionary GAs, achieved an Accuracy of 95.75% and an MCC of 0.8966 on a compact operational dataset, this study successfully refuted the conventional assumption that sophisticated machine learning modeling is reserved exclusively for enterprises possessing massive, multi-million-record big data repositories. The perfect detection rate achieved by the ensemble for dissatisfied accounts (Class 2) held profound managerial significance: it proved that operational data natively generated during routine wholesale transactions contained highly structured, predictive behavioral signals that could be systematically exploited to deploy preemptive customer recovery strategies before relationship dissolution occurred.

When contextualized within contemporary academic literature, the findings of this study provided both crucial validations and distinct points of divergence. In contrast to the prominent research conducted by Zaghloul et al. (2024), which utilized an immense dataset exceeding 100,000 e-commerce orders to predict customer

satisfaction via standalone RF and Support Vector Machine (SVM) models—achieving an accuracy of approximately 92%—our framework demonstrated that an evolutionarily tuned ensemble architecture could extract significantly higher predictive accuracy (95.75%) and a vastly superior multi-criteria balance (validated via TOPSIS) from a small data footprint (n=346*n*=346). Furthermore, our SHAP-based feature importance hierarchy strongly reinforced the empirical conclusions of Tang (2025) and Hui et al. (2025), who argued that within B2B industrial environments, operational logistics signals, transaction recency, and purchase volume metrics carry exponentially greater predictive weight regarding customer sentiment than traditional demographic segmentation or basic pricing mechanisms.

The extraordinary dominance of the feature of "Days Since Last Purchase" within our SHAP explainability pipeline underscored a vital behavioral reality in the Iranian industrial wholesale sector. In this context, an increasing temporal gap between orders did not merely signify a passive pause in purchasing; rather, owing to intense competition and macro-environmental supply disruptions, it served as an active, early-warning indicator of client alienation, low engagement, or silent switching to alternative suppliers. Conversely, the moderate significance of "KG" (physical weight) and "Total Spend" confirmed that clients who consolidated large, heavy industrial orders established deeper, more integrated operational ties with the wholesaler, rendering them highly sensitive to logistics execution quality, packaging reliability, and delivery schedules. The finding that demographics and commission structures exerted almost zero predictive influence sent a strong signal to wholesale executives: broad financial discounting policies or rigid demographic marketing campaigns were fundamentally ineffective levers for securing long-term loyalty. Instead, corporate interventions had to be aggressively redirected toward recency-based customer activation systems and the systematic optimization of fulfillment workflows.

In conclusion, this study successfully established an optimized, highly accurate, and mathematically explainable machine learning pipeline for predicting customer satisfaction within the online B2B industrial wholesale sector. By fusing CatBoost, RF, and MLP models through a hard-voting ensemble, and optimizing their internal structures via a G A, the framework provided a highly reliable predictive tool that remained fully transparent through SHAP value interpretation. From a managerial perspective, the model offered an analytical blueprint for shifting corporate strategies from reactive grievance resolution to proactive client retention, prioritizing interaction recency and fulfillment reliability over superficial financial adjustments.

However, certain structural limitations must be acknowledged. This study relied on a compact, single-organization dataset from an industrial setting in Iran, which might limit the direct generalizability of the exact parameter values to substantially different economic sectors. Additionally, the classification pipeline employed a standard multi-class formulation without deploying specialized ordinal classification algorithms and the feature space lacked micro-level logistical timestamps and textual client reviews. To address these constraints, future research trajectories should focus on: (1) expanding data collection to cross-organizational, multi-sector wholesale databases to validate model generalizability; (2) implementing dedicated ordinal machine learning models to capture the intrinsic mathematical progression of satisfaction levels; (3) enriching the feature space by integrating automated Natural Language Processing (NLP) text-mining models applied to raw telephonic and digital customer review transcripts; and (4) executing randomized, controlled field pilots to empirically measure the financial and operational Return On Investment (ROI) of SHAP-guided, proactive retention interventions.

تحقیقات بازاریابی نوین
سال شانزدهم، شماره دوم، شماره پیاپی (۶۱) ۱۴۰۵
تاریخ دریافت: ۱۴۰۳/۱۱/۲۳ تاریخ پذیرش: ۱۴۰۵/۳/۲۴
ص: ۱۲۴- ۹۵

مقاله پژوهشی

پیش‌بینی رضایت مشتری در عمده‌فروشی B2B آنلاین: رویکرد گروهی با تنظیم ابرپارامتر الگوریتم ژنتیک و تبیین‌پذیری SHAP

امیرحسین اسمعیل‌پور [۱] ، مریم عاملی [۲] * ، احسان گلچین [۳]

۱- دانشجوی دکتری، گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه خوارزمی، تهران، ایران
Amir.esmaielpour@khu.ac.ir
۲- استادیار گروه مهندسی صنایع، دانشکده فنی و مهندسی، دانشگاه خوارزمی، تهران، ایران
m.ameli@khu.ac.ir
۳- دانشجوی کارشناسی ارشد، گروه مدیریت بازرگانی، دانشکده مدیریت، دانشگاه خوارزمی، تهران، ایران
Ehsangolchin@khu.ac.ir

چکیده

با شتاب‌گیری تجارت الکترونیک، پیش‌بینی رضایت مشتری برای تصمیم‌گیری‌های داده‌محور ضروری است. این پژوهش با تکیه بر داده‌های ۳۴۹ سفارش از یک عمده‌فروشی آنلاین صنعتی در ایران، مسئله را به‌صورت طبقه‌بندی چندکلاسه (ماهیت ترتیبی) صورت‌بندی کرده و یک چارچوب گروهی رأی‌گیری اکثریت (Majority Voting) بر فراز سه مدل CatBoost، جنگل تصادفی (RF) و شبکهٔ عصبی چندلایه (MLP) ارائه می‌کند. ابرپارامترهای (Hyperparameters) هر مدل به‌طور جداگانه با الگوریتم ژنتیک تنظیم و ارزیابی بر مبنای داده‌های تفکیک‌شده و معیارهای استاندارد انجام شده است. نتایج، نشان می‌دهد مدل گروهی رأی‌گیری اکثریت در مقایسه با بهترین مدل منفرد، عملکرد متوازن‌تری دارد؛ به‌طور مشخص به Accuracy = 0.9575 دست یافته و در ارزیابی TOPSIS با وزن‌های برابر برای مقایسهٔ چهار مدل پیشنهادی، رتبهٔ نخست را کسب کرده است. تحلیل توضیح‌پذیری با میانگین قدرمطلق مقادیر SHAP در کل کلاس‌ها نشان می‌دهد «روزهای گذشته از آخرین خرید» و شاخص‌های ارزش/حجم سبد (مانند وزن کل خرید) مهم‌ترین محرک‌های پیش‌بینی رضایت/نارضایتی هستند، درحالی‌که متغیرهای صِرف مالی (نظیر کمیسیون) و جمعیت‌شناختی (دموگرافیک) نقش کم‌اثرتری دارند. چارچوب پیشنهادی -با اتکا به ترکیب مدل‌ها، تنظیم ژنتیکی و تبیین‌پذیری- انتقال‌پذیر به سایر بافت‌های عمده‌فروشی آنلاین است و به‌عنوان ابزاری کاربردی برای بهبود رضایت مشتری پیشنهاد می‌شود.

کلید واژه‌ها: پیش‌بینی رضایت مشتری، یادگیری ماشین، الگوریتم ژنتیک، تحلیل رفتار مصرف‌کننده، تجارت الکترونیک.

---

* نویسنده مسؤول

اسمعیل‌پور، امیرحسین، عاملی، مریم و گلچین، احسان . (۱۴۰۵). پیش‌بینی رضایت مشتری در عمده‌فروشی B2B آنلاین: رویکرد گروهی با تنظیم فراپارامتر الگ وریتم ژنتیک و تبیین‌پذیری SHAP ، *تحقیقات بازاریابی نوین*، ۱۶ (۲)، ۹۵- ۱۲۴ .

10.22108/nmrj.2026.144327.3159

۱- مقدمه

در عصر حاضر، رضایت مشتری به‌عنوان یکی از کلیدی‌ترین شاخص‌های موفقیت کسب‌وکارها مورد توجه قرار گرفته است. این رضایت نه‌تنها به حفظ مشتریان و افزایش وفاداری آن‌ها کمک می‌کند، بلکه تأثیر مستقیمی بر سودآوری و رشد سازمان‌ها دارد (Obafemi et al., 2023). با گسترش فضای دیجیتال و ظهور تجارت الکترونیک، تحلیل و درک رفتار مشتریان از اهمیت بیشتری برخوردار شده است. در این میان، عمده‌فروشی‌های آنلاین به‌دلیل پیچیدگی‌های خاص، مانند مدیریت حجم زیاد سفارش‌ها، رقابت شدید، و نیاز به ارائهٔ تجربه‌ای متمایز و مثبت برای مشتریان، با چالش‌های بیشتری مواجه‌اند (Quintus et al., 2024).

یکی از دلایل اصلی تأکید بر رضایت مشتری در عمده‌فروشی آنلاین، نقش آن در ایجاد مزیت رقابتی است. در شرایط جهانی‌سازی و افزایش دسترسی مشتریان به گزینه‌های متنوع، تجربهٔ مشتری به یک عامل تعیین‌کننده در انتخاب و وفاداری آن‌ها تبدیل شده است. این تجربه شامل عواملی، مانند سرعت تحویل، کیفیت بسته‌بندی، هزینهٔ حمل‌ونقل، و سطح تعاملات انسانی در خدمات مشتری می‌شود؛ ازاین‌رو، شرکت‌ها نیازمند ابزارهای دقیق و پیشرفته‌ای برای پیش‌بینی رفتار و رضایت مشتریان هستند تا بتوانند راهبردهای خود را بهینه‌سازی و در بازارهای رقابتی جایگاه خود را حفظ کنند (Chen et al., 2018). با شتاب‌گیری پژوهش‌های داده‌محور، تمرکز مبانی از تبیین نظری رضایت به پیش‌بینی رضایت مشتری با تکیه بر یادگیری ماشین و عمیق جابه‌جا شده است. مطالعات اخیر نشان می‌دهند که ترکیب داده‌های رفتاری و تراکنشی با مدل‌های کلاسیک و عمیق -از RF و SVM تا MLP و رویکردهای چندمدلی- می‌تواند امتیاز یا سطح رضایت را با دقت زیاد پیش‌بینی کند و برای تصمیم‌سازی مدیریتی به کار رود (Zaghloul et al., 2024). در این میان، بازخوردهای آنلاین (امتیازها و متن مرورها)، سیگنال‌های غنی برای مدل‌های پیش‌بینی‌اند. پژوهش‌های جدید از تحلیل احساس و ترنسفورمرها تا چارچوب‌های ترکیبی، نشان داده‌اند که گنجاندن متن نظر مشتری به‌طور معناداری خطای پیش‌بینی رضایت را کاهش می‌دهد (Bellar et al., 2024). برای توضیح‌پذیری و استخراج محرک‌های کلیدی رضایت نیز به‌صورت گسترده به کار می‌رود؛ حتی ترکیب‌هایی مانند RFE-SHAP و مدل‌سازی (LDA) گزارش شده که دقت پیش‌بینی و تبیین اثر ویژگی‌ها بر رضایت را بهبود می‌دهد (Lamane et al., 2025).

باوجود این پیشرفت‌ها، بستر آنلاین با چالش‌هایی نظیر رقابت شدید، نبود شفافیت اطلاعات، و نگرانی‌های امنیتی و حریم خصوصی روبه‌روست که مستقیماً بر اعتماد مشتری و در نتیجه بر رضایت اثر می‌گذارد. مرورها و مطالعات بین‌بازاری جدید، نقش شفافیت، کیفیت خدمت و امنیت داده را برای حفظ اعتماد و رضایت تأیید می‌کنند (Quintus et al., 2024)؛ بنابراین، به‌کارگیری فرایندهای (پایپ‌لاین‌های) تحلیلی پیشرفته (ترکیب داده‌های تراکنشی و متنی، مدل‌های گروهی/عمیق، و ارزیابی دقیق) همراه با تبیین‌پذیری، مسیر مطمئنی را برای تصمیم‌گیری آگاهانه و بهبود پایدار تجربهٔ مشتری فراهم می‌کند (GhorbanTanhaei et al., 2024).

در ایران، بازارهای آنلاین با مشکلات

منحصربه‌فردی مواجه‌اند که بخشی از آن‌ها ریشه در فرهنگ خرید محلی، زیرساخت‌های فناوری و شرایط اقتصادی دارد (Emami et al., 2023). فروشندگان عمده در ایران، به‌ویژه در حوزه‌های صنعتی و تجاری، علاوه‌بر رقابت داخلی، با مسائلی همچون تغییرات مداوم قوانین و مقررات، نوسانات ارزی و اختلال در زنجیرهٔ تأمین مواجه‌اند (Abedini et al., 2025; Nikghadam & Reza Rabi, 2013). ازسوی‌دیگر، مشتریان ایرانی معمولاً به شفافیت بیشتر اطلاعات محصولات، زمان تحویل دقیق و هزینه‌های معقول حمل‌ونقل اهمیت می‌دهند. این ویژگی‌ها، پیش‌بینی رفتار و رضایت مشتری را دشوارتر کرده و ضرورت بهره‌گیری از ابزارهای پیشرفتهٔ تحلیل داده را برای شناسایی الگوهای رفتاری و ارتقای تجربهٔ مشتریان دوچندان می‌سازد.

در نهایت، با رشد روزافزون تجارت الکترونیک، تمرکز بر درک رفتار مشتریان و پیش‌بینی رضایت آن‌ها به نیازی حیاتی برای شرکت‌ها تبدیل شده است. این تمرکز، نه‌تنها موجب افزایش رضایت و وفاداری مشتریان می‌شود، بلکه به کسب‌وکارها کمک می‌کند تا با بهینه‌سازی منابع خود، در رقابت باقی بمانند (Nisar & Prabhakar, 2017). این پژوهش به بررسی چگونگی استفاده از الگوریتم‌های هوش مصنوعی و روش‌های پیشرفتهٔ تحلیل داده، برای پیش‌بینی رفتار مشتریان و ارتقای سطح رضایت آن‌ها می‌پردازد.

۱-۱ پیشینهٔ نظری

پیش‌بینی رضایت مشتری به‌عنوان یکی از اهداف راهبردی کلیدی سازمان‌ها مطرح است، درحالی‌که خودِ رضایت مشتری به قضاوت نهایی او دربارهٔ میزان برآورده شدن انتظاراتش اشاره دارد. مدل‌های نظری مختلفی این مفهوم را تبیین می‌کنند و مبنایی تحلیلی برای پیش‌بینی آن فراهم می‌سازند (Oliver, 1980). در پژوهش‌های اولیهٔ حوزهٔ رفتار مصرف‌کننده، رضایت مشتری عموماً به‌عنوان تابعی از مقایسهٔ انتظارات پیشین مشتری با تجربهٔ واقعی پس‌از مصرف تعریف می‌شود. در این چارچوب، مدل عدم تطابق انتظارات (Expectancy–Disconfirmation Theory) یکی از بنیان‌های نظری اصلی برای تبیین رضایت است که بیان می‌کند رضایت یا نارضایتی نتیجهٔ فاصلهٔ میان عملکرد ادراک‌شده و انتظارات اولیهٔ مشتری است. این رویکرد، امکان تبدیل سازه‌های ذهنی مشتری به متغیرهای سنجش‌پذیر را فراهم کرده، اما تمرکز آن بیشتر بر تبیین علّی بوده و نه پیش‌بینی داده‌محور رضایت (Chandra & Fitriyanto, 2024).

در مدل کانو (Kano Model)، طبقه‌بندی نیازهای مشتری به پایه‌ای، عملکردی و انگیزشی به درک ماهیت غیرخطی و ترتیبی تأثیر ویژگی‌ها بر رضایت کمک می‌کند. این طبقه‌بندی در طراحی مدل‌های پیش‌بینی برای اولویت‌بندی متغیرها و تفسیر نتایج، حیاتی است؛ زیرا نشان می‌دهد برخی عوامل فقط از نارضایتی جلوگیری می‌کنند، گروهی رضایت را به‌طور خطی افزایش می‌دهند و برخی نیز تأثیر تصاعدی و شادی‌آفرین دارند (Sinemus et al., 2025).

در فضای دیجیتال، مفاهیمی، مانند کیفیت خدمات الکترونیک و اعتماد آنلاین به‌عنوان متغیرهای پیش‌بین قوی عمل می‌کنند. در حوزهٔ تجارت عمده‌فروشی آنلاین (B2B)، شاخص‌های عملیاتی و رفتاری (دقت/زمان تحویل، حجم سفارش، و تناوب خرید) به‌دلیل عینی و سنجش‌پذیر بودن، نسبت به متغیرهای

جمعیت‌شناختی، قدرت پیش‌بینی‌پذیری بیشتری دارند (Amanda Ardhani & Tania, 2025).

رویکردهای یادگیری ماشین (Machine Learning) برای مدل‌سازی روابط پیچیده بین این پیش‌بینی‌ها و رضایت مشتری، بر مدل‌های خطی سنتی برتری دارند. الگوریتم‌هایی، مانند مدل‌های درختی (جنگل تصادفی) و شبکه‌های عصبی مصنوعی (Multilayer Perceptron (MLP)) توانایی بیشتری در پردازش داده‌های با ابعاد وسیع و غیرخطی دارند (Shen et al., 2023).

برای افزایش دقت و پایداری، از مدل‌های گروهی (Ensemble Methods) مانند رأی‌گیری اکثریت استفاده می‌شود که از ترکیب چند مدل پایه برای کاهش واریانس خطا و بهبود تصمیم نهایی بهره می‌برد (Lin et al., 2022). ازآنجاکه این مدل‌های پیچیده می‌توانند به‌عنوان جعبه سیاه (Black Box) تلقی شوند، حوزهٔ تبیین‌پذیری هوش مصنوعی و به‌طور خاص چارچوب SHAP و TOPSIS برای تفسیر خروجی مدل و رتبه‌بندی اهمیت متغیرها ضروری است (Hassija et al., 2024). براین‌اساس، چارچوب نظری این پژوهش بر این فرض استوار است که ترکیب مدل‌های یادگیری ماشین، ارزیابی چندمعیارهٔ عملکرد و ابزارهای تفسیرپذیری می‌تواند پیش‌بینی دقیق‌تر و کاربردی‌تری از رضایت مشتری در تجارت الکترونیک ارائه دهد.

۲-۱. پیشینهٔ تجربی

در سال‌های اخیر، رشد تجارت الکترونیک و افزایش رقابت در بازارهای آنلاین، اهمیت درک و بهبود رضایت مشتریان را بیش از پیش نمایان کرده است. در این راستا، هوش مصنوعی و یادگیری ماشین به‌عنوان ابزارهای مؤثر برای تحلیل داده‌ها و پیش‌بینی رفتار مشتریان مطرح شده‌اند. یادگیری ماشین، به‌عنوان شاخه‌ای از هوش مصنوعی، به کامپیوترها امکان می‌دهد ازطریق داده‌ها آموزش ببینند و بدون نیاز به برنامه‌ریزی صریح، پیش‌بینی‌های دقیقی ارائه دهند (نوروزی و همکاران، ۱۴۰۲؛ Rane et al., 2024). این فناوری می‌تواند با تحلیل الگوهای رفتاری مشتریان، پیش‌بینی رضایت آن‌ها را تسهیل کرده و به مدیران خرده‌فروشی و عمده‌فروشی آنلاین در تصمیم‌گیری‌های بهینه و اجرای راهبردهای کارآمد کمک کند (Cavalcante Siebert et al., 2021).

مطالعات گوناگونی به بررسی نقش الگوریتم‌های یادگیری ماشین در تحلیل رضایت مشتری پرداخته‌اند؛ برای مثال، بالاکریشنان و همکاران در پژوهشی از الگوریتم‌های بیزین ساده (Naive Bayes)، رگرسیون لاجستیک (Logistic Regression)، ماشین بردار پشتیبانی (SVM)، و جنگل تصادفی (Random Forest) برای طبقه‌بندی رضایتمندی مشتریان در تجارت الکترونیک استفاده کردند (Balakrishnan et al., 2022). نتایج این پژوهش نشان داد که الگوریتم جنگل تصادفی با دقت ۹۰.۶۶درصد بهترین عملکرد را داشته است، درحالی‌که بیزین ساده با دقت ۵۴.۸۴درصد کمترین دقت را ارائه داده است.

فوآد و همکاران نیز مدلی برای پیش‌بینی امتیازات دیدگاه مشتریان در تجارت الکترونیک براساس داده‌های تراکنش تاریخی پیشنهاد کردند (Fouad et al., 2023). این مدل با استفاده از ویژگی‌های عددی، به‌صورت یک مسئلهٔ طبقه‌بندی چندکلاسه با پنج سطح

رتبه‌بندی تعریف شد. الگوریتم جنگل تصادفی پس‌از انتخاب ویژگی‌ها، با امتیاز F1 (F1-score) برابر با ۰.۶۷، بهترین عملکرد را داشت. همچنین، عوامل کلیدی مؤثر بر رضایت شامل زمان تحویل، ارزش محصول، و هزینهٔ حمل‌ونقل شناسایی شدند.

در پژوهشی دیگر، مامتا و سانگوان چارچوبی ترکیبی برای تحلیل رفتار مشتری ارائه دادند که شامل دو ماژول تحلیل نیت خرید و تحلیل رهاسازی بود. برای این چارچوب از ترکیب سه الگوریتم یادگیری عمیق شامل CNN، GAN، و LSTM-RNN استفاده کردند (Mamta, & Sangwan, 2024). این مدل توانست دقت ۹۷درصد را به دست آورد و بینش‌های ارزشمندی درزمینهٔ توصیه‌های شخصی‌سازی‌شده برای مشتریان ارائه دهد. زغلول و همکاران نیز رفتار مصرف‌کنندگان آنلاین را ازطریق مدل‌سازی مقایسه‌ای یادگیری ماشین تحلیل کردند. این مطالعه از داده‌های بیش از ۱۰۰هزار سفارش آنلاین استفاده و مدل‌های جنگل تصادفی و SVM را با تکنیک‌های یادگیری عمیق مقایسه کرد. جنگل تصادفی با دقت ۹۲درصد بهترین عملکرد را در پیش‌بینی رضایت آیندهٔ مشتری ارائه داد و در مقایسه با مدل‌های یادگیری عمیق عملکرد بهتری داشت. این پژوهش، همچنین، عوامل کلیدی مؤثر بر رضایت، مانند زمان تحویل و دقت سفارش را شناسایی کرد (Zaghloul et al., 2024).

به‌طور مشابه، پژوهش‌های تازه نشان می‌دهند که مدل‌های یادگیری ماشین و ترکیبی در پیش‌بینی/تبیین رضایت مشتری در تجارت الکترونیک کارآمدند؛ برای نمونه، کومار در مطالعه‌ای با بهره‌گیری از یک رویکرد ترکیبی مبتنی‌بر یادگیری عمیق و روش‌های ensemble بر داده‌های مرورهای آمازون، عوامل یا ریشه‌های نارضایتی مشتریان را استخراج و عملکردی بهتر نسبت به مدل‌های پایه کلاسیک گزارش کرده است (Kumar et al., 2025). همچنین، داودی و همکاران در مطالعه‌ای با بهره‌گیری از تحلیل احساس مبتنی بر جنبه‌ها (Aspect-Based Sentiment Analysis: ABSA) نشان می‌دهند که مدل‌های مبتنی‌بر معماری ترنسفورمر (RoBERTa) در یادگیری سیگنال‌های مرتبط با رضایت مشتری عملکرد برتری دارند و به دقتی بیش از ۹۰درصد در طبقه‌بندی جنبه–احساس دست می‌یابند. این نتایج را می‌توان برای بهبود رضایت و تصمیم‌های عملیاتی استفاده کرد (Davoodi et al., 2025).

این مطالعات نشان می‌دهند که استفاده از الگوریتم‌های پیشرفتهٔ یادگیری ماشین و یادگیری عمیق می‌تواند بینش‌های مهمی دربارهٔ رفتار مشتریان ارائه داده و به کسب‌وکارها کمک کند تا راهبردهای مؤثرتری برای افزایش رضایت مشتریان و بهبود تجربهٔ خرید طراحی کنند. هدف پژوهش، پیش‌بینی میزان رضایت پیش‌از دریافت بازخورد است (Nurul Ain et al., 2024).

شکل ۱ مربوط به ابر واژگان (Word Cloud) مبانی نظری موضوع است. این ابر واژگان، تمرکز پژوهش بر رضایت مشتریان در فضای آنلاین و تحلیل آن‌ها را با استفاده از یادگیری ماشین و روش‌های مدرن نشان می‌دهد.

شکل ۱. ابر واژگان پژوهش شامل برجسته‌ترین مفاهیم مرتبط با تحلیل رضایت مشتری آنلاین در فضای دیجیتال

Figure 1. Word Cloud of the Study Showing the Most Prominent Concepts Related to Online Customer Satisfaction Analysis in the Digital Environment

در سال‌های اخیر، محور بسیاری از مطالعات از تحلیل احساسات پس‌از خرید به‌سمت پیش‌بینی داده‌محور رضایت مشتری جابه‌جا شده است؛ بااین‌حال، بخش عمده‌ای از مبانی نظری موجود بر خرده‌فروشی آنلاین متمرکز است و کمتر به اقتضائات عمده‌فروشی B2B در ایران می‌پردازد؛ جایی که رضایت تحت‌تأثیر تعاملات با مسئول فروش، شرایط اعتباری و محدودیت‌های تحویل شکل می‌گیرد. بر این مبنا، مسئلهٔ این پژوهش، پیش‌بینی میزان رضایت مشتری پیش‌از دریافت بازخورد است تا امکان مداخلهٔ پیشگیرانه برای حفظ یا ارتقای تجربه فراهم شود.

برای صورت‌بندی مسئله، از داده‌های عملیاتی سفارش‌ها استفاده می‌شود؛ مجموعهٔ ویژگی‌ها حوزه‌های محصول/سبد خرید، فرایند تحویل و سابقهٔ تعامل را پوشش می‌دهد (وزن کل اقلام، فاصلهٔ زمانی تا خرید قبلی، وضعیت و زمان تحویل). برچسب هدف نیز به‌صورت چندکلاسه با ماهیت ترتیبی تعریف شده تا طیف رضایت را واقع‌بینانه‌تر بازنمایی کند.

نوآوری این پژوهش در پرداختن به خلأی مشخص در مبانی نظری نهفته است: پیش‌بینی رضایت مشتری در بستر عمده‌فروشی آنلاین B2B با استفاده از داده‌های عملیاتی کوچک، که معمولاً بدون چارچوب مقایسه‌ای منسجم و بدون کنترل بیش‌برازش (Overfitting) گزارش می‌شود. با صورت‌بندی مسئله به‌عنوان طبقه‌بندی چندکلاسه و طراحی یک مدل گروهی رأی‌گیری اکثریت بر فراز سه خط‌پایه به‌روز (CatBoost، جنگل تصادفی، MLP) -که هرکدام به‌طور مستقل با الگوریتم ژنتیک تنظیم ابرپارامتر شده‌اند- نوآوری را در ترکیب پایداری عملکرد و کنترل پیچیدگی ارائه می‌کنیم. به‌جای اتکا به معیاری منفرد یا ادعاهای کلی، ارزیابی چندمعیاره با روش TOPSIS انجام شده است تا مدل‌ها بر مبنای مجموعه‌ای از شاخص‌ها (Accuracy، Precision، Recall، F1، MCC و CK) و پایداری در اعتبارسنجی متقاطع پنج‌بخشی (5-fold) رتبه‌بندی شوند؛ بدین‌ترتیب، برتری مدل گروهی در برابر بهترین

خط‌پایه به‌صورت تصمیم‌گیری چندمعیاره مستند می‌شود. در نهایت، با SHAP (کلاس‌به‌کلاس) پیوندی تحلیلی میان نتایج و مبانی نظری برقرار می‌کنیم و نشان می‌دهیم چرا شاخص‌های عملیاتی/لجستیکی (وزن سبد و فاصله تا خرید قبلی) در این حوزه نسبت به متغیرهای جمعیت‌شناختی/قیمتی صِرف اهمیت بیشتری می‌یابند. بدین‌سان، سهم پژوهش فراتر از توصیف، و به‌صورت چارچوبی اتکاپذیر برای گزینش مدل و اقدام مدیریتی پیشگیرانه تثبیت می‌شود.

۲- روش‌شناسی پژوهش

این بخش، به تشریح مراحل انجام پژوهش و روش‌های به‌کاررفته برای تحلیل داده‌ها و ساخت مدل‌های یادگیری ماشین می‌پردازد. ابتدا داده‌های مربوط به مشتریان ازقبیل ویژگی‌های جمعیت‌شناختی، اطلاعات خرید و سایر متغیرهای مرتبط با میزان رضایت، جمع‌آوری و پیش‌پردازش شدند. در فرایند پیش‌پردازش، داده‌های ناقص حذف و ویژگی‌های عددی نرمال‌سازی شدند تا تأثیر مقیاس‌های مختلف کاهش یابد؛ سپس با استفاده از روش یادگیری گروهی رأی اکثریت و مدل‌های جنگل تصادفی، CatBoost و شبکهٔ عصبی چندلایه (MLP)، مدلی برای پیش‌بینی میزان رضایت مشتریان توسعه یافت. به‌منظور بهینه‌سازی ابرپارامترهای این مدل‌ها، از الگوریتم ژنتیک استفاده شد.

داده‌های استفاده‌شده در این پژوهش مربوط به ۳۴۹ مشتری شرکت آرمان روشن فرتاک (فروشندهٔ انواع چسب صنعتی) است که شامل ویژگی‌های مختلفی ازجمله جنسیت مسئول خرید، سن مسئول خرید، شهر، مجموع هزینه‌های انجام‌شده (تا قبل‌از آخرین خرید)، وزن کل محصولات خریداری‌شده (تا قبل‌از آخرین خرید)، کمیسیون فروش، روزهای گذشته از آخرین خرید، نوع شرکت و میزان رضایت (به‌عنوان ستون هدف) است. با توجه به تعداد کم داده‌ها، برای کاهش خطر بیش‌برازش ناشی از اندازهٔ نمونهٔ محدود، از ۵-fold طبقه‌بندی‌شده و محدودسازی پیچیدگی مدل‌ها استفاده شده است. این داده‌ها به شرکت کمک می‌کند تا با تحلیل رفتار و ویژگی‌های مشتریان، بتواند الگوهای مرتبط با رضایت آن‌ها را شناسایی کرده و از این اطلاعات برای بهبود خدمات و ارتقا میزان رضایت مشتریان بهره‌برداری کند. این ویژگی‌ها بر دو اساسِ در دسترس بودن و نظر متخصص انتخاب شده‌اند. جدول ۱ توصیف دقیقی از ویژگی‌های اصلی داده‌ها را ارائه می‌دهد که در مدل یادگیری ماشین به کار گرفته شده است. داده‌های مورد بررسی شامل ترکیبی از ویژگی‌های عددی و طبقه‌ای است که هرکدام نقش مهمی در توصیف رفتار مشتریان ایفا می‌کنند. داده‌های طبقه‌ای (Categorical) داده‌هایی هستند که به دسته‌ها یا گروه‌های خاصی تقسیم می‌شوند و هر دسته یک مقدار گسسته دارد. در مقابل، داده‌های عددی (Numerical) داده‌هایی هستند که به‌صورت پیوسته یا گسسته مقادیر عددی دارند و می‌توانند عملیات ریاضی، مانند میانگین یا انحراف معیار بر آن‌ها اعمال کرد (Kampezidou et al., 2024).

جدول ۱. توصیف کامل ویژگی‌های مجموعه داده‌ها

Table 1. Complete description of dataset features

ویژگی	نوع	میانگین/فراوانی	انحراف معیار
جنسیت (۰ = مرد، ۱ = زن)	طبقه‌ای	۲۱/۵درصد زن، ۷۸/۵درصد مرد	–
سن	عددی	۴۴ سال و ۴ ماه	۵/۱
شهر (کلان‌شهر و شهرستان)	طبقه‌ای	۷۷ = کلان‌شهر، ۲۳ = شهرستان	–
مجموع هزینه‌های انجام‌شده	عددی	۱۵۰۴۰۳۴۴۸ ریال	۷۰۴۷۱۳۱۰/۱۲
وزن کل محصولات خریداری‌شده	عددی	۵۱۳ کیلو و ۴۰۰ گرم	۲۳۴/۹
کمپیسون (۰ = خیر، ۱ = بله)	طبقه‌ای	۷۰/۷ بله، ۲۹/۳ خیر	–
روزهای گذشته از آخرین خرید	عددی	۲۶ هفته و ۶ روز	۱۳/۴
نوع شرکت (خصوصی یا دولتی)	طبقه‌ای	۷۹/۵درصد خصوصی، ۲۰/۵درصد دولتی	–
سطح رضایت (۰ = راضی، ۱ = بی‌سو، ۲ = ناراضی)	طبقه‌ای	۱۷/۶درصد ناراضی، ۳۰/۸درصد بی‌سو، ۵۱/۶درصد راضی	–

(منبع داده‌های پژوهش)

فروشنده میزان رضایت مشتریان را به کمک نظرسنجی آنلاین و نیز به‌صورت تلفنی دریافت و محاسبه کرده است. همچنین، سه سطح برای مشتریان در نظر گرفته شده است (۰ = راضی، ۱ = بی‌سو، ۲ = ناراضی)، به این معنا که مسئلهٔ یادگیری ماشین مورد بررسی یک مسئلهٔ طبقه‌بندی چنددسته‌ای (Multi-class Classification) است. این صورت‌بندی با سازوکار ارزیابی داخلی سازمان هم‌خوان است و امکان مداخلهٔ پیشگیرانه را پیش‌از ثبت بازخورد فراهم می‌کند. در این نوع مسائل، هدف مدل یادگیری ماشین آن است که داده‌های ورودی را در یکی از چندین دستهٔ ممکن طبقه‌بندی کند. چالش اصلی در مسائل طبقه‌بندی چنددسته‌ای، یادگیری یک مدل دقیق است که بتواند تفاوت‌های جزئی میان کلاس‌ها را به‌خوبی شناسایی و آن‌ها را به‌درستی طبقه‌بندی کند (Adnane & Zerari, 2024). به‌طور مشخص، مدل باید بتواند براساس ویژگی‌های مشتری، میزان رضایت مشتری را در یکی از سه دستهٔ زیر پیش‌بینی کند:

- راضی (کلاس ۰): مشتریانی که از خدمات یا محصولات رضایت دارند؛
- بی‌سو (کلاس ۱): مشتریانی که به خدمات یا محصولات، نه رضایت کامل دارند و نه ناراضی‌اند؛
- ناراضی (کلاس ۲): مشتریانی که از خدمات یا محصولات ناراضی‌اند.

در این پژوهش، پیش‌پردازش داده‌ها یکی از مراحل مهم آماده‌سازی داده‌ها برای مدل‌سازی به شمار می‌رود. داده‌های اولیه معمولاً شامل مقادیری گمشده (Missing Values) و مقیاس‌های متفاوت در ویژگی‌ها هستند که می‌تواند عملکرد مدل یادگیری ماشین را به‌شدت تحت‌تأثیر قرار دهد؛ بنابراین، چندین مرحلهٔ کلیدی در پیش‌پردازش داده‌ها اجرا شده است:

۱. حذف رکوردهای دارای مقادیر گمشده:

در بسیاری از موارد، برخی ویژگی‌ها یا رکوردها ممکن است دارای مقادیر گمشده باشند. وجود آن‌ها می‌تواند باعث کاهش دقت مدل و ایجاد اختلال در فرایند یادگیری شود. برای مقابله با این مشکل، رکوردهایی که دارای مقادیر گمشده در ویژگی‌های اصلی بودند، حذف شدند. در مجموعه‌دادهٔ ۳۴۹

رکوردی، فقط ۳ رکورد (≈۰/۸۶درصد) به‌دلیل نقص جبران‌ناپذیر در فیلدهای کلیدی حذف شد.

۲. نرمال‌سازی داده‌ها (Data Normalization):

ویژگی‌های مختلف در داده‌های خام، دارای مقیاس‌های متفاوتی‌اند. این تفاوت مقیاس‌ها می‌تواند باعث شود مدل به‌اشتباه، ویژگی‌هایی با مقیاس بزرگ‌تر را مهم‌تر تلقی کند. برای رفع این مشکل، داده‌ها با استفاده از تکنیک نرمال‌سازی (Normalization) استاندارد شدند. در این پروژه، از نرمال‌سازی استاندارد (StandardScaler) استفاده شد، که ویژگی‌ها را به‌گونه‌ای مقیاس‌بندی می‌کند که میانگین هر ویژگی صفر و انحراف معیار آن برابر با یک باشد. این کار باعث می‌شود که تمام ویژگی‌ها در مقیاس یکسان قرار بگیرند و مدل بتواند به‌درستی از تمام ویژگی‌ها استفاده کند (Efendi et al., 2024). به‌منظور جلوگیری از نشت اطلاعات در مرحلهٔ آزمایش مربوط به ۸۰درصد از داده‌ها، از اعتبارسنجی k-fold طبقه‌بندی‌شده (K=۵) استفاده کردیم.

۱-۲. روش‌شناسی مدل پیشنهادی

در این پژوهش، از یادگیری گروهی رأی اکثریت به‌عنوان روشی قوی برای بهبود عملکرد مدل‌های یادگیری ماشین استفاده شده است. مدل پیشنهادی شامل ترکیب سه الگوریتم مختلف است: جنگل تصادفی، CatBoost و MLP. ترکیب این سه مدل با استفاده از روش رأی اکثریت صورت گرفته است. در یادگیری گروهی رأی اکثریت، چندین مدل با ساختارها و الگوریتم‌های مختلف ترکیب می‌شوند تا از نقاط قوت هر مدل بهره‌برداری شود. در این پژوهش، روش رأی اکثریت برای ترکیب نتایج مدل‌ها به کار گرفته شده است. هریک از مدل‌ها (جنگل تصادفی، CatBoost و MLP) یک پیش‌بینی انجام می‌دهد و پیش‌بینی نهایی براساس اکثریت آرا بین این مدل‌ها تعیین می‌شود (اسمعیل‌پور و همکاران، ۱۴۰۳). این روش می‌تواند خطاهای ناشی از مدل‌های تکی را کاهش داده و پایداری بیشتری ایجاد کند. مدل‌هایی که عملکرد ضعیف‌تری دارند توسط رأی اکثریت تعدیل می‌شوند. با ترکیب مدل‌ها با دیدگاه‌های متفاوت، دقت کلی مدل افزایش می‌یابد. ترکیب مدل‌ها همچنین می‌تواند از بیش‌برازش جلوگیری کند؛ زیرا مدل‌های مختلف گرایش‌های بیش‌برازش مختلفی دارند و ترکیب آن‌ها منجر به تعادل در نتایج می‌شود (Esmaielpour et al., 2026).

برای مقایسه hard voting و soft voting یک آزمون سادهٔ نشانه (Sign Test) انجام دادیم: در یک ۵-fold CV، در هر فولد معیارهای اصلی برای هر دو رویکرد محاسبه و سپس «آزمون نشانه» غیررسمی اجرا شد (شمارش تعداد فولدهایی که در آن، hard بهتر/ضعیف‌تر/برابر است). نتیجه نشان داد hard voting در اکثریت فولدها برتری یا برابری دارد و اختلاف‌ها در همهٔ فولدها کوچک و زیر ۱درصد باقی می‌ماند؛ با توجه به اندازهٔ نمونه و حساسیت soft به کالیبراسیون احتمال، رویکرد hard به‌دلیل سادگی و پایداری به‌عنوان مدل نهایی انتخاب شد.

جنگل تصادفی یک الگوریتم یادگیری گروهی است که از ترکیب چندین درخت تصمیم‌گیری برای بهبود دقت پیش‌بینی استفاده می‌کند. جنگل تصادفی با ترکیب چندین درخت تصمیم‌گیری و استفاده از نمونه‌های تصادفی داده‌ها، به کاهش واریانس و افزایش دقت کمک می‌کند. جنگل تصادفی همچنین می‌تواند با داده‌هایی که شامل ترکیبی از ویژگی‌های پیوسته و گسسته‌اند، به‌خوبی کار کند و اهمیت

ویژگی‌های مختلف را نیز ارزیابی کند (Mirzahossein & Rezashoar, 2025).

Catboost یک الگوریتم یادگیری گروهی مبتنی‌بر تقویت گرادیان (Gradient Boosting) است که به‌طور ویژه برای کار با داده‌های دسته‌ای طراحی شده است. این مدل با بهینه‌سازی فرایند یادگیری و استفاده از تکنیک‌های پیشرفته، مانند پردازش داده‌های ناپیوسته، می‌تواند به دقت بیشتری در پیش‌بینی دست یابد. این الگوریتم می‌تواند مدل‌هایی با دقت زیاد تولید کند و نسبت به سایر مدل‌های تقویت گرادیان، زمان اجرای کمتری دارد (Zhang & Guo, 2024).

شبکه‌های عصبی چندلایه نوعی شبکهٔ عصبی عمیق‌اند که با استفاده از لایه‌های گوناگون نورون‌ها، می‌توانند الگوهای پیچیدهٔ موجود در داده‌ها را فراگیرند. این روش به‌دلیل ساختار انعطاف‌پذیر خود، برای پیش‌بینی داده‌های پیچیده و غیرخطی مناسب است (Zhang et al., 2024).

۲-۲. بهینه‌سازی ابرپارامترها

در این پژوهش، برای بهبود عملکرد مدل‌های یادگیری ماشین و بهینه‌سازی ابرپارامترها، از الگوریتم ژنتیک استفاده شده است. الگوریتم ژنتیک روشی مبتنی‌بر تکامل و انتخاب طبیعی است که به‌دنبال یافتن بهترین ترکیب از پارامترها برای بهینه‌سازی مدل است. الگوریتم ژنتیک از جستجوی تصادفی هدایت‌شده استفاده می‌کند، که این موضوع به آن اجازه می‌دهد تا فضای جستجو را به‌طور مؤثری کاوش کرده و راه‌حل‌های مناسبی برای تنظیم ابرپارامترها پیدا کند. درحالی‌که روش‌های کلاسیکی، مانند جستجوی شبکه‌ای (Grid Search) و جستجوی تصادفی (Random Search) به‌شکل کور جستجو می‌کنند، الگوریتم ژنتیک با استفاده از انتخاب طبیعی و شبیه‌سازی فرایند تکامل، جستجویی هدایت‌شده‌تر و کارآمدتر انجام می‌دهد. در مسائلی که تنظیم ابرپارامترها پیچیده و شامل ترکیبات مختلف از پارامترهاست (مدل‌های جنگل تصادفی، CatBoost و MLP)، الگوریتم ژنتیک می‌تواند به‌طور تدریجی و با ارزیابی هر نسل از جمعیت، به‌سمت بهترین ترکیب ممکن از پارامترها حرکت کند. این ویژگی به‌ویژه برای مدل‌هایی با تعداد زیاد پارامترهای تنظیمی حائز اهمیت است (Bakır & Ceviz, 2024).

پارامترهای الگوریتم ژنتیک به شرح زیرند:

(جمعیت=۲۰، تعداد نسل‌ها=۴۰، کراس‌اور=۰.۸۵، جهش=۰.۱۰، انتخاب Tournament=3، توقف زودهنگام=۱۰ نسل بی‌بهبود)

۳-۲. معیارهای ارزیابی

همچنین، به‌منظور ارزیابی دقت و کارایی الگوریتم‌ها در پیش‌بینی، شاخص‌هایی ازقبیل دقت (Accuracy)، صحت (Precision)، بازخوانی (Recall)، امتیاز F1-، ضریب همبستگی مَتیوز (Matthews Correlation Coefficient: MCC) و ضریب کاپای کوهن (Cohen's Kappa: CK) بررسی شده‌اند (Reddy et al., 2024)؛ ازاین‌رو، محاسبه هرکدام از شاخص‌ها به شرح زیر است:

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

$$Precision = \frac{tp}{tp + fp}$$

$$Recall = \frac{tp}{tp + fn}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$MCC = (tp \times tn - fp \times fn) / \sqrt{((tp + fp) \times (tp + fn) \times (tn + fp) \times (tn + fn))}$$

$$CK = (Po - Pe) / (1 - Pe)$$

tp = الگوریتم، نمونه را در دستهٔ مثبت طبقه‌بندی کرده و نمونه هم مثبت است؛

tn = الگوریتم، نمونه را در دستهٔ منفی طبقه‌بندی کرده و نمونه هم منفی است؛

fp = الگوریتم، نمونه را در دستهٔ مثبت طبقه‌بندی کرده، اما نمونه منفی است؛

fn = الگوریتم، نمونه را در دستهٔ منفی طبقه‌بندی کرده، اما نمونه مثبت است؛

Po = نسبت توافق مشاهده‌شده (دقت کلی)؛

Pe = احتمال توافق تصادفی.

۲-۴. منطق انتخاب روش‌ها و سنجه‌ها

برای دادهٔ عملیاتی کوچک و ترکیبی، CatBoost به‌دلیل کارایی زیاد با ویژگی‌های عددی/رده‌ای و کنترل پیش‌فرض پیچیدگی، Random Forest به‌عنوان خط‌پایه نیرومند و پایدار غیرخطی، و MLP برای ثبت تعاملات پیچیده‌تر برگزیده شدند؛ سپس به‌منظور بهره‌گیری از ناهمگونی خطاها و کاستن از واریانس، بدون افزودن پارامترهای اضافی، این سه مدل در قالب یک Ensemble با استفاده از روش رأی‌گیری اکثریت (Majority Voting) از نوع رأی‌گیری سخت (Hard Voting) تجمیع شدند (از روش‌های پیچیده‌تری چون Stacking عمداً پرهیز شد تا ریسک بیش‌برازش در n=349 افزایش نیابد). ابرپارامترهای هر مدل نیز به‌صورت جداگانه با الگوریتم ژنتیک و تابع برازندگی مبتنی‌بر F1–وزنی در k-fold طبقه‌بندی‌شده t تنظیم شد تا تعادل دقت/پایداری حفظ شود. ازسوی سنجه‌ها، صِرف Accuracy کفایت ندارد؛ لذا Precision و Recall –به‌ویژه برای کلاس ۲=ناراضی– و F1–وزنی گزارش شدند.

۲-۵. مقایسهٔ مدل‌ها به کمک روش TOPSIS[1]

روش TOPSIS نوعی رتبه‌بندی چندمعیاره است که هر گزینه را هم‌زمان با «راه‌حل ایدئال» (بهترین مقادیر معیارها) و «راه‌حل ضدایدئال» (بدترین مقادیر) مقایسه می‌کند. مراحل به‌اختصار عبارت‌اند از: ۱) تشکیل ماتریس معیارها و نرمال‌سازی برداری ستون‌ها؛ ۲) وزن‌دهی به معیارها و ساخت ماتریس وزن‌دار؛ ۳) استخراج بردار ایدئال و ضدایدئال (برای همهٔ معیارهای سود، مقدار بیشینه ایدئال است)؛ ۴) محاسبهٔ فاصلهٔ اقلیدسی هر گزینه از ایدئال و ضدایدئال؛ ۵) محاسبهٔ ضریب نزدیکی $C = \frac{S^-}{S^+ + S^-}$ که هرچه بزرگ‌تر باشد، گزینه مطلوب‌تر است. در این پژوهش برای جلوگیری از سوگیری و تأکید یکسان بر Accuracy، Precision، Recall و F1، وزن‌ها برابر در نظر گرفته شد؛ لذا رتبهٔ نهایی فقط براساس عملکرد هم‌زمان در چهار معیار و بدون ترجیح پیشینیِ هیچ معیاری، به دست آمده است (Taherdoost & Madanchian, 2024).

۲-۶. روش SHAP Values

با توجه به اینکه مدل‌های استفاده‌شدهٔ شبکه‌های عصبی عمیق یا مدل‌های گروهی (جنگل تصادفی یا CatBoost) تفسیرناپذیر و پیچیده‌اند، از SHAP برای تفسیر بهتر رابطهٔ هر ویژگی با پیش‌بینی‌های مدل استفاده شده است تا سهم هر ویژگی در نتیجهٔ نهایی مشخص شود. این روش از نظریهٔ بازی‌ها الهام گرفته شده و هدف آن، محاسبهٔ سهم هر ویژگی در پیش‌بینی یک مدل است. SHAP به‌طور دقیق مشخص می‌کند که هر ویژگی چگونه و تا چه حد در نتیجهٔ نهایی یک مدل نقش دارد.

[1] Technique for Order Preference by Similarity to Ideal Solution

اهمیت کاربرد SHAP در این است که شفافیت و تفسیرپذیری مدل‌های پیچیده را ممکن می‌سازد؛ زیرا، به ما این امکان را می‌دهد که سهم و تأثیر هر ویژگی بر پیش‌بینی مدل را بررسی کنیم و مدل را شفاف‌تر کنیم. همچنین، SHAP تعیین می‌کند که کدام ویژگی‌ها بیشترین تأثیر را بر نتایج دارند، که این موضوع برای تحلیل دقیق‌تر داده‌ها و بهبود عملکرد مدل‌ها بسیار ارزشمند است. در مسائل مهمی مانند پزشکی، بانکداری، و تجارت که تصمیمات براساس پیش‌بینی‌های مدل گرفته می‌شود، SHAP با ارائهٔ توضیحات برای هر پیش‌بینی، به تصمیم‌گیران کمک می‌کند تا به درک بهتری از دلایل پیش‌بینی‌های مدل برسند (Choudhary et al., 2024).

۳- یافته‌ها

در این بخش، یافته‌های حاصل از تحلیل داده‌ها و نتایج عملکرد مدل‌های یادگیری ماشین ارائه می‌شود. در بخش پیشین با ارائهٔ توصیف آماری داده‌ها، شامل ویژگی‌های طبقه‌ای و عددی، توزیع ویژگی‌های مختلف بررسی شد. برای متغیرهای عددی، میانگین و انحراف معیار و برای متغیرهای طبقه‌ای، فراوانی درصدی گزارش ارائه شد؛ سپس با استفاده از روش‌های یادگیری گروهی رأی اکثریت شامل مدل‌های جنگل تصادفی، CatBoost و MLP، مسئلهٔ پیش‌بینی میزان رضایت مشتریان در سه سطح ناراضی، بی‌سو و راضی مدل‌سازی شد. برای بهبود عملکرد مدل‌ها، از الگوریتم ژنتیک برای تنظیم ابرپارامترها استفاده شده است. همچنین، برای تفسیر بهتر تأثیر ویژگی‌ها بر نتایج مدل، روش SHAP به کار گرفته شد روشی که نقش هر ویژگی در پیش‌بینی‌های مدل را به‌صورت شفاف و تفسیرپذیر بیان می‌کند. در ادامه، نتایج حاصل از این مدل‌ها و تفسیرهای مربوط به آن‌ها ارائه می‌شود.

پس‌از اجرای الگوریتم ژنتیک و با هدف تعیین بهترین ابرپارامترها، جدول ۲ مجموعه‌ای از پارامترهای بهینه‌شده برای سه مدل یادگیری ماشین (جنگل تصادفی، CatBoost و MLP) را نشان می‌دهد که برای موضوع طبقه‌بندی چندکلاسه استفاده شده‌اند. برای هر مدل، مجموعه‌ای از پارامترها تعیین و بهترین مقدار برای هریک با استفاده از الگوریتم ژنتیک به دست آمده است. این مقادیر بهینه در نهایت باعث بهبود عملکرد کلی مدل‌ها و دقت پیش‌بینی‌های آن‌ها شده است. پارامترهای کلیدی و دامنه‌های جستجو براساس تأثیر چشمگیر آن‌ها بر عملکرد و کارآیی محاسباتی مدل تعیین شدند، که ضمن امکان ایجاد فضای جستجوی هدفمند، نتایج بهینه‌سازی مطلوب را نیز تضمین می‌کنند.

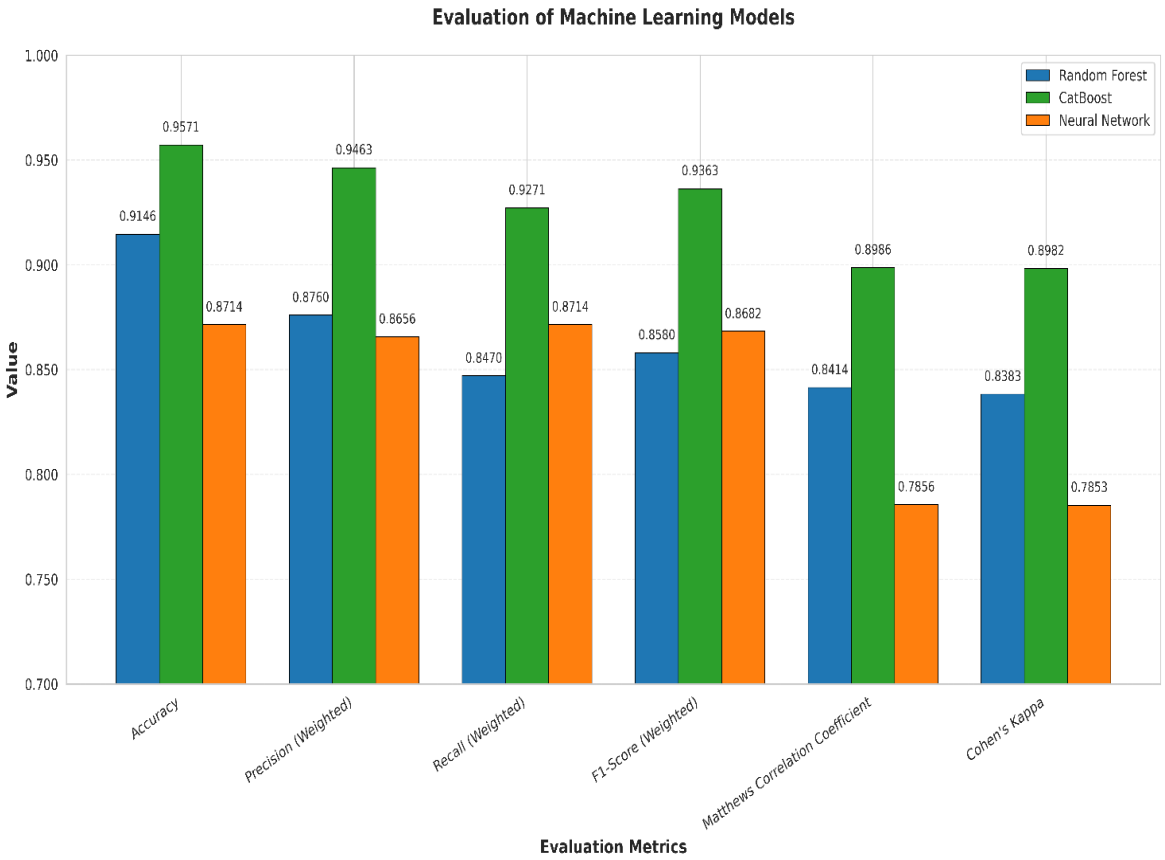
جدول ۲. بهترین پارامترهای بهینه‌شده برای مدل‌های جنگل تصادفی، CatBoost و MLP

Table 2. The optimal hyperparameters for the Random Forest, CatBoost, and MLP models

ویژگی‌ها	پارامترها	بهترین پارامتر
جنگل تصادفی		
تعداد درخت‌ها (n_estimators)	[۱۰، ۱۰۰۰]	۱۱۰
حداکثر عمق درخت (max_depth)	[۵, ۱۵, ۲۵, ۳۵, ۴۵, ۱۱۱۱]	۵
حداقل تعداد نمونه‌ها برای تقسیم (min_samples_split)	[۲، ۲۲]	۱۰
حداقل تعداد نمونه‌ها در برگ (min_samples_leaf)	[۱، ۲۱]	۱۷
معیار (criterion)	gini, entropy	gini
CatBoost		
تابع زیان (loss_function)	MultiClass, MultiClassOneVsAll	MultiClass
میزان یادگیری (learning_rate)	[۰/۰۱، ۰/۰۰۱]	۰/۰۶۳
عمق (depth)	[۲، ۸]	۳
تعداد تکرارها (iterations)	[۱۰۰، ۱۰۰۰]	۵۰۰
تنظیم l2 برگ (l2_leaf_reg)	[۱، ۱۰]	۱
دمای bagging (bagging_temperature)	[۰، ۱]	۰/۴
استحکام تصادفی (random_strength)	[۱، ۲۰]	۱۱
تعداد مرزها (border_count)	[۳۲، ۲۵۵]	۹۵
تعداد تکرارهای برآورد برگ (leaf_estimation_iterations)	[۱، ۱۰]	۹
MLP		
روش بهینه‌سازی (optimizer_method)	Adam, SGD, RMSprop, Adagrad, Nadam	Adagrad
تعداد لایه‌ها (num_layers)	[۱، ۱۰]	۴
تابع فعال‌سازی (activation)	relu, tanh, sigmoid, softmax	softmax
میزان یادگیری (learning_rate)	[۰/۱، ۰/۰۰۱]	۰/۳
تابع زیان (loss_function)	categorical_crossentropy, mean_squared_error	categorical_crossentropy
اندازه دسته (batch_size)	[۱۰۰، ۱۰۰۰]	۱۷۰
تعداد دوره‌ها (epochs)	[۱۰۰، ۱۰۰۰]	۲۷۰

(Nematzadeh et al., 2022)

با توجه به انتخاب بهترین پارامترهای این سه مدل، در شکل ۲ می‌توان اجرا و مقایسهٔ عملکرد مدل‌های جنگل تصادفی، CatBoost و MLP را در بحث طبقه‌بندی رضایت مشتری مشاهده کرد.



شکل ۲. مقایسهٔ عملکرد سه روش اصلی جنگل تصادفی، Catboost و MLP براساس شاخص‌های اصلی

Figure 2. Performance comparison of three main methods (Random Forest, CatBoost, and MLP) based on main metrics

همان‌طور که مشاهده می‌شود، CatBoost در تمامی معیارها عملکرد بهتری در مقایسه با جنگل تصادفی و MLP از خود نشان داده است. این مدل به‌ویژه در معیارهای بازیابی و دقت مثبت، که برای تصمیم‌گیری‌های درست در طبقه‌بندی چندکلاسه بسیار مهم‌اند، بهترین نتایج را ارائه کرده است. جنگل تصادفی نیز در برخی موارد عملکرد پذیرفتنی داشته، اما به‌طور کلی پایین‌تر از CatBoost قرار می‌گیرد. MLP، نسبت به دیگر مدل‌ها عملکرد پایین‌تری دارد، اما همچنان نتایج مناسبی ارائه داده است و در برخی موارد با جنگل تصادفی رقابت می‌کند.

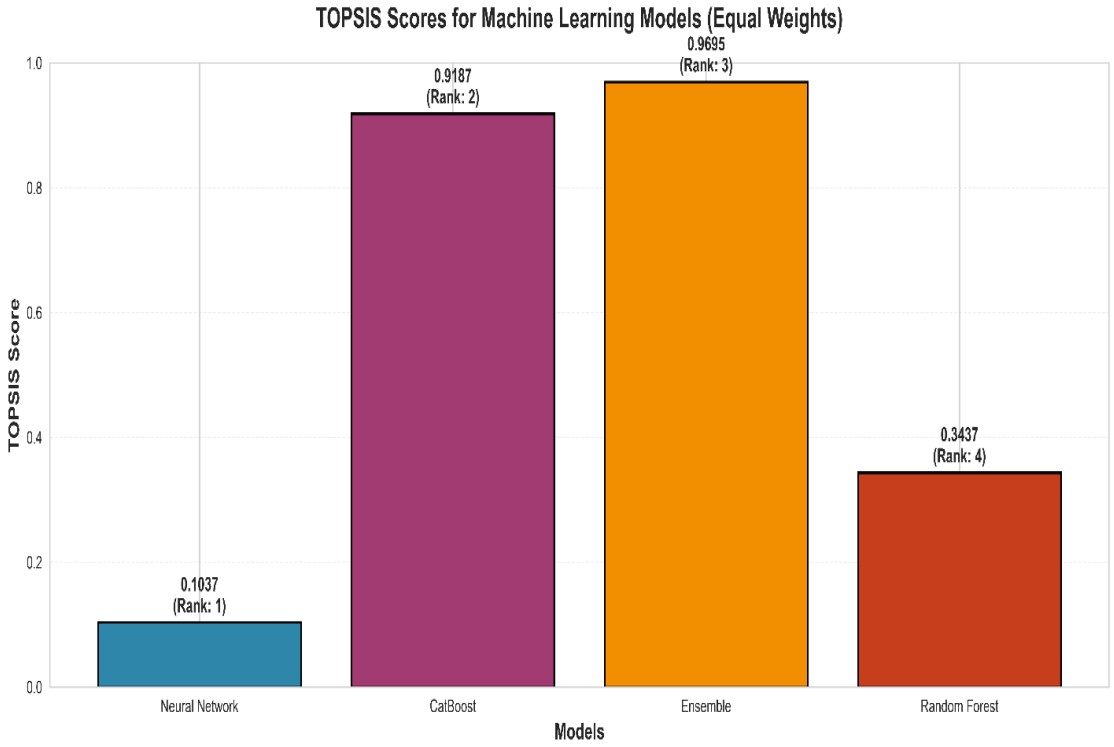
پس‌از اجرای مدل گروهی رأی اکثریت براساس سه مدل یادشده، نتایج شاخص‌های مربوط به آن در جدول ۳ آمده است.

جدول ۳. نتایج عملکرد شاخص‌های مدل گروهی رأی اکثریت براساس مدل‌های جنگل تصادفی، Catboost و MLP

Table 3. Performance results of the majority-voting ensemble model based on Random Forest, CatBoost, and MLP base learners.

مدل	دقت	صحت	یادآوری	امتیاز-F1	MCC	CK
مدل گروهی رأی اکثریت	۰/۹۵۷۵	۰/۹۳۹۶	۰/۹۴۷۶	۰/۹۳۷۹	۰.۸۹۶۶	۰.۸۹۵۷

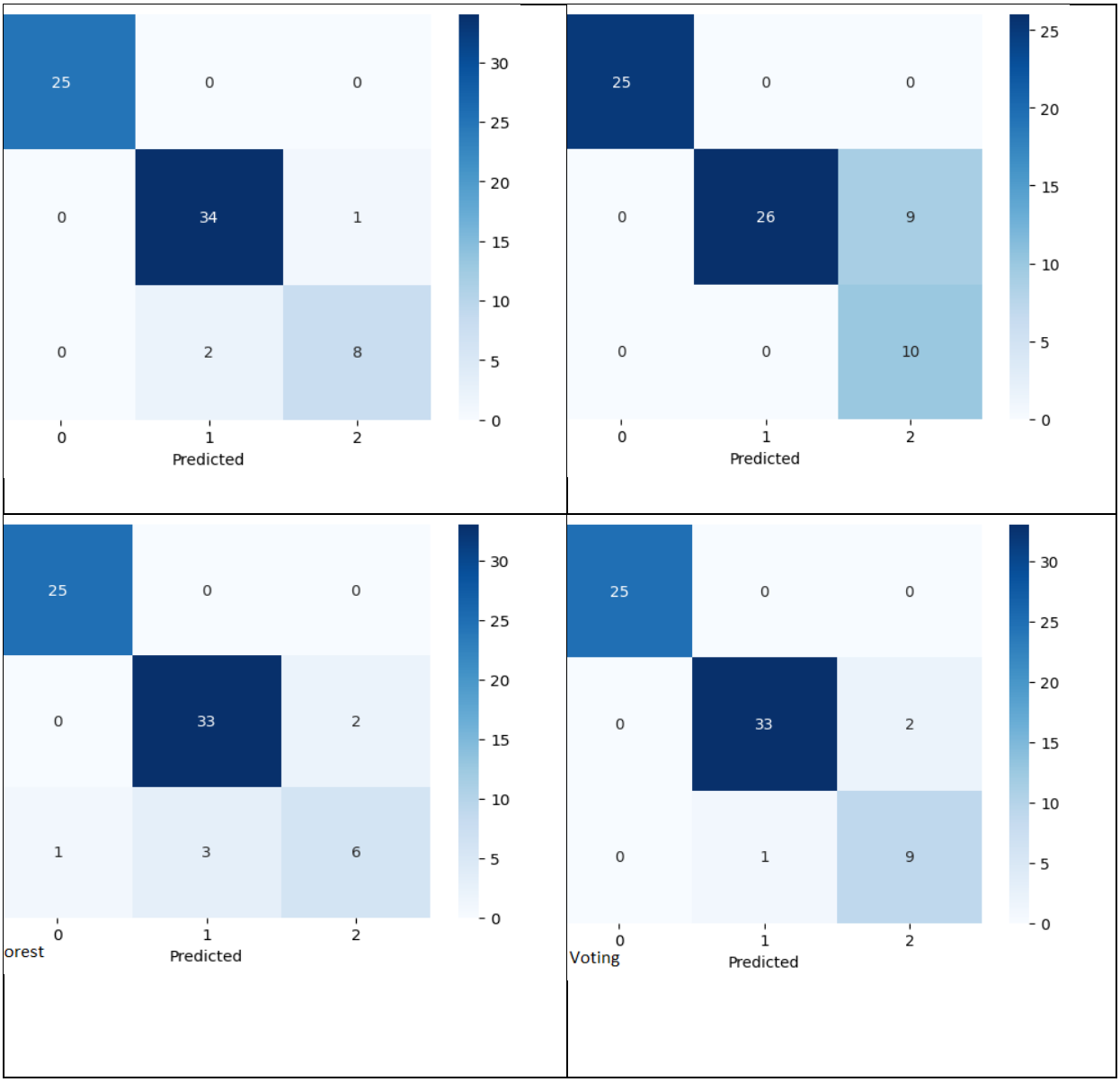
در شکل ۳ نتایج حاصل از روش TOPSIS جهت مقایسهٔ عملکرد چهار مدل ارائه شده است که محور افقی مدل‌ها (models) و محور عمودی امتیاز topsis score را نشان می‌دهد. امتیازها براساس نرمال‌سازی برداری و وزن‌دهی برابر چهار معیار محاسبه شده‌اند؛ یعنی امتیاز بیشتر، راه‌حل مطلوب‌تری در پی دارد. نتایج بیان‌کنندهٔ برتری پایدار مدل گروهی مبتنی‌بر رأی‌گیری اکثریت نسبت به بهترین مدل منفرد (CatBoost) و نیز فاصلهٔ معنادار عملکرد آن با مدل‌های RF و MLP هستند؛ بنابراین، برای دادهٔ موجود، Ensemble گزینهٔ نهایی مناسب‌تری است.



شکل ۳. مقایسهٔ امتیاز TOPSIS مدل‌ها (وزن‌های برابر برای همه شاخص‌های اصلی)

Figure 3. Comparison of models' TOPSIS scores (equal weights applied across all primary metrics).

شکل ۴ ماتریس درهم‌ریختگی (Confusion Matrix) حاصل از هر چهار مدل استفاده‌شده را ارائه می‌دهد. با توجه به شکل، این ماتریس‌ها نتایج عملکرد مدل‌ها در پیش‌بینی سطوح رضایت مشتری (ناراضی، بی‌سو، راضی) را به تصویر می‌کشند. مدل Catboost در پیش‌بینی کلاس ۱ (بی‌سو) و مدل گروهی بهترین عملکرد را در پیش‌بینی کلاس ۲ (ناراضی) دارد.

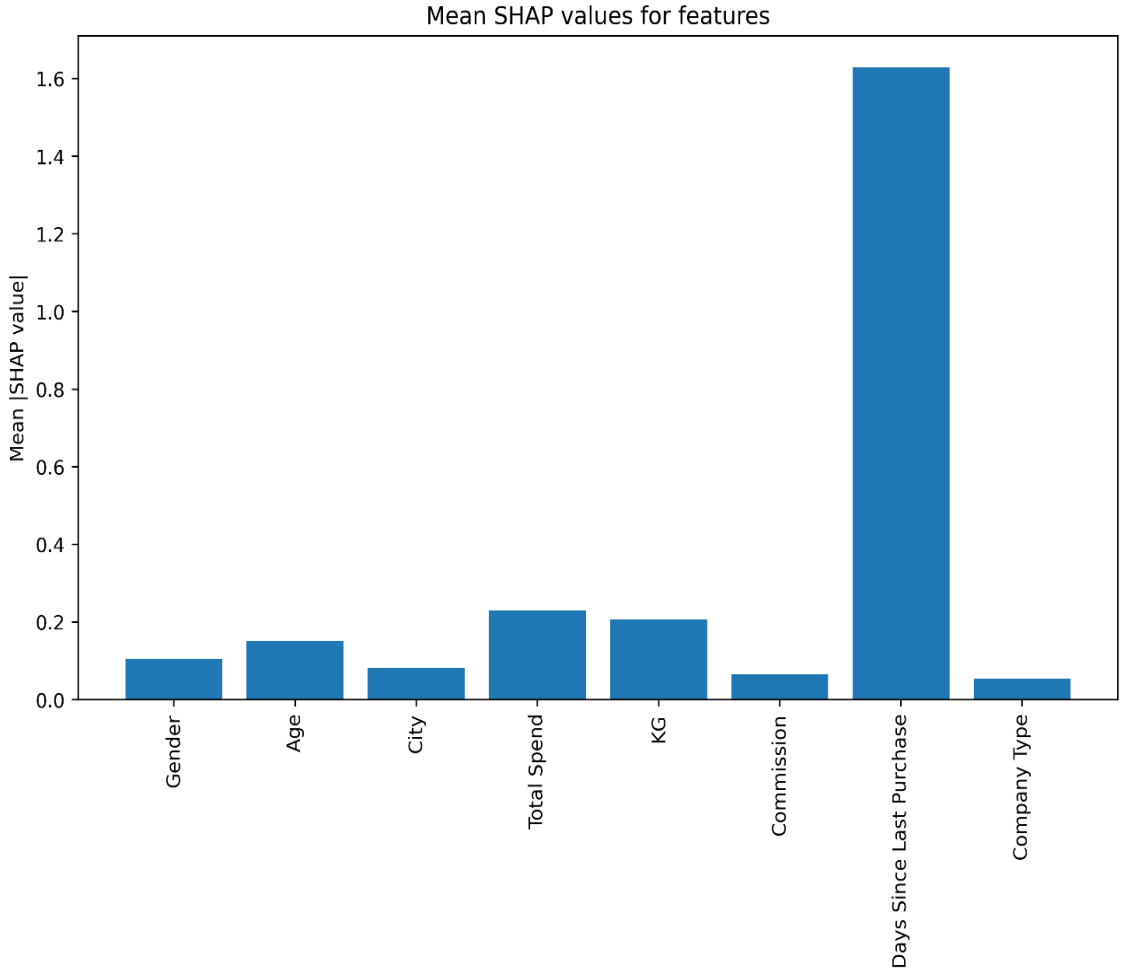


شکل ۴. ماتریس درهم‌ریختگی مربوط به هر چهار مدل اصلی

Figure 4. Confusion matrices for the four main models

به‌طور کلی، مدل CatBoost و مدل گروهی عملکرد بهتری در پیش‌بینی درست تمامی کلاس‌ها نشان داده‌اند، درحالی‌که MLP در تمایز بین نمونه‌های کلاس ۱ و ۲ ضعیف‌تر عمل کرده است.

شکل ۵ نمودار SHAP را نشان می‌دهد که بینش‌هایی دربارهٔ تأثیر ویژگی‌های مختلف بر پیش‌بینی‌های انجام‌شده را توسط مدل ترکیبی رأی‌گیری متشکل از CatBoost، جنگل تصادفی و MLP ارائه می‌دهد. مقادیر SHAP نشان‌دهندهٔ بزرگی و جهت تأثیر هر ویژگی بر خروجی مدل برای هر نمونه‌اند.



شکل ۵. میانگین قدرمطلق مقادیر SHAP برای همهٔ ویژگی‌ها

Figure 5. Global mean absolute SHAP values for all dataset features

نتایج نشان می‌دهد فاصلهٔ زمانی از آخرین خرید قوی‌ترین محرک نارضایتی است؛ هرچه این فاصله بیشتر باشد، احتمال برچسب «ناراضی» افزایش می‌یابد. شاخص‌های مرتبط با ارزش/حجم سفارش (Total Spend و KG) اثر میانی دارند و می‌توانند در کنار سیاست‌های تحویل/خدمت برای پیشگیری هدف‌گذاری شوند. در مقابل، Commission و ویژگی‌های جمعیت‌شناختی (جنسیت، سن، شهر، نوع شرکت) تأثیر اندکی در توضیح نارضایتی دارند؛ بنابراین، اولویت مداخله باید بر فعال‌سازی مجدد مشتریان کم‌تکرار و بهبود تجربهٔ عملیاتی خرید متمرکز شود، نه بر تغییرات صِرف قیمتی یا بخش‌بندی جمعیت‌شناختی.

۴- بحث و نتیجه‌گیری

این پژوهش با هدف بررسی عوامل مؤثر بر رضایت مشتری و ارائهٔ مدلی ترکیبی برای پیش‌بینی دقیق‌تر میزان رضایت انجام شد. استفاده از مدل ترکیبی که شامل CatBoost، جنگل تصادفی، و MLP بود، نتایج چشمگیری را در مقایسه با مدل‌های منفرد ارائه داد. روش رأی‌گیری اکثریت برای ترکیب خروجی مدل‌ها نشان داد که این رویکرد می‌تواند قدرت تشخیصی هر مدل را

ترکیب کند و نتایج بهتری ازنظر دقت، یادآوری و امتیاز-F1 و نهایی مدل TOPSIS به دست آورد.

۴-۱. تأثیر الگوریتم ژنتیک در بهینه‌سازی

یکی از ویژگی‌های مهم این پژوهش، استفاده از الگوریتم ژنتیک برای بهینه‌سازی ابرپارامترها بود. این روش توانست با بهره‌گیری از سازوکار جستجوی تصادفی هدفمند، محدودیت‌های روش‌های سنتی، مانند جستجوی شبکه‌ای و جستجوی تصادفی را برطرف کند. الگوریتم ژنتیک امکان جستجوی مؤثر در فضای بزرگ و پیچیدهٔ پارامترها را فراهم کرد و به نتایج بهینه‌ای دست یافت که به بهبود عملکرد مدل‌ها انجامید.

۴-۲. تحلیل نتایج مدل‌های منفرد

در این پژوهش، سه مدل CatBoost، جنگل تصادفی، و MLP بررسی و تحلیل شدند. CatBoost در تمامی معیارهای عملکرد، ازجمله دقت، صحت، بازیابی، امتیاز-F1، MCC و CK عملکرد برتری نسبت به سایر مدل‌ها داشت. این مدل به‌ویژه در بازیابی مثبت و دقت، که اهمیت بسیاری در تصمیم‌گیری‌های درست در طبقه‌بندی چندکلاسه دارند، نتایج بسیار مطلوبی را ارائه داد؛ به‌عنوان مثال، این مدل به‌خوبی توانست نمونه‌های طبقهٔ «راضی» را تشخیص دهد و میزان خطای کمی در پیش‌بینی کلاس‌ها داشت. جنگل تصادفی نیز باوجود عملکرد نسبتاً پذیرفتنی در مقایسه با CatBoost، در تمامی شاخص‌ها عملکرد ضعیف‌تری نشان داد؛ بااین‌حال، این مدل در شناسایی کلاس‌های «خنثی» و «ناراضی» رقابت نزدیکی با CatBoost داشت و توانست دقت مناسبی برای این گروه‌ها ارائه کند. در مقابل، MLP که مدلی پیچیده‌تر، اما حساس‌تر به تنظیمات است، عملکرد ضعیف‌تری نسبت به دو مدل دیگر داشت. باوجوداین، در برخی مواقع، مانند پیش‌بینی دستهٔ «خنثی»، توانست با جنگل تصادفی رقابت کند. این موضوع نشان می‌دهد که استفاده از MLP در یک مدل ترکیبی می‌تواند تکمیل‌کنندهٔ پیش‌بینی‌های دیگر مدل‌ها باشد.

۴-۳. تحلیل نتایج مدل گروهی

برای بهبود دقت و عملکرد کلی پیش‌بینی، از یک مدل گروهی رأی اکثریت استفاده شد که ترکیبی از پیش‌بینی‌های سه مدل CatBoost، جنگل تصادفی، و MLP است. این مدل با ترکیب خروجی‌های مختلف، توانست عملکرد کلی بهتری ارائه دهد.

طبق جدول ۳، مدل گروهی رأی اکثریت در معیار دقت، صحت، بازیابی، و امتیاز-F1 از سایر مدل‌ها پیشی گرفت؛ برای نمونه، دقت مدل گروهی ۰/۹۵۷۵ است که در مقایسه با دقت ۰/۹۵۴۰ مدل CatBoost کمی بهتر است. این افزایش جزئی در دقت، نشان‌دهندهٔ توانایی مدل گروهی در ترکیب مزایای مدل‌های منفرد است. ازسوی‌دیگر، بازیابی مدل گروهی برابر ۰/۹۴۷۶ بود که در مقایسه با ۰/۹۲۷۱ مدل CatBoost بهبود چشمگیری را نشان داد. این موضوع نشان‌دهندهٔ قدرت مدل گروهی در شناسایی تمامی موارد درست در هر سه کلاس است. نکتهٔ مهم دیگر، امتیاز-F1 مدل گروهی است که معادل ۰/۹۳۷۹ بوده و در مقایسه با ۰/۹۳۶۳ مدل CatBoost عملکرد بهتری داشته است. مقادیر ۰/۸۹۶۶ برای ضریب همبستگی متیوز (MCC) و ۰/۸۹۵۷ برای ضریب کاپای کوهن (CK) که مدل گروهی (Ensemble) در بازه توافق قوی (Strong Agreement) و بسیار نزدیک به حداکثر مقدار ممکن

(۱.۰) قرار دارند، تأیید می‌کنند که مدل نهایی نه‌تنها در پیش‌بینی کلی (Accuracy) موفق است، بلکه در شرایط واقعی و احتمالاً نامتعادل داده‌ها نیز پایدار است؛ چراکه هر دو شاخص، همهٔ انواع خطاهای طبقه‌بندی را به دقت محاسبه و اثر طبقه‌بندی تصادفی را حذف می‌کنند. این نتایج با ارزیابی چندمعیاره نیز تأیید شد. با به‌کارگیری روش TOPSIS و وزن‌دهی برابر به هر شش معیار، امتیاز نزدیکی مدل‌ها به راه‌حل ایدئال برای Ensemble، Catboost، Forest Random و MLP به‌ترتیب ۰/۹۶۹۵، ۰/۹۱۸۷، ۰/۳۴۳۷ و ۰/۱۰۳۷ به دست آمد؛ بنابراین، مدل گروهی رأی اکثریت رتبهٔ اول را کسب کرد و پس‌از آن CatBoost قرار گرفت. این شاخص نشان‌دهندهٔ تعادل بهتر مدل گروهی میان دقت و بازیابی است که منجر به پیش‌بینی متعادل‌تری برای تمامی کلاس‌ها شده است. براساس ماتریس سردرگمی مدل گروهی، می‌توان مشاهده کرد که این مدل توانسته است تمامی نمونه‌های دستهٔ «ناراضی» را به‌درستی پیش‌بینی کند (True Positive = ۲۵) و هیچ خطایی در این دسته رخ نداده است. در دستهٔ «خنثی»، مدل گروهی عملکرد خوبی داشته و از ۳۶ نمونه، فقط ۲ نمونه به‌اشتباه در دستهٔ «راضی» قرار گرفته‌اند. در نهایت، در دستهٔ «راضی»، عملکرد مدل گروهی نسبت به دو دستهٔ دیگر اندکی ضعیف‌تر است؛ زیرا ۱ نمونه از این دسته به‌اشتباه در دستهٔ «خنثی» طبقه‌بندی و میزان خطای بیشتری برای این دسته مشاهده شده است.

۴-۴. تحلیل ویژگی‌ها با SHAP

در شکل ۴ رتبه‌بندی جهانی اهمیت ویژگی‌ها بر مبنای میانگین قدرمطلق SHAP در کل کلاس‌ها ارائه می‌شود؛ بنابراین، این شکل فقط شدت اثر هر ویژگی را نشان می‌دهد و جهت اثر (مثبت/منفی) را باید در نمودارهای کلاس‌به‌کلاس دید. مطابق این شکل ۴، «روزهای گذشته از آخرین خرید» مهم‌ترین سیگنال پیش‌بینی است و پس‌از آن شاخص‌های ارزش/حجم سبد، مانند «وزن/حجم خرید ’KG‘» و «Total Spend» در رده‌های بعد قرار می‌گیرند؛ در مقابل، متغیرهای جمعیت‌شناختی (سن/جنسیت/شهر/نوع شرکت) و کمیسیون کم‌اثرترند. این الگو با مبانی نظری پژوهش نیز هم‌سوست که نشان می‌دهد ابعاد عملیاتی و شدت تعامل از عوامل قیمتی و جمعیت‌شناختی، در تبیین رضایت آنلاین اثرگذارترند.

۵-۴. مقایسهٔ نتایج با مبانی نظری موضوع

در قیاس با مطالعهٔ زغلول (Zaghloul, 2024) که رضایت آینده را از روی مرورها/امتیازها و نیز داده‌ای بسیار بزرگ (>۱۰۰هزار سفارش) پیش‌بینی کرده و با اتکا به ویژگی‌های عملیاتی کلیدی مثل زمان تحویل، ارزش سفارش و موقعیت، به دقت حدود ۹۲درصد (برتری RF نسبت به DL) رسیده است، پژوهش حاضر بر داده‌های عملیاتیِ سفارش در یک بافت B2B و با نمونهٔ کوچک‌تر (۳۴۹ رکورد) انجام شد، اما با Ensemble رأی‌گیری (CatBoost+RF+MLP) به Accuracy=95.75% و توازن بهتر در معیارهای Recall/F1 دست یافت؛ ضمن آنکه TOPSIS با وزن برابر نیز برتری جمعی Ensemble را تأیید می‌کند.

نتایج این پژوهش نشان می‌دهد که «وزن کل محصولات خریداری‌شده» را مهم‌ترین متغیر نشان می‌دهد، با جریان پژوهشی مبتنی‌بر شاخص‌های رفتاری–تراکنشی هم‌سوست؛ مطالعات اخیر که از چارچوب RFM و مدل‌های پیش‌بینی استفاده می‌کنند، مؤلفهٔ Monetary/ارزش سبد را از قوی‌ترین

سیگنال‌های رفتار و پیامدهای پسینی مشتری (رضایت/نگهداشت) گزارش کرده‌اند (Tang, 2025). در همین راستا، شواهد بخشی (خرده‌فروشی و حمل‌ونقل) نیز نشان می‌دهد ابعاد عملیاتی سفارش و شدت تعامل مشتری (که وزن/ارزش خرید نماد آن است) با برآورد رضایت هم‌بستگی معناداری دارد (Hui et al., 2025). دربارهٔ «روزهای گذشته از آخرین خرید»، الگوی غالب منفی در نتایج این پژوهش (و گاه اثر معکوس در تعامل با سایر ویژگی‌ها) با مبانی نظری Recency انطباق دارد: هرچه فاصلهٔ زمانی میان خریدها بیشتر باشد، به‌عنوان نشانه‌ای از کاهش درگیری یا رضایت مشتری تلقی می‌شود؛ بااین‌حال، پژوهش‌ها تأکید می‌کنند این اثر می‌تواند وابسته به زمینه و برهم‌کنش با سایر متغیرها باشد –دقیقاً همان چیزی که در SHAP مشاهده کرده‌ایم (Zaghloul et al., 2025).

۴-۶. یافته‌های غیرمنتظره

یکی از یافته‌های جالب این پژوهش، تأثیر کمِ عواملی، مانند «کمیسیون» و «مجموع هزینه‌های انجام‌شده» بر رضایت مشتری بود. این نتایج نشان می‌دهد که عوامل مالی به‌طور مستقیم، تأثیر کمتری بر رضایت مشتری دارند و رضایت بیشتر به تجربیات ذهنی و تعاملاتی مشتری با شرکت وابسته است. این یافته اهمیت توسعهٔ راهبردهایی را که بر تجربیات مثبت مشتریان تمرکز دارند، برجسته می‌کند.

۴-۷. اهمیت مدیریتی و راهبردی

این پژوهش می‌تواند نقش کلیدی در تدوین راهبردهای بازاریابی و مدیریت ارتباط با مشتریان ایفا کند. نتایج مدل ترکیبی به مدیران فروشگاه‌های آنلاین کمک می‌کند تا با تمرکز بر ویژگی‌های کلیدی، میزان رضایت مشتریان را بهبود بخشند. یافته‌های ما نشان می‌دهد Recency (‘روزهای گذشته از آخرین خرید) قوی‌ترین محرک رضایت/نارضایتی است و پس‌از آن، شاخص‌های ارزش/حجم سبد (KG ،Total Spend) قرار می‌گیرند، درحالی‌که کمیسیون/هزینه تأثیرگذار است؛ اما به‌طور معناداری کم‌اهمیت‌تر از عوامل عملیاتی است.

این پژوهش نشان داد که مدل ترکیبی مبتنی‌بر رأی‌گیری اکثریت، شامل CatBoost، جنگل تصادفی و شبکه عصبی مصنوعی، که پارامترهای آن با استفاده از الگوریتم ژنتیک بهینه‌سازی شده‌اند، با توجه به نتایج مقایسهٔ مدل‌ها با استفاده از TOPSIS، عملکرد موفقی در بهبود دقت و صحت پیش‌بینی در مسائل طبقه‌بندی چندکلاسه، مانند پیش‌بینی رضایت مشتری، دارد.

این مدل در شاخص‌های کلیدی، مانند دقت، یادآوری و امتیاز F1 نسبت به مدل‌های منفرد برتری داشت و توانست در دسته‌بندی سطوح مختلف رضایت مشتری نتایج مطلوبی ارائه دهد. تحلیل SHAP نیز نشان داد که «وزن کل محصولات خریداری‌شده» قوی‌ترین ویژگی تأثیرگذار بر رضایت مشتری است، درحالی‌که ویژگی‌هایی نظیر «کمیسیون» و «مجموع هزینه‌های انجام‌شده» اهمیت کمتری داشتند. این یافته‌ها می‌تواند راهنمایی برای طراحی راهبردهای بهینه در بازاریابی و افزایش تعاملات مشتری باشد. این پژوهش بر داده‌ای کوچک و تک‌سازمانی (عمده‌فروشی B2B در ایران) متکی است؛ برچسب «رضایت» به‌صورت چندکلاسهٔ ترتیبی ساخته شده، اما از مدل‌های ترتیبی اختصاصی استفاده نشده است. جایگزینی مقادیر گمشده با روش‌های ساده انجام شد، درحالی‌که ویژگی‌های عملیاتی ریزدانه (شاخص‌های دقیق تحویل/ مرجوعی/

اعلان وضعیت) و متن بازخوردها در دسترس نبوده‌اند؛ بنابراین، خطر بیش‌برازش و محدودیت در تعمیم‌پذیری وجود دارد. بر همین مبنا، برای پژوهش‌های آینده پیشنهاد می‌شود گردآوری نمونه‌ای بزرگ‌تر و چندسازمانی همراه با ارزیابی بیرونی در بازه‌های زمانی متفاوت مدنظر قرار گیرد. همچنین، به‌کارگیری مدل‌های ترتیبی (Ordinal)، بهبود مدیریت داده‌های گمشده با استفاده از Multiple Imputation، غنی‌سازی داده‌ها با سیگنال‌های لجستیکی و عملیاتی، ادغام داده‌های عملیاتی با تحلیل متن دیدگاه‌ها، و اجرای پایلوت‌های میدانی هدفمند برای سنجش اثر مداخلات مبتنی‌بر Recency و ارزش/وزن سبد بر شاخص‌های رضایت و رفتار خرید پیشنهاد می‌شود.

این مسیرها مستقیماً محدودیت‌های فعلی را هدف می‌گیرند و کاربردپذیری مدیریتی نتایج را افزایش می‌دهند.

۵- منابع

اسمعیل‌پور، امیرحسین، عاملی، مریم، مزدگیر، اشکان، احمدی ارد، و ضرابی، مرتضی (۱۴۰۳). پیش‌بینی لخته شدن خون بند ناف پیش‌از جمع‌آوری با کمک الگوریتم‌های یادگیری ماشین پیشرفته. *فصلنامه پژوهشی خون*، *۲۱*(۲)، ۱۵۹-۱۵۱.
http://bloodjournal.ir/article-1-1520-en.html

نوروزی، محمد، ماستری فراهانی، محمد، و سعادت، مهدی (۱۴۰۲). تحلیل رضایت مشتریان با رویکرد متن‌کاوی (مورد مطالعه: مشتریان عسل‌های ارگانیک). *تحقیقات بازاریابی نوین*، *۱۳*(۳)، ۴۹-۷۲.
https://doi.org/10.22108/nmrj.2023.137845.2908

References

Abedini, M., Moeini, H., & Abbasi, R. (2025). Challenges in e-commerce adoption within the Iranian paper industry: A barrier analysis. *Knowledge Economy Studies*, *2*(2), 71–80. https://doi.org/10.22034/kes.2025.2070230.1076

Adnane, Y. I., & Zerari, M. (2024). Optimizing business intelligence classification rule mining using quantum-inspired genetic algorithm. *IEEE Access*, *12*, 137041–137053. https://doi.org/10.1109/ACCESS.2024.3463506

Amanda Ardhani, D., & Tania, K. D. (2025). Knowledge discovery on e-commerce customer churn using interpretable machine learning: A comparative study of SHAP-based classifiers. *Journal of Applied Informatics and Computing*, *9*(5), 2695–2702. https://doi.org/10.30871/jaic.v9i5.10811

Bakır, H., & Ceviz, Ö. (2024). Empirical enhancement of intrusion detection systems: A comprehensive approach with genetic algorithm-based hyperparameter tuning and hybrid feature selection. *Arabian Journal for Science and Engineering*, *49*(9), 13025–13043. https://doi.org/10.1007/s13369-024-08949-z

Balakrishnan, V., Shi, Z., Law, C. L., Lim, R., Teh, L. L., & Fan, Y. (2022). A deep learning approach in predicting products' sentiment ratings: A comparative analysis. *The Journal of Supercomputing*, *78*(5), 7206–7226. https://doi.org/10.1007/s11227-021-04169-6

Bellar, O., Baina, A., & Ballafkih, M. (2024). Sentiment analysis: Predicting product reviews for e-commerce recommendations using deep learning and transformers. *Mathematics*, *12*(15), 2403. https://doi.org/10.3390/math12152403

Cavalcante Siebert, L., Bianchi Filho, J. F., Silva Júnior, E. J. d., Kazumi Yamakawa, E., & Catapan, A. (2021). Predicting customer satisfaction for distribution companies using machine learning. *International Journal of Energy Sector Management*, *15*(4), 743–764. https://doi.org/10.1108/IJESM-10-2018-0007

Chandra, D., & Fitriyanto, A. (2024). An

empirical analysis of student satisfaction with lecturer teaching quality: Applying the expectation-disconfirmation theory. *Indonesian Journal of Economics and Management*, *4*(3), 460–474. https://doi.org/10.35313/ijem.v4i3.6425

Chen, L., Nan, G., & Li, M. (2018). Wholesale pricing or agency pricing on online retail platforms: The effects of customer loyalty. *International Journal of Electronic Commerce*, *22*(4), 576–608. https://doi.org/10.1080/10864415.2018.1485086

Choudhary, N., Huang, E. W., Subbian, K., & Reddy, C. K. (2024). *An interpretable ensemble of graph and language models for improving search relevance in e-commerce*. In Companion Proceedings of the ACM Web Conference 2024 (WWW '24 Companion) (pp. 206–215). Association for Computing Machinery. https://doi.org/10.1145/3589335.3648318

Davoodi, L., Mezei, J., & Heikkilä, M. (2025). Aspect-based sentiment classification of user reviews to understand customer satisfaction of e-commerce platforms. *Electronic Commerce Research, 26,* 1417–1459. https://doi.org/10.1007/s10660-025-09948-4

Efendi, A., Fitri, I., & Nurcahyo, G. W. (2024). Improvement of machine learning algorithms with hyperparameter tuning on various datasets. In *IEEE International Conference on Artificial Intelligence*. https://doi.org/10.1109/ICFTSS61109.2024.10691354

Emami, A., Farshad Bakhshayesh, E., & Rexhepi, G. (2023). Iranian communities e-business challenges and value proposition design. *Journal of Enterprising Communities: People and Places in the Global Economy*, *17*(2), 479–497. https://ideas.repec.org/a/eme/jecpps/jec-09-2021-0141.html

Esmaielpour, A. H., Ameli, M., Mozdgir, A., Ahmadi, O., & Zarabi, M. (2024). Predicting pre-collection umbilical cord blood clotting using advanced machine learning algorithms. *Journal of Iranian Blood Transfusion*, *21*(2), 151. http://bloodjournal.ir/article-1-1520-en.html [In Persian]

Esmaielpour, A. H., Ameli, M., Mozdgir, A., Ahmadi, O., & Zarabi, M. (2026). Prenatal prediction of umbilical cord blood CD34(+) adequacy and identifying influential factors via an ensemble machine learning and TOPSIS ranking: A Retrospective study. *Cell Journal*, *27*(1), 1–10. https://doi.org/10.22074/cellj.2025.2064741.1878

Fouad, M., Barakat, S., & Rezk, A. (2023). *Effective e-commerce based on predicting the level of consumer satisfaction*. International Conference on Advanced Intelligent Systems and Informatics. https://doi.org/10.1007/978-981-99-4764-5_17

Ghorban Tanhaei, H., Boozary, P., Sheykhan, S., Rabiee, M., Rahmani, F., & Hosseini, I. (2024). Predictive analytics in customer behavior: Anticipating trends and preferences. *Results in Control and Optimization*, *17*, 100462. https://doi.org/10.1016/j.rico.2024.100462

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Hui, G., Al Mamun, A., Reza, M. N. H., & Hussain, W. M. H. W. (2025). An empirical study on logistic service quality, customer satisfaction, and cross-border repurchase intention. *Heliyon*, *11*(1), e41156. https://doi.org/10.1016/j.heliyon.2024.e41156

Kampezidou, S. I., Tikayat Ray, A., Bhat, A. P., Pinon Fischer, O., & Mavris, D. (2024). Fundamental components and principles of supervised machine learning workflows with numerical and categorical data. *Eng*, *5*, 384–416. https://doi.org/10.3390/eng5010021

Kumar, R., Mukherjee, S., & Choudhary, D. (2025). Uncovering the roots of customer dissatisfaction via Amazon reviews: A hybrid ensemble-deep learning approach for e-commerce quality management. *Annals of Operations Research*, *353*(2), 545–574. https://doi.org/10.1007/s10479-025-06770-x

Lamane, H., Mouhir, L., Moussadek, R., Baghdad, B., Kisi, O., & El Bilali, A. (2025). Interpreting machine learning models based on SHAP values in predicting suspended sediment concentration. *International Journal of Sediment Research*, *40*(1), 91–107. https://doi.org/10.1016/j.ijsrc.2024.10.002

Lin, S., Zheng, H., Han, B., Li, Y., Han, C., & Li, W. (2022). Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction. *Acta Geotechnica*. *17*, 1477–1502. https://doi.org/10.1007/s11440-021-01440-1

Mamta, K., & Sangwan, S. (2024). AaPiDL: An ensemble deep learning-based predictive framework for analyzing customer behaviour and enhancing sales in e-commerce systems. *International Journal of Information Technology*, *16*(5), 3019–3025. https://doi.org/10.1007/s41870-024-01796-z

Mirzahossein, H., & Rezashoar, S. (2025). Feature importance analysis of optimized machine learning modeling for predicting customers satisfaction at the United States Airlines. *Machine Learning with Applications*, *22*, 100734. https://doi.org/10.1016/j.mlwa.2025.100734

Nematzadeh, S., Kiani, F., Torkamanian-Afshar, M., & Aydin, N. (2022). Tuning hyperparameters of machine learning algorithms and deep neural networks using metaheuristics: A bioinformatics study on biomedical and biological cases. *Computational Biology and Chemistry*, *97*, 107619. https://doi.org/10.1016/j.compbiolchem.2021.107619

Nikghadam Hojjati, S., & Reza Rabi, A. (2013). Effects of Iranian online behavior on the acceptance of internet banking. *Journal of Asia Business Studies*, *7*(2), 123–139. https://doi.org/10.1108/15587891311319422

Nisar, T. M., & Prabhakar, G. (2017). What factors determine e-satisfaction and consumer spending in e-commerce retailing? *Journal of Retailing and Consumer Services*, *39*, 135–144. https://doi.org/10.1016/j.jretconser.2017.07.010

Noruzi, M., Masteri Farahani, M., & Saadat, M. (2023). Customer satisfaction analysis with a text mining approach (Case study: Customers of organic honeys). *New Marketing Research Journal*, *13*(3), 49–72. https://doi.org/10.22108/nmrj.2023.137845.2908 [In Persian]

Nurul Ain, M., Maslina Abdul, A., & Shuzlina Abdul, R. (2024). Predicting consumer behavior in e-commerce using decision tree: A case study in Malaysia. *Information Management and Business Review*, *16*(3), 201-209. https://ideas.repec.org/a/rnd/arimbr/v16y2024i3p201-209.html

Obafemi, O., Onyebuchi, O., & Oiku, O. P. (2023). Impact of customer loyalty on organizational performance. *IIARD International Journal of Economics and Business Management*, *8*, 56–62. https://doi.org/10.56201/ijebm.v8.no5.2022.pg56.62

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460–469. https://doi.org/10.1177/002224378001700405

Quintus, M., Mayr, K., Hofer, K. M., & Chiu, Y. T. (2024). Managing consumer trust in e-commerce: Evidence from advanced versus emerging markets. *International Journal of Retail & Distribution Management*, *52*(10-11), 1038–1056. https://doi.org/10.1108/ijrdm-10-2023-0609

Rane, N., Paramesha, M., Choudhary, S., & Rane, J. (2024). Artificial intelligence, machine learning, and deep learning for advanced business strategies: A review. *Partners Universal International Innovation Journal* *2*(03), 147–171. https://doi.org/10.5281/zenodo.12208298

Reddy, B. H., Vickram, A. S., & Karthikeyan, P. R. (2024). *Classification of fire and smoke images using random forest algorithm in comparison with decision tree to measure accuracy, precision, recall, F-score*. In AIP Conference Proceedings (Vol. 2853, No. 1, Article 020077). AIP Publishing. https://doi.org/10.1063/5.0198489

Shen, X., Yin, F., & Jiao, C. (2023). Predictive models of life satisfaction in older people: A machine learning approach. *International Journal of Environmental Research and Public Health*, *20*(3), 2445. https://doi.org/10.3390/ijerph20032445

Sinemus, K., Zielke, S., & Dobbelstein, T. (2025). Improving consumer satisfaction through shopping app features: A Kano-based approach. *Journal of Retailing and Consumer Services*, *85*, 104243. https://doi.org/10.1016/j.jretconser.2025.104243

Taherdoost, H., & Madanchian, M. (2024). A comprehensive survey and literature review on TOPSIS. *International Journal of Service Science, Management, Engineering, and Technology*, *15*(1). https://doi.org/10.4018/IJSSMET.347947

Tang, J. (2025). Unlocking retail insights: Predictive modeling and customer segmentation through data analytics. *Journal of Theoretical and Applied Electronic Commerce Research*, *20*(2), 59. https://doi.org/10.3390/jtaer20020059

Zaghloul, M., Barakat, S., & Rezk, A. (2024). Predicting e-commerce customer satisfaction: Traditional machine learning vs. deep learning approaches. *Journal of Retailing and Consumer Services*, *79*, 103865. https://doi.org/10.1016/j.jretconser.2024.103865

Zaghloul, M., Barakat, S., & Rezk, A. (2025). Enhancing customer retention in online retail through churn prediction: A hybrid RFM, K-means, and deep neural network approach. *Expert Systems with Applications*, *290*, 128465. https://doi.org/10.1016/j.eswa.2025.128465

Zhang, J., Yu, Q., Chen, Y., Zhou, G., Liu, Y., Sun, Y., Liang, C., Huzhang, G., Ni, Y., Zeng, A., & Yu, H. (2024). *An e-commerce dataset revealing variations during sales*. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1162–1171). Association for Computing Machinery. https://doi.org/10.1145/3626772.3657870

Zhang, X., & Guo, C. (2024). Research on multimodal prediction of e-commerce customer satisfaction driven by big data. *Applied Sciences*, *14*(18), 8181. https://doi.org/10.3390/app14188181