

Article

Customer Satisfaction in City Delivery Systems and Its Implications for Delivery Efficiency and Environmental Impacts: A Machine Learning Analysis

Adisa Medić ^{1,*} , Amel Kosovac ¹ , Ermin Muharemović ¹ , Mladen Krstić ² , Muhamed Begović ¹ ,
Snežana Tadić ²  and Aida Kalem ¹ 

¹ Faculty of Traffic and Communications, University of Sarajevo, Zmaja od Bosne 8, 71000 Sarajevo, Bosnia and Herzegovina; amel.kosovac@fsk.unsa.ba (A.K.); ermin.muharemovic@fsk.unsa.ba (E.M.); muhamed.begovic@fsk.unsa.ba (M.B.); aida.kalem@fsk.unsa.ba (A.K.)

² Logistics Department, Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Belgrade, Serbia; m.krstic@sf.bg.ac.rs (M.K.); s.tadic@sf.bg.ac.rs (S.T.)

* Correspondence: adisa.medic@fsk.unsa.ba

Abstract

The rapid growth of e-commerce has intensified last-mile delivery activities in urban areas, creating challenges for city logistics systems related to operational efficiency, congestion, and environmental impacts. In this context, understanding the factors that influence customer satisfaction with logistics operators is increasingly important, as mismatches between customer expectations and delivery service characteristics may lead to operational inefficiencies such as failed delivery attempts and repeated delivery rounds. This study proposes a machine learning framework for predicting customer satisfaction with postal and logistics operators in urban delivery systems using survey data on customer characteristics, preferences, and service perceptions. Several machine learning algorithms were developed and evaluated to identify the key determinants of customer satisfaction and assess their predictive performance. Beyond predictive accuracy, the study interprets customer satisfaction as an indicator of the alignment between customer expectations and delivery service configurations. Improved alignment may support service configurations that reduce delivery mismatches and repeated delivery attempts, which are recognized as a significant source of additional transport activity in urban freight systems. By identifying customer segments whose expectations are not adequately addressed by existing delivery services, the proposed framework can support more informed service design and operational decision-making. From a city logistics perspective, the potential reduction in failed deliveries and repeated delivery rounds may contribute to lower vehicle kilometers travelled, congestion, energy consumption, and emissions associated with urban freight transport, although these operational and environmental indicators were not directly measured in this study. The proposed approach therefore provides a data-driven decision-support tool that can help operators improve service quality and serve as a basis for future integration with operational and environmental indicators in sustainable last-mile delivery planning.

Keywords: machine learning; postal; city logistics; operator selection; urban sustainability; predictive modeling; customer satisfaction



Academic Editor: Maria Rosa Trovato

Received: 23 June 2026

Revised: 21 July 2026

Accepted: 22 July 2026

Published: 1 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The rapid growth of e-commerce has significantly transformed urban freight transport systems, particularly in relation to last-mile delivery operations. Increasing volumes of small parcel shipments have intensified delivery activities in densely populated urban areas, contributing to higher traffic congestion, increased vehicle kilometers travelled, and greater environmental pressure associated with urban freight transport [1–3]. As cities continue to experience growth in online shopping demand, improving the efficiency and sustainability of last-mile logistics has become a major challenge for both logistics operators and urban planners.

One of the key operational challenges in urban parcel delivery systems is the occurrence of failed delivery attempts. When deliveries cannot be successfully completed during the first attempt, additional delivery rounds are required, increasing operational costs and generating additional transport activity within already congested urban environments. Repeated delivery attempts therefore represent an important source of inefficiency in last-mile logistics and contribute to increased emissions and environmental pressures associated with urban freight transport [4,5]. In this context, customer behavior and expectations play a significant role in shaping the performance of urban delivery systems [6]. Misalignment between customer preferences and service characteristics may increase the likelihood of delivery disruptions and unsuccessful delivery attempts. Therefore, understanding the determinants of customer satisfaction can provide valuable insights for improving delivery service design and operational planning in city logistics systems [7–9].

However, in urban contexts, customer-related decisions extend beyond market outcomes [10]. Customer choices regarding logistics operators and delivery options may indirectly influence delivery success rates and the frequency of repeated delivery attempts, thereby affecting overall transport activity. In densely populated urban environments, even small variations in delivery success can lead to significant changes in vehicle kilometers travelled and the intensity of freight movements. Consequently, customer behavior should be considered as a driver of urban logistics system performance, with implications for both operational efficiency and environmental impact [11,12].

While contemporary logistics increasingly adopts customer-oriented approaches supported by CRM and data-driven tools [13,14], these approaches are still predominantly applied in the context of improving operator performance. In urban delivery systems, however, the relationship between customer expectations and service configurations has broader implications, as inconsistencies between them may contribute to inefficient delivery processes and increased transport activity.

Despite research on sustainable urban logistics and freight-related environmental impacts [15,16], the role of customer-driven operator selection in shaping these outcomes remains insufficiently addressed. Existing machine learning applications are largely fragmented and predominantly oriented towards operator side optimization, with limited emphasis on customer satisfaction as a predictive and system relevant variable.

These limitations point to the need for approaches that explicitly integrate customer behavior into the analysis of urban logistics systems. Accordingly, this study develops a customer profile based model for postal and logistics operator selection, with the aim of improving the alignment between demand and service provision. By positioning customer satisfaction as a predictive component, the proposed approach contributes to a better understanding of demand allocation mechanisms in urban freight systems and their potential indirect implications for transport efficiency and environmental pressure. Accordingly, the empirical scope of this study is limited to the prediction of customer satisfaction and the selection of postal and logistics operators. The proposed relationships with delivery efficiency, failed and repeated delivery attempts, transport activity, and environmental

outcomes are conceptually and theoretically grounded, but they are not empirically validated within the present study because the corresponding operational and environmental indicators were not included in the available dataset.

The paper is structured into six sections. Following this introduction, Section 2 provides a review of relevant literature on city logistics, operator performance indicators, environmental impacts, customer profiling, and the application of machine learning in postal and logistics systems. Section 3 describes the research methodology, including data collection, preprocessing, and model construction. Section 4 presents the results and model performance evaluation. Section 5 discusses the research findings, with particular emphasis on sustainability implications and decision-support relevance for operators and users. Finally, Section 6 summarizes the main conclusions of the study, outlines its limitations, and suggests directions for future research.

2. Literature Review

City logistics is no longer just an operational segment of supply chains. In modern cities, it takes place within a complex environment shaped by intensified consumption patterns, spatial constraints and growing concerns about air quality and transport emissions. Postal and logistics activities, while essential for the functioning of the urban economy, simultaneously contribute to cumulative pressures on mobility systems and environmental performance. Their role is therefore twofold: they enable urban life but also affect its ecological balance [1,4,17].

These interdependencies cannot be examined from a single analytical angle. A more layered perspective is needed, one that views operators, customers, performance structures and technological instruments as interconnected components of a wider urban system. In this context, urban and sustainability aspects are not treated as a separate analytical layer, but are embedded within each thematic segment of the review. Accordingly, the literature review is structured into three interrelated sections. The first section examines postal and logistics operators together with the conceptual foundations of customer profiling, emphasizing their relevance within urban service environments. The second section analyzes performance indicators, considering them not only as internal management measures but also as variables with wider operational and sustainability implications. The third section addresses the application of machine learning in postal and logistics systems, focusing on its predictive capabilities and its role in supporting structured decision-making in complex urban contexts.

2.1. Roles of Postal and Logistics Operators and Customer Profile Definition

Postal and logistics systems represent a key link between supply and demand, enabling the movement of shipments from sender to recipient and supporting the continuity of supply chains [18]. Logistics operators play a central role by providing transport, warehousing, and increasingly integrated services adapted to diverse customer requirements [19,20]. In urban environments, these systems extend beyond purely operational functions, as delivery activities interact with traffic conditions, spatial constraints, and environmental factors. As a result, both operators and users can be understood as participants in broader urban systems whose actions influence congestion patterns, the spatial distribution of freight flows, and local environmental exposure [1,2,4,21].

The structure and capabilities of logistics operators vary depending on the level of service integration, ranging from basic service provision to advanced coordination of supply chain activities supported by information and communication technologies (ICT) [22,23]. Similarly, postal and courier services operate within different institutional and market frameworks, from public service provision to competitive courier, express, and

parcel (CEP) markets. These differences affect service flexibility, delivery organization, and responsiveness to customer demand.

Customer segments in postal and logistics systems generate different types of delivery flows (e.g., B2B, B2C, C2C), which vary in volume, spatial distribution, and service requirements [24,25]. These heterogeneous demand patterns reflect the diversity of user needs and represent an important component in the organization of urban freight systems.

Customer profiling provides a structured basis for analyzing user behavior in postal and logistics systems. These profiles are commonly defined through demographic and geodemographic characteristics that influence service usage patterns [26,27]. Previous research indicates that factors such as gender, age (for individuals) or years of operation (for legal entities), status, income, and education affect the adoption and intensity of e-commerce and delivery services [28–31], while spatial context further shapes service use depending on urban or rural environments and the availability of alternative channels [32–35]. These demographic and geodemographic characteristics provide a general understanding of service usage patterns; however, they do not fully explain customer decision-making. Therefore, the literature emphasizes service related requirements and preferences as key determinants of operator selection and satisfaction. Among these, delivery time is frequently emphasized as a key determinant of customer satisfaction [7,36], followed by shipment security [37] and price [7], while the use of ICT is increasingly recognized as a relevant factor in service evaluation [38]. In addition to these core factors, the literature also highlights the role of convenience, reliability, and operator accessibility in shaping customer satisfaction [9]. Further relevant aspects include strategic partnerships, employee and information quality, as well as network coverage and service range, which contribute to overall service evaluation [7,38].

Although the existing literature identifies relevant demographic, geodemographic, and service-related factors, these elements have not been systematically integrated into a formal customer profile suitable for quantitative modelling. Building on these literature-derived dimensions, this study develops a structured mathematical representation of the customer profile in Section 3.2.1.

2.2. Performance Indicators of Postal and Logistics Operators and Their Sustainability Implications

Performance indicators are used to evaluate service provision and support decision-making in postal and logistics systems, including their broader operational and environmental implications [39]. In the literature, they are commonly grouped into operational, financial, and customer-oriented dimensions, reflecting different aspects of system performance [40,41]. Among these, customer satisfaction and profiling are increasingly examined due to their relationship with service quality, loyalty, and long-term business outcomes [42–45]. Previous studies show that delivery performance, reliability, and information transparency influence customer perceptions and service evaluation, while trust and cooperation contribute to more stable customer–operator relationships [46,47].

From a system perspective, these dimensions are not independent. Operational performance, financial outcomes, and customer satisfaction interact and jointly shape the functioning of postal and logistics services. In this context, customer satisfaction can be interpreted not only as an outcome of service quality, but also as a factor influencing demand patterns and service utilization. Accordingly, this paper considers customer satisfaction as a measurable component that enables the joint analysis of user preferences and operator performance in the process of operator selection.

In recent research, postal and logistics performance has also been examined in relation to environmental outcomes. Empirical evidence based on the Logistics Performance Index (LPI) indicates that improvements in different operational dimensions may have varying environmental effects [5]. For instance, while certain aspects of logistics organization are

associated with lower emissions, increased emphasis on delivery speed and strict scheduling may lead to higher transport intensity and energy use. These findings suggest that improvements in logistics performance do not necessarily correspond to reduced environmental impact. Instead, their effects depend on how service operations are structured and implemented. Additional studies further confirm the interdependence between logistics efficiency, energy consumption, and environmental pressure at both regional and international levels [48,49].

Within urban contexts, such relationships become more pronounced, as logistics activities are embedded in existing traffic systems and spatial constraints. Consequently, performance indicators should be interpreted not only as internal management tools, but also as variables reflecting broader patterns of urban flows and environmental exposure [50,51].

2.3. Machine Learning in City Logistics Systems

Machine learning (ML) represents a core component of artificial intelligence, enabling the extraction of patterns from data and the generation of predictive insights without explicit programming [52,53]. Its application typically involves data preparation, model training, validation, and implementation within operational environments [2,54].

In postal and logistics systems, ML has been primarily applied to operator-oriented tasks, including customer churn prediction [55], price and demand estimation [56], freight classification and product clustering [57], and competitiveness analysis [58]. These approaches contribute to improved operational efficiency and competitiveness, while customer-related aspects are often addressed indirectly through service performance metrics such as time, cost, and reliability. Some studies extend this perspective by incorporating customer satisfaction through hybrid or perception-based approaches, including survey-based evaluation frameworks and analysis of user-generated data [59,60]. Nevertheless, customer satisfaction is rarely treated as an explicit predictive variable within decision-support models. While these applications demonstrate strong predictive performance, they typically rely on single-operator or proprietary datasets and rarely report interpretability or robustness checks (e.g., cross-validation, predictor-importance analysis), which limits their transferability and managerial applicability beyond the specific operational context studied.

Traditional evaluation approaches, such as SERVQUAL and Kansei-based methods, provide structured insights into service quality and customer perception [61,62]. However, these approaches are generally descriptive and limited in their ability to support predictive and data-driven decision-making.

Several studies attempt to integrate operator and customer perspectives within broader logistics frameworks. These include applications focused on delivery time prediction, network design, and urban distribution planning, often combining machine learning with optimization and spatial analysis techniques [63–65]. More recent approaches incorporate sustainability considerations through the use of AI-based optimization and alternative delivery concepts, demonstrating improvements in operational performance alongside reductions in energy use and emissions [66,67].

At the same time, the role of ML in logistics has expanded beyond efficiency-oriented applications towards the analysis of environmental and urban system dynamics. Recent studies apply ML techniques to emission prediction, energy optimization, route planning, and the analysis of freight movement patterns in urban environments [2,68–71]. In parallel, methods such as sentiment analysis and natural language processing are used to examine user perceptions of logistics systems and their impacts within cities [60]. However, many of these models rely on complex, low-interpretability architectures (e.g., deep learning, LSTM) without accompanying explainability analysis, and they are rarely linked back to customer-level behavioral drivers, leaving a disconnect between environmentally oriented modelling and the demand-side factors that ultimately shape delivery activity.

From a methodological perspective, a wide range of techniques has been applied, including classification, regression, ensemble, and deep learning approaches. These methods differ in terms of interpretability, flexibility, and predictive performance, and are often selected depending on the complexity of the problem and the structure of available data. To provide a clearer comparative synthesis of the existing literature and to distinguish the present study from previous machine learning and data-driven applications in postal and urban logistics, Table 1 summarizes representative studies according to their data type and source, applied methods, and main contribution.

Table 1. Comparative overview of selected machine learning and data-driven studies in postal and urban logistics.

Ref.	Data Type and Source	Applied Method(s)/ Algorithm(s)	Main Contribution
[2]	Published studies	Literature review of ML and ANN applications	Reviews ML and ANN applications for reducing urban greenhouse gas emissions.
[55]	E-commerce transactions, reviews, geolocation, and census data	Logistic Regression, XGBoost, topic modelling, DBSCAN, PCA, and XAI	Predicts e-commerce customer churn.
[56]	Purchase history, transactions, and customer feedback	Multiclass Decision Forest, Logistic Regression, and One-vs-All	Predicts product prices, demand, discounts, and continuity from customer behavior.
[57]	Electronic waybill and order data from transport company	IBk, SMO, and LMT	Predicts freight type from electronic waybill data.
[58]	Questionnaire and competitiveness indicators for 12 enterprises	SEM, Nested Logit, SEM-NL, factor analysis, and PCA	Evaluates logistics competitiveness and logistics-mode choice.
[59]	Delivery-service and customer satisfaction data	Blockchain, LSTM, and smart contracts	Predicts delivery satisfaction and supports automated compensation decisions.
[60]	111,265 city logistics-related tweets	NLP, SVD, K-Means, t-SNE, and VADER	Identifies city logistics topics and public sentiment.
[61]	Questionnaire data from 400 postal users	SERVQUAL and non-parametric statistical tests	Measures gaps between expected and perceived postal service quality.
[62]	Questionnaire data from logistics users	Modified Kansei, SERVQUAL, EFA, and CFA	Examines service quality, satisfaction, commitment, and loyalty.
[63]	Retail delivery, customer, traffic, and location data	Clustering, optimization, simulation, and deep reinforcement learning	Develops a data-driven framework for improving urban last-mile routing, vehicle use, and decision making.
[64]	Large-scale Canada Post parcel, origin–destination, and weather data	Convolutional neural networks and baseline ML models	Develops an end-to-end deep-learning model for predicting last-mile parcel delivery time.
[65]	More than 30,000 pickup and delivery records from a courier operator	K-means++, Elbow and Silhouette methods, and GIS	Determines the optimal number and locations of postal centers.

Table 1. *Cont.*

Ref.	Data Type and Source	Applied Method(s)/ Algorithm(s)	Main Contribution
[66]	Pre- and post-implementation logistics data from a logistics operator	AI route optimization, fleet coordination, energy management, and EV-based delivery	Demonstrates that AI–EV integration can improve delivery time, energy efficiency, fleet use, customer satisfaction, and emissions.
[67]	Expert assessments from 43 stakeholders in seven countries	Fuzzy Delphi, Fuzzy FARE, and Fuzzy COBRA	Ranks barriers to drone use in last-mile logistics and strategies for overcoming them.
[68]	Inbound logistics planning data on inventory, sourcing, time, strategy, and network design	KNN, Random Forest, and SVM	Applies ML to support planning and prediction in inbound logistics processes.
[69]	Simulated urban freight-demand data	Hybrid ML framework for urban zoning	Develops a demand-based zoning approach for planning urban consolidation networks.
Present study	Survey data from users of three postal and logistics operators	Linear Regression, Random Forest, Gradient Boosting, and Friedman test	Develops a comparative ML framework for predicting customer satisfaction and supporting operator selection based on formalized customer profiles.

Table 1 confirms that machine learning and related data-driven approaches have been applied to a broad range of logistics problems, from customer churn, demand forecasting, and freight classification to delivery-time prediction, route planning, urban zoning, and postal-center location. The usefulness of these methods, however, depends strongly on the nature of the task and the structure of the available data. Models developed on large transactional and operational datasets can capture complex patterns that are difficult to identify using conventional statistical procedures, but they are often designed for one narrowly defined objective and provide limited insight into customer satisfaction as a distinct behavioural outcome. Deep-learning and ensemble methods offer considerable predictive flexibility, especially when nonlinear relationships and interactions are present, yet they generally require larger and better-balanced datasets and may be less transparent to logistics managers who need to understand the basis of a recommendation. Clustering, GIS, simulation, and optimization methods are particularly valuable for network design, routing, facility location, and the spatial organization of delivery systems, but customer satisfaction is usually treated only indirectly through time, cost, reliability, or service availability. Survey-based approaches such as SERVQUAL, SEM, EFA, CFA, and Kansei Engineering provide a clearer interpretation of service-quality dimensions and customer perceptions. Their main strength lies in explanation rather than prediction: they identify relationships and evaluation gaps, but they are not primarily designed to estimate satisfaction for a new customer profile or to compare several operators within the same decision procedure. Social-media and text-mining approaches extend the analysis to large volumes of user-generated content and can capture public perceptions in near real time, although the data are less controlled, may reflect self-selection, and are not always representative of the wider customer population. Taken together, the existing literature offers strong solutions for specific operational, spatial, and customer-related problems, but the integration of these perspectives remains limited. Among the studies reviewed, customer satisfaction is either examined as an explanatory outcome, inferred indirectly through service-performance

indicators, or incorporated into a narrowly defined predictive task. None of these studies develops a procedure in which satisfaction is predicted for the same formalized customer profile across several postal and logistics operators and then used as the basis for operator selection. In addition, although the studies presented in Table 1 apply a wide range of statistical, machine learning, and deep-learning methods across related logistics problems, none of the reviewed applications compares Linear Regression, Random Forest, and Gradient Boosting within the same modelling framework. This distinction is important because the proposed framework contributes at two connected levels. Methodologically, it compares three complementary regression approaches within a single operator-selection procedure. Practically, it serves both sides of the service relationship: customers receive a structured recommendation of the operator whose service characteristics are most closely aligned with their profile and expectations, while operators gain insight into the customer segments, service conditions, and preference patterns associated with higher or lower predicted satisfaction. The present study therefore brings together a distinct comparative modelling design, formalized customer profiling, satisfaction prediction, and operator recommendation within one integrated framework.

2.4. Customer Satisfaction and Environmental Implications in Urban Logistics

Customer satisfaction in logistics services has traditionally been analyzed from the perspective of service quality, marketing performance, and customer relationship management [72]. Numerous studies have emphasized that factors such as delivery reliability, timeliness, communication, and flexibility of delivery options significantly influence users' perceptions of logistics service quality and their overall satisfaction with service providers (e.g., [73,74]). In the context of parcel and postal services, customer satisfaction is therefore often considered a key determinant of customer loyalty and competitive positioning among logistics operators [75].

However, in urban logistics systems, customer satisfaction may also have broader operational implications that extend beyond traditional service quality considerations. In particular, mismatches between customer expectations and delivery service characteristics can increase the probability of unsuccessful delivery attempts [76]. When customers are not available at the time of delivery, or when delivery arrangements do not correspond to their preferences, logistics operators are often required to perform additional delivery rounds or redirect shipments to alternative delivery points [77]. Such operational disruptions represent an important source of inefficiency in last-mile logistics systems.

From an environmental perspective, failed deliveries and repeated delivery attempts contribute to increased transport activity in urban areas [78]. Additional delivery rounds increase vehicle kilometers travelled, fuel consumption, and associated emissions, thereby intensifying the environmental pressures generated by urban freight transport [79]. Previous research on city logistics has highlighted that improving delivery success rates is an important strategy for reducing the environmental footprint of last-mile distribution systems (e.g., [80,81]). Studies on urban freight transport emphasize that inefficient delivery processes can significantly increase congestion, energy consumption, and greenhouse gas emissions in densely populated urban environments [82]. The relationships between failed deliveries, additional transport activity, and environmental pressure summarized above are empirically documented in the cited studies. The present study draws on this established evidence to motivate its conceptual framework, rather than re-testing these relationships within its own dataset.

Consequently, improving the alignment between customer expectations and delivery service configurations represents a potentially important mechanism for enhancing both operational efficiency and environmental sustainability in urban logistics systems. A better

understanding of the determinants of customer satisfaction can therefore support logistics operators in designing delivery services that are better aligned with customer expectations and may reduce the likelihood of failed deliveries and unnecessary delivery rounds. Advances in data-driven analytics and machine learning methods provide new opportunities for analyzing complex relationships between customer characteristics, preferences, and service perceptions, enabling more accurate prediction of customer satisfaction and improved service design decisions [78].

Despite the growing body of research on both urban logistics sustainability and customer satisfaction in logistics services, relatively limited attention has been given to integrating these perspectives through predictive analytics approaches. In particular, there is a lack of studies that examine how machine learning-based prediction of customer satisfaction can be used as a decision-support tool for improving delivery efficiency and potentially reducing the environmental impacts associated with last-mile delivery activities in urban logistics systems. Addressing this research gap, the present study proposes a machine learning framework for predicting customer satisfaction with logistics operators and explores how such predictions may support decision-making for more efficient urban delivery operations and provide a behavioural basis for future sustainability-oriented analyses.

3. Materials and Methods: Comparative LR–RF–GB Modeling Framework for City Logistics Operator Selection

This section presents the methodological framework of the study, including the conceptual structure and a comparative machine learning approach developed to analyze customer satisfaction patterns and support operator selection in urban parcel delivery systems.

3.1. Conceptual Framework

The conceptual framework adopted in this study links several analytical layers that connect customer behavior with logistics and environmental outcomes in urban delivery systems. Figure 1 illustrates the conceptual framework linking customer characteristics, predicted satisfaction, decision and logistics outcomes, and environmental impacts in urban delivery systems.

The first layer refers to customer-related factors, including demographic, geodemographic characteristics, customer preferences, and responses collected through the survey. These variables represent the input features used in the predictive modelling process. In this study, customer preferences refer to the service-related requirements, priorities, and expected service conditions expressed by respondents when evaluating or selecting a postal and logistics operator. These preferences include the preferred basis for service pricing, the priority criterion used for operator selection, and the expected delivery time, together with broader service-related factors such as delivery speed, shipment security, price, ICT-supported services, operator availability, staff flexibility, information quality, network coverage, and service range. The second layer consists of the customer satisfaction prediction model, where machine learning techniques are applied to estimate the expected level of customer satisfaction based on the identified customer characteristics and preferences. In addition to the predictive component, the framework incorporates an intermediate layer that links model outputs with subsequent analytical and decision-making processes. Predicted satisfaction levels are interpreted through two complementary perspectives. From the user side, they support the selection of a logistics operator aligned with individual preferences, while from the operator side they indicate service dimensions that may require adjustment or improvement.

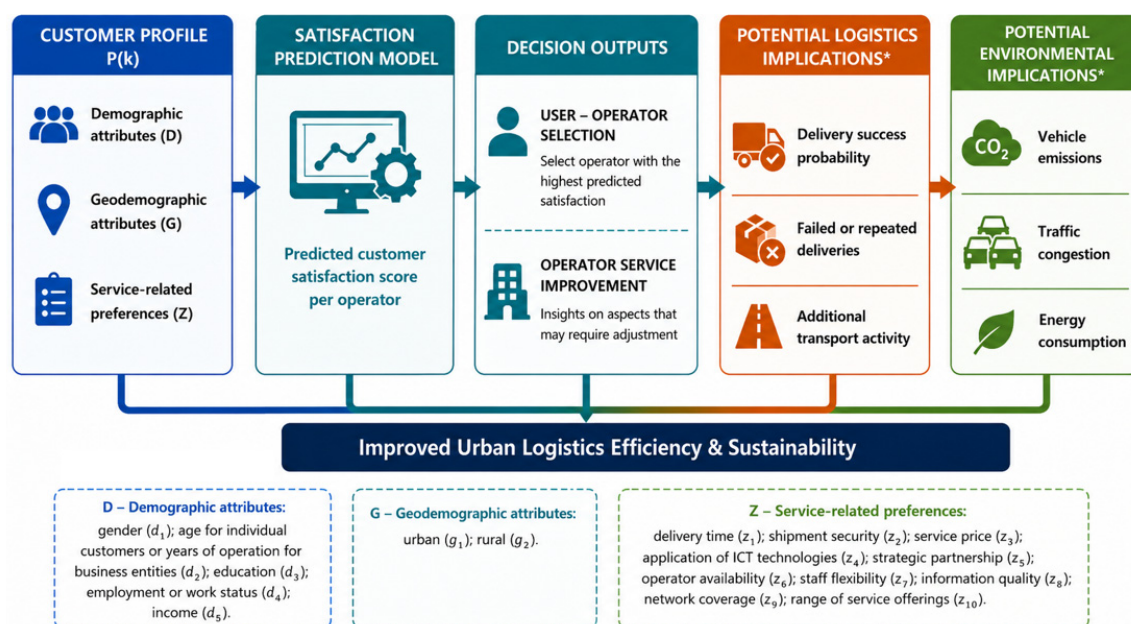


Figure 1. Conceptual framework. * The logistics and environmental implications are conceptually grounded in prior literature but were not directly estimated because the dataset did not include operational or environmental indicators.

The predicted satisfaction levels are then conceptually associated with logistics outcomes, including delivery success probability, the likelihood of failed deliveries, and the resulting transport activity generated by delivery operations. These logistics outcomes may further be linked to environmental impacts, such as vehicle emissions, traffic congestion, and energy consumption associated with urban freight transport. However, these operational and environmental indicators are not directly measured or modelled in the present study. By linking these layers, the proposed framework highlights how understanding and predicting customer satisfaction may support improved alignment between customer expectations and logistics service configurations, providing a basis for future integration with operational and environmental performance indicators. The links depicted between logistics outcomes and environmental impacts represent a conceptual structure informed by prior empirical literature. They are not statistical relationships estimated from the data collected in this study.

3.2. Proposed Comparative ML Framework

This study proposes a comparative machine learning framework based on the application and evaluation of three regression methods: Linear Regression (LR), Random Forest (RF), and Gradient Boosting (GB). The term “comparative framework” is used because the three algorithms are not combined into a single ensemble or meta-model; instead, they are trained and evaluated independently, and the best-performing method is selected for the operator selection stage. By structuring operator selection through data-driven prediction, the framework contributes to reducing mismatches between customer expectations and service characteristics, which may lead to repeated delivery attempts, prolonged vehicle presence, and increased transport activity in urban areas. Since the objective is to predict customer satisfaction ratings and compare predicted satisfaction levels across logistics operators, the problem is formulated as a regression task. Although customer satisfaction was measured on a five-point Likert scale, which is ordinal in nature, the satisfaction score was treated as a quasi-interval outcome for modelling purposes. This decision was made because the main objective of the framework is not to conduct statistical inference on an ordinal scale or to assign respondents to discrete satisfaction classes, but to estimate relative satisfaction levels.

and compare predicted satisfaction across postal and logistics operators. More specifically, the final phase of the proposed framework requires the ranking of several operators for the same customer profile and the selection of the operator with the highest expected satisfaction level. Regression modelling enables continuous satisfaction estimates, which allow more precise differentiation between operators that could otherwise be assigned to the same ordinal satisfaction category in a classification-based approach. For example, predicted values such as 4.72 and 4.39 may both correspond to a high satisfaction category, but they still provide useful information for ranking operators in the decision-support stage. Therefore, the regression approach was selected because continuous satisfaction estimates are functionally more suitable for operator ranking and recommendation, while still acknowledging the ordinal nature of the original Likert-scale measurement. Regression modelling also enables the use of standard prediction-error metrics, such as MAE, MSE, and R^2 , which support objective comparison of model performance.

Treating ordinal satisfaction scores as numerical values involves several methodological considerations. First, it assumes approximately equal distances between adjacent satisfaction categories, although the perceived difference between ratings 1 and 2 may not be equivalent to the difference between ratings 4 and 5. Second, since regression models were used, the predicted satisfaction values are expressed as continuous numerical estimates rather than discrete Likert-scale categories. Therefore, these values should not be interpreted as exact survey responses, but as estimated relative satisfaction levels used for comparison across operators. When categorical interpretation is required, continuous estimates can be mapped to the nearest Likert category; however, in this study they are primarily used to compare predicted satisfaction levels across operators.

The selection of the applied methods was based on the structure of the available survey data and the complementary characteristics of the algorithms. LR was included as an interpretable baseline model for examining linear relationships between customer attributes and satisfaction ratings. RF was selected because it can effectively capture complex nonlinear relationships and interactions among heterogeneous demographic, geodemographic, shipment-related, and preference-related variables, while also being relatively robust to noise and overfitting. GB was included because it sequentially improves prediction by focusing on the residual errors of previously constructed trees and is therefore suitable for identifying more subtle patterns in structured tabular data. The joint application of these methods enables a comparison between a transparent linear baseline and two tree-based ensemble approaches with different learning mechanisms, allowing the most reliable model to be selected empirically for the operator-selection stage. In this research, multiple linear regression [83] is applied to estimate the relationship between customer attributes and satisfaction ratings, expressed as:

$$Y = f\{x, \beta\} \quad (1)$$

where Y denotes the dependent variable representing customer satisfaction with the service, f function dependent on input variables x and parameters β , x refers to the independent variables (demographic, geodemographic, and preference related attributes), and β represents the unknown model parameters. While LR enables interpretation of relationships between variables, its performance may be limited in the presence of nonlinear patterns. To address this, the model incorporates RF, an ensemble method based on aggregating multiple decision trees [84]. The prediction is obtained as the average output of individual trees:

$$g(x) = \frac{1}{n} \sum_{i=1}^n f_i(x) \quad (2)$$

where $f_i(x)$ represents the prediction generated by the i -th decision tree for input sample x , while n denotes the total number of trees in the ensemble. Further improvement is achieved through the GB algorithm [85], which constructs decision trees sequentially, with each iteration reducing the residual error of the previous model:

$$F_N(x_t) = \sum_{n=1}^N f_n(x_t) \quad (3)$$

where $f_n(x_t)$ represents the prediction of the N -th weak learner (decision tree), while $F_N(x_t)$ denotes the cumulative output of the model after N iterations. The ensemble of trees is constructed iteratively, where each new decision tree $f_{n+1}(x_t)$ is evaluated through the minimization of the loss function, as shown in the following equation:

$$\operatorname{argmin} \sum_t L(y_t F_N(x_t) + f_{n+1}(x_t)) \quad (4)$$

In the above equations, y_t denotes the actual value of the target variable for the current sample, $F_N(x_t)$ represents the cumulative prediction of the ensemble after N iterations, and $f_{n+1}(x_t)$ is the new weak learner added to minimize the residual error. The function $L(y_t F_N(x_t))$ represents the loss function that quantifies the difference between the true and predicted values, while the iterative minimization process ensures that each subsequent model focuses on reducing the remaining prediction errors. The overall comparative LR–RF–GB framework developed in this study is shown in Figure 2 using a BPMN (Business Process Model and Notation) diagram. The process comprises five main phases:

Phase I: Data collection from postal and logistics service. It involves the collection of data using a structured questionnaire developed on the basis of a formalized customer profile created in this study. The profile systematizes demographic, geodemographic, and service-related factors identified in the literature into an integrated analytical structure suitable for predictive modelling. Accordingly, the questionnaire captures customer characteristics, service requirements, priorities, expected service conditions, shipment-related information, and satisfaction ratings, thereby providing the empirical basis for the subsequent modelling phases.

Phase II: Data preparation and selection of relevant parameters for modeling. It includes data cleaning, correction of inconsistencies, consolidation and regrouping of insufficiently represented categories, selection of relevant variables, and preparation of separate datasets for the selected operators. Categorical variables are transformed using one-hot encoding, the target variable is adjusted through winsorization, and the prepared datasets are divided into training and test subsets for model development and evaluation.

Phase III: Prediction of customer satisfaction ratings per operator using regression methods. It involves training and applying the LR, RF, and GB regression models to predict customer satisfaction separately for the three selected postal and logistics operators. The same modelling procedure is also applied to the extended resampled dataset in order to examine model behaviour under more balanced target-variable conditions.

Phase IV: Evaluation and verification of model performance and selection of the final method for application in the operator selection phase. It includes the evaluation and verification of model performance using Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R^2). In addition, the absolute prediction errors of the LR, RF, and GB models are statistically compared within each dataset using the Friedman test. For the best-performing model, a cross-validated permutation importance analysis is also conducted to identify the customer and service-related predictors that contribute most consistently to customer satisfaction predictions across different validation folds. The combined evaluation, statistical comparison, and predictor-importance results

are then used to identify, interpret, and select the most reliable method for application in the final operator-selection phase.

Phase V: Selection of the operator for customer k based on the highest predicted rating. It applies the best-performing model to identical customer profiles across the selected operators. For each customer profile, the model generates a predicted satisfaction score for every operator, and the operator with the highest predicted value is selected as the most suitable option. This phase transforms the predictive results into a practical decision-support recommendation for operator selection.

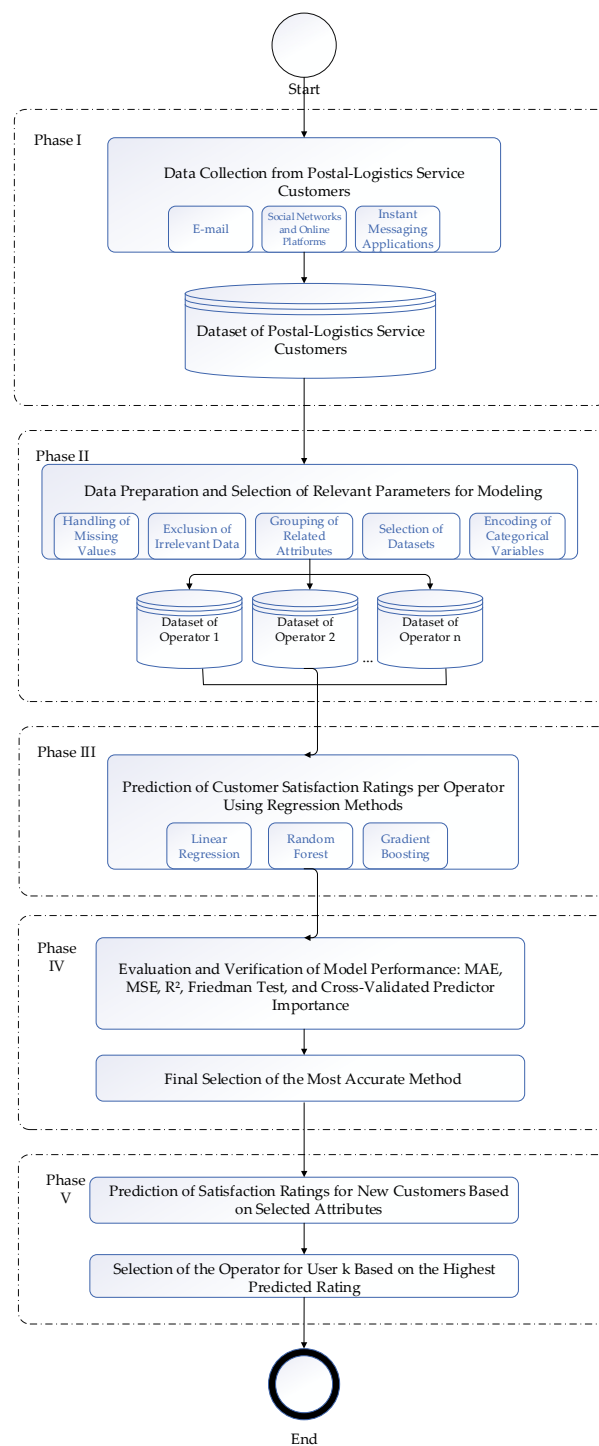


Figure 2. Structure of proposed comparative LR–RF–GB machine learning framework.

Beyond individual operator recommendation, the outputs of this phase can also support operator-specific service adjustment. Customer profiles associated with lower predicted satisfaction can be identified and used to guide the adaptation of communication channels, delivery options, service packages, and other service conditions. When combined with operational and spatial logistics data, the same procedure may also provide an analytical basis for broader demand-responsive urban freight planning.

3.2.1. Phase I: Data Collection from Customers of Postal and Logistics Services

Data collection was conducted using a structured questionnaire as the primary research instrument. Due to confidentiality constraints, direct access to operator data was not available, which motivated the development of a survey framework capturing demographic and geodemographic characteristics, together with customers' service-related requirements, priorities, and expected service conditions, hereafter referred to as customer preferences. The questionnaire design was informed by the demographic, geodemographic, and service-related factors identified in the literature review presented in Section 2.1, ensuring consistency between the theoretical background and the empirical analysis. Since the existing literature identifies these factors mainly as separate determinants, but does not integrate them into a single formalized customer profile suitable for predictive modelling, this study combines them into a structured analytical representation. The customer profile is formally represented as:

$$P_{(k)} = \{D, G, Z\} \quad (5)$$

where:

$$D = \{d_1, d_2, d_3, d_4, d_5\} \wedge G = \{g_1, g_2\} \wedge Z = \{z_1, z_2, z_3, \dots, z_{10}\} \quad (6)$$

The set D includes the demographic characteristics of the customer, such as gender (d_1), age for individual customers or years of operation for business entities (d_2), education (d_3), employment or work status (d_4), and income (d_5). The set G encompasses spatial components referring to urban (g_1) or rural areas (g_2). The set Z includes customer preferences, defined as service-related requirements, priorities, and expected conditions considered when evaluating or selecting a postal and logistics operator, among which are delivery time (z_1), shipment security (z_2), service price (z_3), application of ICT (z_4), strategic partnership (z_5), operator availability (z_6), staff flexibility (z_7), information quality (z_8), network coverage (z_9), and the range of service offerings (z_{10}). In this way, the defined profile $P_{(k)}$ enables a systematic representation of the customer through the most relevant demographic, geodemographic, and preference-related factors, which form the basis for customer satisfaction modeling in subsequent stages of the research.

The questionnaire was distributed via Google Forms through multiple communication channels, including email and social media. Respondents selected among 18 postal and logistics operators. The survey was conducted between February and October 2023, resulting in 1239 valid responses. The collected data reflect service usage patterns in dense urban environments, where delivery activities interact with existing traffic conditions and infrastructure constraints.

The dataset includes 14 attributes describing customer characteristics, shipment features, service preferences, and satisfaction ratings. The detailed categorization of variables is presented in the subsequent phase of data preparation. Due to confidentiality restrictions, the dataset is not publicly available; however, anonymized data and analysis scripts can be provided upon request.

3.2.2. Phase II: Data Preparation for Modeling and Selection of Relevant Parameters

Data preparation was conducted to ensure consistency and enable reliable model generalization [86,87]. The process included data cleaning and correction of minor inconsistencies. Attributes and subcategories with insufficient representation were removed [88]. In cases where sparse categories retained analytical relevance, responses were reassigned to related groups. To reduce data fragmentation, conceptually similar categories were consolidated. The variables Age and Years of operation were grouped into six intervals (Table 2). Product categories were aggregated into broader groups to enable more stable pattern recognition and reduce variability in the dataset.

Table 2. Categorization of the attributes “Age” for individuals and “Years of operation” for legal entities.

Category	Individual (C2 and 2C)	Legal Entity (B2 and 2B; G2 and 2G)
I	18–25	0–5
II	26–30	6–10
III	31–40	11–25
IV	41–50	26–35
V	51–60	36–50
VI	Over 60	Over 50

An imbalance in the distribution of responses among operators was observed, with a predominance of courier services. This distribution was treated as an empirical characteristic of the urban logistics market rather than as a statistical irregularity. Model development was therefore focused on the three operators with the largest number of observations. The final dataset included ten customer-related attributes used as input variables, while the operator identifier was excluded to ensure objective evaluation. The structure, categorization, and types of variables are summarized in Table 3.

Table 3. Categorization and types of variables for analyzing customer satisfaction with provided services.

Variable Name	Variable Categorization	Variable Values	Variable Type
Gender	Independent	M (Male) F (Female) C (Company)	Categorical
Customer category	Independent	Individual customer (C2 and 2C) Legal entity—company/enterprise (B2 and 2B) Legal entity—institution (G2 and 2G)	Categorical
Age/Years of operation	Independent	Category I, II, III, IV, V, VI	Categorical
Education	Independent	Secondary school Short-cycle tertiary Bachelor’s degree Master’s degree PhD	Categorical
Type of shipment	Independent	Letter/document Parcel Pallet	Categorical

Table 3. Cont.

Variable Name	Variable Categorization	Variable Values	Variable Type
Product category	Independent	Document Clothing Footwear Cosmetics Electronics Auto parts Consumer goods and food Entertainment & children	Categorical
Billing preference	Independent	Item Weight Distance (kilometers traveled)	Categorical
Priority criterion for operator selection	Independent	Delivery time Delivery price Service availability Payment method and options Professional and courteous staff	Categorical
Expected service delivery time	Independent	D (same-day delivery) D + 1 (within two days) D + 2 (within three days) D + 3 (within four days) D + 4 (within five days) Other	Categorical
Customer satisfaction	Dependent	1 (very dissatisfied) 2 (dissatisfied) 3 (neutral) 4 (satisfied) 5 (very satisfied)	Ordinal treated as quasi-interval for regression modelling

Categorical variables were transformed using One-Hot encoding, while the dependent variable (customer satisfaction) was adjusted using winsorization to limit the influence of extreme values [89]. This approach preserves variability while improving model stability. The prepared datasets were subsequently used for model training and evaluation using LR, RF, and GB algorithms, applying standard regression metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Determination (R^2). The implementation was performed in a Python 3.10 environment using Google Colab and scikit-learn, with Pipeline and GridSearchCV for model optimization and five-fold cross-validation. Reproducibility was ensured by fixing the random seed (random_state = 42), while detailed parameter settings are provided in Appendix A.

4. Results and Interpretation

This section presents the empirical results of the predictive modeling conducted within the third phase of the methodological framework. The analysis covers three postal/logistics operators (Operator 1, Operator 2, and Operator 3), whose datasets were used to evaluate the performance of three machine learning regression methods: Linear Regression (LR), Random Forest (RF), and Gradient Boosting (GB). Although the predictive models developed in this study focus on estimating customer satisfaction levels, the interpretation of the obtained results can also be extended to a broader urban logistics context. As discussed in the conceptual framework, customer satisfaction can be interpreted as an indicator of the alignment between customer expectations and delivery service characteristics offered by

logistics operators. Higher levels of alignment may support delivery service configurations that are more consistent with customer expectations, while mismatches between expectations and service conditions may result in delivery disruptions and repeated delivery attempts. From a city logistics perspective, such repeated delivery attempts are recognized as an important source of additional transport activity and environmental pressure in urban freight systems; however, these operational and environmental indicators were not directly measured in this study.

In line with the methodological approach defined in Section 3, all datasets were divided into a training set (80%) and a testing set (20%) to ensure consistency in model evaluation. The dataset for Operator 1 consists of 381/96 samples, for Operator 2 a total of 252/64, while the dataset for Operator 3 contains 240/60 samples.

Based on the obtained results, a noticeable imbalance of the target variable was observed, particularly in the lower satisfaction ratings (1 and 2), which could have affected model performance and generalization capability. Therefore, as a complement to the initial analyses, an extended dataset was generated using a resampling technique, with the aim of achieving a more balanced distribution of satisfaction ratings and increasing the reliability of the models. The extended dataset contains a total of 10,000 samples, of which 80% were used for training and 20% for model testing.

4.1. Phase III: Prediction of Customer Satisfaction Ratings for Operator 1

Table 4 provides an overview of sample results for Operator 1. The first column presents the customer identification number (denoted as C1, C2, ..., C381), while the second column contains the actual satisfaction ratings. The remaining three columns show the predicted values generated using the LR, RF, and GB methods. This overview enables a direct comparison between actual and predicted values, forming the basis for analyzing deviations and the accuracy of each method.

Table 4. Results of applied LR, RF, and GB methods in predicting customer satisfaction with Operator 1's service.

Customer ID	Actual Customer Satisfaction Ratings	LR Predicted Ratings	RF Predicted Ratings	GB Predicted Ratings
C1	4	4.11	3.92	4.12
C2	4	4.04	4.61	4.11
C3	5	4.39	4.62	4.11
C4	4	4.19	4.12	4.09
C5	3	3.69	3.24	4.11
...	...	...	...	...
...	...	...	...	...
C380	5	4.86	4.78	4.40
C381	5	4.62	4.54	4.12

Figure 3 provides a visual representation of the model results through four graphs. Parts (a–c) show the curves comparing predicted and actual values for the LR, RF, and GB models, while part (d) illustrates the overall prediction accuracy by individual satisfaction rating. This visualization offers both numerical and graphical assessments of model performance, providing a comprehensive insight into their predictive capability.

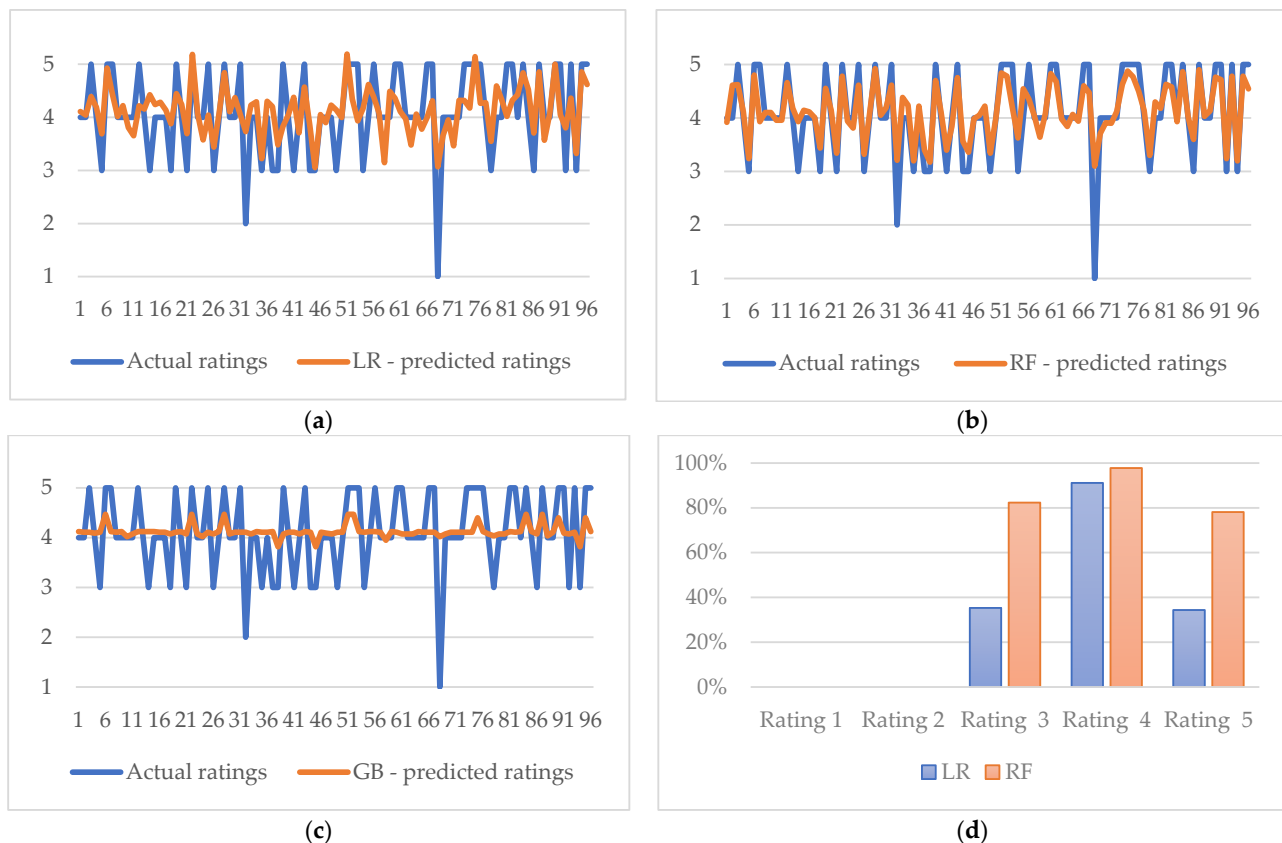


Figure 3. Prediction results for Operator 1: (a) Comparison of actual and predicted customer satisfaction ratings using LR method; (b) Comparison of actual and predicted customer satisfaction ratings using RF method; (c) Comparison of actual and predicted customer satisfaction ratings using GB method; (d) Prediction accuracy per satisfaction rating across models.

The analysis of the test dataset comprising a total of 96 samples reveals clear differences in the performance of the applied methods. The LR model (Figure 3a) demonstrates limited ability in accurately predicting customer satisfaction. Statistical data show that for rating 5, out of 32 test samples, 11 were correctly predicted, yielding an accuracy of 34.38%. For rating 4, the model performed significantly better—out of 45 samples, only 4 were misclassified, corresponding to an accuracy of 91.11%. In contrast, for rating 3, out of 17 test samples, 6 were correctly predicted, giving an accuracy of 35.30%. Low ratings proved particularly challenging: for ratings 1 and 2, each with only 4 samples, the model did not achieve a single correct prediction. Figure 3b presents the results of the RF regression method. It is evident that the model based on this method demonstrates the best results for predicting medium and high levels of customer satisfaction with Operator 1's services. This conclusion is further supported by statistical data. For rating 5, out of 32 test samples, 25 were correctly predicted, resulting in an accuracy of 78.13%. For rating 4, the model achieved nearly perfect results—out of 45 samples, only one was misclassified, corresponding to 97.78% accuracy. Similarly, for rating 3, out of 17 samples, 14 were correctly predicted, i.e., 82.35%. These values confirm a high degree of alignment between actual and predicted ratings, which is visually evident through almost parallel curves in Figure 3b. From an operational perspective, such predictive accuracy is particularly relevant because it enables a more reliable identification of customer segments that are likely to be satisfied with a specific logistics operator. In practical urban delivery systems, this information may support a better alignment between customer expectations and service configurations. Such alignment may contribute to reducing delivery mismatches

and repeated delivery rounds, which are recognized as a significant source of additional vehicle movements and operational inefficiencies in urban freight transport systems.

However, as with the LR method, the model could not accurately predict ratings 1 and 2 due to the very small number of samples in the test set. As shown in Figure 3c, the GB model was unable to predict accurate customer satisfaction ratings. Compared with the two previously discussed models, it demonstrated very poor performance both in predicting high ratings and in estimating medium and low satisfaction levels. Therefore, the GB model showed the weakest predictive performance for this dataset, although it was retained in the subsequent comparative evaluation in order to provide a complete statistical comparison of all three methods. From a decision-support perspective, the identification of more reliable predictive models is essential because inaccurate satisfaction predictions could lead to inappropriate operator recommendations. Such mismatches between customers and logistics operators may increase the likelihood of delivery disruptions, including missed deliveries or the need for repeated delivery attempts.

4.2. Phase III: Prediction of Customer Satisfaction Ratings for Operator 2

Following the same approach used for Operator 1, this section presents the results of applying the LR, RF, and GB methods to the dataset of Operator 2, both in tabular form (Table 5) and through graphical representations (Figure 4a–d).

Table 5. Results of applied LR, RF, and GB methods in predicting customer satisfaction with Operator 2'nd service.

Customer ID	Actual Customer Satisfaction Ratings	LR Predicted Ratings	RF Predicted Ratings	GB Predicted Ratings
C1	3	4.12	3.64	4.30
C2	4	4.08	4.16	4.29
C3	5	4.93	4.82	4.31
C4	5	4.97	4.74	4.44
C5	4	4.62	4.24	4.31
...	...	...	...	...
...	...	...	...	...
C63	3	4.11	3.56	4.30
C64	4	4.13	4.2	4.33

The analysis of individual examples from the Operator 2 test dataset (a total of 64 test samples) indicates satisfactory precision of the LR model in predicting high customer satisfaction ratings (Figure 4a). Among 27 samples rated 5, the model misclassified 9. Furthermore, out of 32 samples rated 4, it produced inaccurate predictions for 3 samples. Of the four test samples rated 3, all were incorrectly predicted, as was the single sample rated 2. No samples were recorded for rating 1. The application of this method to the Operator 2 dataset resulted in 0% prediction accuracy for ratings 1, 2, and 3, while accuracy levels of 90.63% and 66.67% were achieved for ratings 4 and 5, respectively (Figure 4d). The RF model stands out as exceptionally accurate in predicting ratings 4 and 5 (Figure 4b). This is supported by the descriptive accuracy results: out of 27 test samples rated 5, the RF model misclassified only 4, achieving an accuracy of 85.19%. For rating 4, it provided completely accurate predictions for all 32 test samples, attaining a perfect 100% prediction accuracy (Figure 4d). However, as in the case of Operator 1, due to the same challenge of a limited number of samples for ratings 3, 2, and 1, reduced precision in prediction was observed. As with Operator 1, the GB model for Operator 2 also failed to achieve satisfactory accuracy. In practical delivery systems, insufficient predictive accuracy may lead to suboptimal service allocation and customer–operator mismatches. These mismatches may translate into

operational inefficiencies such as rescheduled deliveries, parcel redirection, or additional delivery attempts. The curves of the predicted values deviate considerably from the actual ones (Figure 4c), confirming that the GB model showed the weakest predictive performance for this dataset (Figure 4d). Therefore, it was not considered a suitable candidate for the final operator selection stage, although it was retained in the subsequent comparative evaluation in order to provide a complete statistical comparison of all three methods. In contrast, the ability of the RF model to accurately identify high satisfaction levels is particularly important from the perspective of urban logistics system efficiency. When logistics operators are able to anticipate customer satisfaction with a high level of confidence, delivery services can be designed or recommended in a way that better reflects customer expectations regarding delivery time, price, and service conditions. Such alignment may reduce the probability of delivery mismatches and may contribute to more stable and predictable delivery operations within urban freight systems.



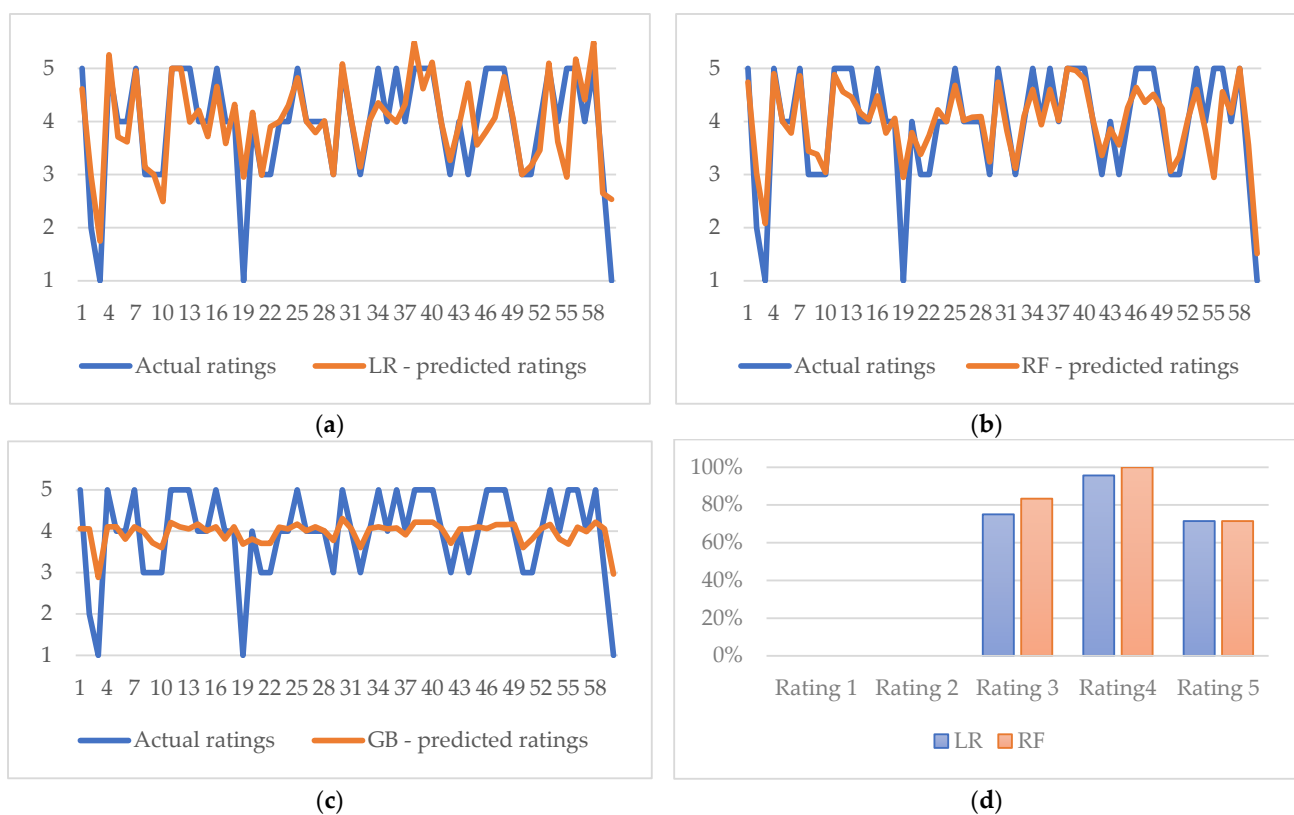
Figure 4. Prediction results for Operator 2: (a) Comparison of actual and predicted customer satisfaction ratings using LR method; (b) Comparison of actual and predicted customer satisfaction ratings using RF method; (c) Comparison of actual and predicted customer satisfaction ratings using GB method; (d) Prediction accuracy per satisfaction rating across models.

4.3. Phase III: Prediction of Customer Satisfaction Ratings for Operator 3

Table 6 provides an overview of the actual customer satisfaction ratings and the predicted values obtained using the LR, RF, and GB methods, while the results are also graphically presented in Figure 5a–d.

Table 6. Results of applied LR, RF, and GB methods in predicting customer satisfaction with Operator 3's service.

Customer ID	Actual Customer Satisfaction Ratings	LR Predicted Ratings	RF Predicted Ratings	GB Predicted Ratings
C1	5	4.61	4.74	4.06
C2	2	2.97	3.02	4.05
C3	1	1.75	2.08	2.88
C4	5	5.25	4.9	4.11
C5	4	3.71	4	4.11
...	...	...	...	...
...	...	...	...	...
C59	3	2.65	3.56	4.05
C60	1	2.53	1.51	2.96

**Figure 5.** Prediction results for Operator 3: (a) Comparison of actual and predicted customer satisfaction ratings using LR method; (b) Comparison of actual and predicted customer satisfaction ratings using RF method; (c) Comparison of actual and predicted customer satisfaction ratings using GB method; (d) Prediction accuracy per satisfaction rating across models.

The analysis shows that the LR method achieved stable results for medium and high satisfaction ratings (3, 4, and 5), while the RF method once again confirmed the highest precision—particularly for ratings 4 and 5, where prediction accuracy was very high. Conversely, as in the previous cases, the GB method exhibited significant deviations and insufficient predictive accuracy. However, it was retained in the subsequent comparative and statistical analyses to enable a complete evaluation of all three methods.

4.4. Phase III: Prediction of Customer Satisfaction Ratings on the Extended Dataset

The analysis of modeling results on the datasets of Operators 1, 2, and 3 revealed a significant imbalance in the target variable distribution, with lower satisfaction ratings

(1 and 2) being considerably underrepresented. In the context of machine learning, such an imbalance can lead to model bias and reduced generalization capability, as majority samples dominate the informational contribution of minority classes [90,91]. The problem of imbalanced datasets, where one of the possible classes has a very small frequency compared to others—is commonly addressed through sampling techniques such as undersampling or upsampling [92]. In regression tasks, the goal of such methods is to balance the distribution of the target variable and thereby improve model accuracy.

In this research, to resolve the imbalance and enhance model performance, a resampling procedure was applied, resulting in the creation of an extended dataset. It should be emphasized that the extended dataset was not intended to replace the empirical survey dataset or to serve as an independent empirical basis for drawing new conclusions. The primary model development and evaluation were conducted on the original survey-based datasets for Operators 1, 2, and 3. The extended dataset was used as an additional modelling step to examine model behavior under a more balanced distribution of customer satisfaction ratings, particularly because lower ratings were underrepresented in the empirical data.

A larger dataset generally contributes to improved reliability and generalization capacity of machine learning models [93]. Synthetic data [94] were generated using the dataset of Operator 1, which had the highest number of original samples (477). A resampling technique was applied to increase the number of instances with ratings 1 and 2 and to adjust the distribution of other ratings (3, 4, and 5) according to predefined target ratios. The resampling procedure was performed at the level of complete observations rather than by independently generating individual attributes. Therefore, each resampled instance retained the original combination of customer characteristics, shipment-related variables, service preferences, and satisfaction rating. In this way, the joint structure of the attributes was preserved, and unrealistic combinations of variables were avoided. Since the distribution of the target variable was intentionally adjusted to reduce the underrepresentation of lower satisfaction ratings, the extended dataset represents a resampled and balanced dataset used for model testing under more balanced target-variable conditions. Thus, the dataset preserves the observed attribute combinations from the original data, while the marginal distribution of satisfaction ratings was modified for modelling purposes.

The extended dataset contains 10,000 samples, of which 80% were used for training and 20% for testing the models. The LR, RF, and GB models were retrained and re-evaluated on this dataset. The results are presented in tabular form (Table 7) and graphically (Figure 6a–d).

Table 7. Results of applied LR, RF, and GB methods in predicting customer satisfaction ratings on the extended resampled dataset.

Customer ID	Actual Customer Satisfaction Ratings	LR Predicted Ratings	RF Predicted Ratings	GB Predicted Ratings
C1	4	3.16	3.97	3.92
C2	4	3.90	3.98	3.87
C3	4	4.79	4.13	4.21
C4	5	6.47	5	4.99
C5	3	2.72	3.20	3.19
...	...	...	...	...
...	...	...	...	...
C1999	2	3.57	2.74	2.87
C2000	2	3.57	2.74	2.87

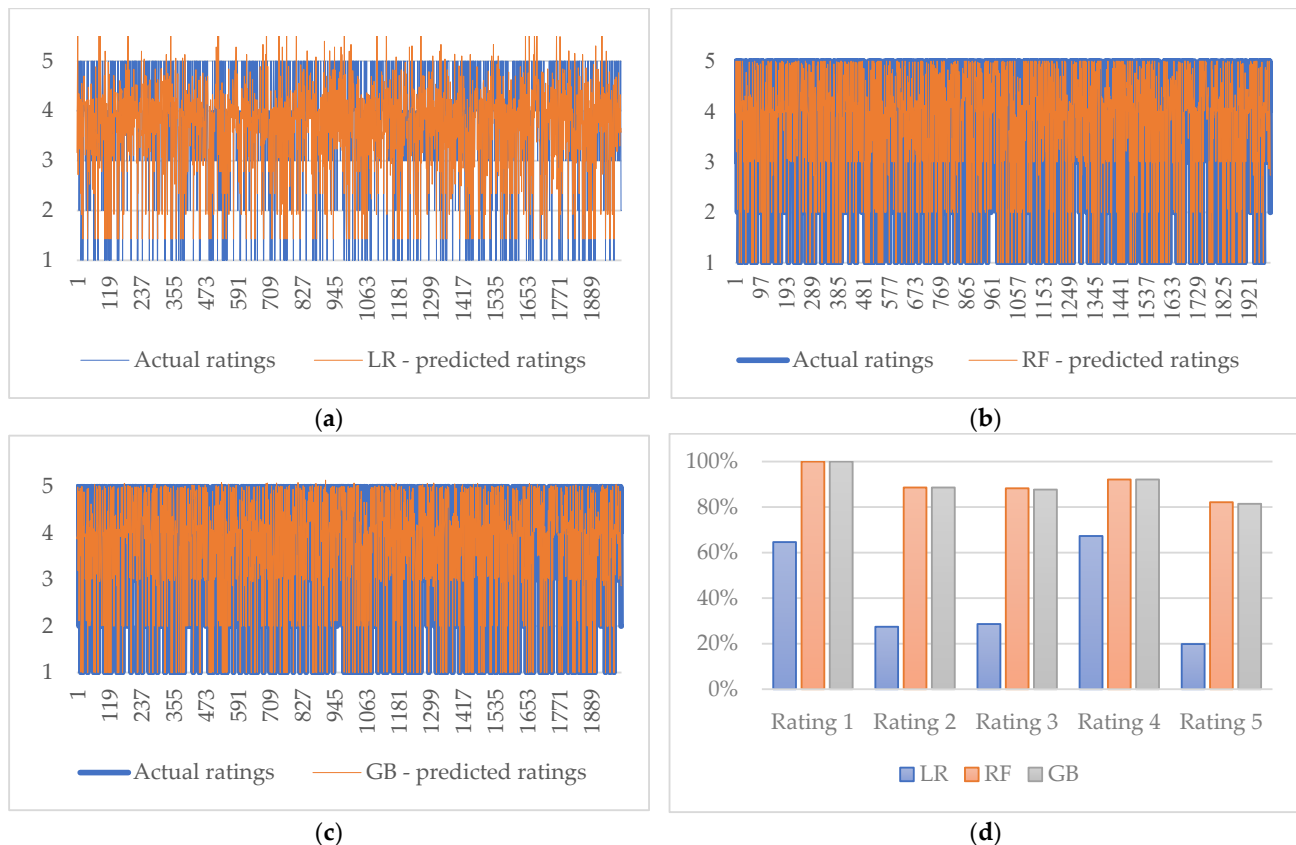


Figure 6. Prediction results on the extended resampled dataset: (a) Comparison of actual and predicted customer satisfaction ratings using LR method; (b) Comparison of actual and predicted customer satisfaction ratings using RF method; (c) Comparison of actual and predicted customer satisfaction ratings using GB method; (d) Prediction accuracy per satisfaction rating across models.

The modeling analysis on the extended dataset shows a significant improvement compared to the results obtained on the original data. By expanding the distribution of the target variable and balancing the lower satisfaction ratings (1 and 2), more precise learning and generalization were achieved. The application of the LR method still demonstrated limited ability to predict lower and midrange satisfaction ratings (Figure 6a). The model achieved 64.62% accuracy for rating 1, 27.43% for rating 2, and 28.61% for rating 3, while higher accuracy was achieved for ratings 4 (67.31%) and 5 (19.79%) (Figure 6d). These values confirm a relatively low level of accuracy and the limited capacity of the LR model to adequately capture complex relationships among variables. In contrast, the RF method exhibited very high predictive precision, with actual and predicted values almost perfectly aligned (Figure 6b). The model achieved 100% accuracy for rating 1, 88.57% for rating 2, 88.25% for rating 3, 92.11% for rating 4, and 82.11% for rating 5 (Figure 6d), confirming its ability to perform robustly even under balanced conditions. Interestingly, the results of the GB model (Figure 6c) show a substantial improvement compared to previous phases of the study. Unlike its poor performance on the original datasets, on the extended dataset, the GB model achieved results almost identical to those of the RF method: 100% accuracy for rating 1, 88.57% for rating 2, 87.65% for rating 3, 92.11% for rating 4, and 81.42% for rating 5 (Figure 6d). From a broader urban logistics perspective, the ability of advanced machine learning methods such as RF and GB to capture complex relationships between customer characteristics and satisfaction outcomes is particularly relevant. Accurate prediction of satisfaction enables a more informed allocation of delivery services and logistics operators, which may contribute to reducing operational disruptions and unnecessary delivery repetitions in urban distribution systems. When comparing the results of the RF and GB

methods, it is evident that both achieved high precision in predicting customer satisfaction ratings. However, some performance differences can be noted. The GB model showed slightly lower precision for rating 3, correctly predicting 87.65% of samples, whereas the RF model achieved 88.25%. The RF model also had a slight advantage in predicting rating 5, with 82.11% accuracy compared to 81.42% achieved by the GB model. For the remaining ratings (1, 2, and 4), the RF and GB methods demonstrated identical accuracy levels.

In the context of urban delivery systems, such predictive consistency is important because it allows logistics operators to rely on data-driven decision-making when designing delivery strategies or recommending service options to customers. Over time, improved alignment between predicted customer satisfaction and the selected logistics operator may support better service allocation and reduce delivery mismatches, providing a basis for future analyses of routing efficiency, delivery disruptions, and transport intensity within urban areas.

4.5. Phase IV: Evaluation and Verification of Model Performance

After predicting customer satisfaction ratings for Operators 1, 2, and 3, as well as for the extended dataset, an evaluation and verification of model performance were conducted to identify the method achieving the best results and suitable for implementation in the final stage of the proposed model—the operator selection process based on customer profiles. For this purpose, standard performance metrics were used, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R^2) [95]. Their mathematical definitions are given by the following expressions:

$$MAE = \frac{1}{N} \sum_{n=1}^N (\hat{Y}_n - Y_n) \quad (7)$$

where MAE represents the mean absolute error, N the number of observations, $\hat{Y}_n$ the predicted values, and Y_n the actual values.

$$MSE = \frac{1}{N} \sum_{n=1}^N (\hat{Y}_n - Y_n)^2 \quad (8)$$

where MSE denotes the mean squared error, where squaring the differences gives greater weight to larger deviations.

$$R^2 = 1 - \frac{SSR}{SST} = \frac{\sum (Y_n - \hat{Y}_n)^2}{\sum (Y_n - \bar{Y})^2} \quad (9)$$

where R^2 represents the coefficient of determination, SSR the sum of squared residuals, and SST the total sum of squares of deviations of actual values from their mean $\bar{Y}$.

Lower MAE and MSE values indicate higher model accuracy and precision [96], while R^2 shows how well the model explains the variability of the dependent variable. An R^2 value close to 1 indicates a wellfitted model with a high level of explanatory power, while values closer to 0 indicate a weaker ability to explain variations in the data [97]. Given the nature of the problem addressed, the model is considered successful and not overfitted if the predicted satisfaction ratings are relatively close to the actual ones. Table 8 presents the evaluation results of the LR, RF, and GB models for Operators 1, 2, and 3, as well as for the extended dataset.

Table 8. Performance comparison of LR, RF, and GB models on the actual and extended datasets.

		Evaluation Metric			
	Applied Method	MAE	MSE	R ²	
Data Collected through Empirical Research	Operator 1	LR	0.469	0.347	0.359
		RF	0.264	0.113	0.791
		GB	0.524	0.464	0.143
	Operator 2	LR	0.356	0.208	0.489
		RF	0.264	0.116	0.714
		GB	0.512	0.343	0.155
	Operator 3	LR	0.407	0.395	0.629
		RF	0.317	0.246	0.769
		GB	0.641	0.718	0.326
Extended datasets	LR	0.686	0.713	0.544	
	RF	0.159	0.100	0.936	
	GB	0.168	0.101	0.935	

The results show that for Operator 1, the RF model proved to be the most accurate, with the lowest MAE (0.264) and MSE (0.113) values and the highest R² (0.791), indicating good explanatory capability. The LR model demonstrated moderate accuracy (MAE = 0.469, MSE = 0.347, R² = 0.359), while the GB method exhibited weaker performance with higher error values (MAE = 0.524, MSE = 0.464, R² = 0.143). A similar pattern was observed for Operators 2 and 3, where RF again achieved the best results (MAE = 0.264 and 0.317, MSE = 0.116 and 0.246, R² = 0.714 and 0.769). The LR model maintained moderate performance, while the GB method showed greater deviations between actual and predicted values. The analysis of the extended dataset further confirms these findings. The RF model achieved the highest accuracy (MAE = 0.159, MSE = 0.100, R² = 0.936), while the GB method achieved nearly identical results (MAE = 0.168, MSE = 0.101, R² = 0.935), showing a significant improvement compared to the individual operator datasets. Conversely, the LR model exhibited a decline in performance (MAE = 0.686, MSE = 0.713, R² = 0.544), indicating a limited ability to model more complex nonlinear relationships. Beyond their statistical interpretation, these evaluation metrics also provide an indication of the reliability of satisfaction predictions when applied in practical decision-support contexts. Reliable prediction models allow logistics operators to better anticipate customer responses to different service configurations, which may support improved delivery planning and reduce the likelihood of operational disruptions. In urban logistics systems, reducing such disruptions can contribute to lowering the number of repeated delivery attempts and associated transport activity, thereby supporting efforts aimed at mitigating the environmental impacts of last-mile delivery operations.

Overall, the results obtained from both the original and extended datasets indicate that the RF method achieves the highest level of stability and predictive precision. Dataset expansion further improved model consistency, particularly for the RF and GB methods, suggesting better generalization under more balanced distributions. On this basis, RF was selected as the most suitable method for structured operator selection. In practical terms, this means that customer profiles can be evaluated in a consistent manner across operators, which is particularly relevant in urban service environments where fragmented decision-making may lead to inefficiencies in service allocation.

4.5.1. Statistical Analysis of Model Differences

In addition to the standard indicators of model accuracy, presented through MAE, MSE and coefficient of determination R^2 , an additional statistical analysis of the differences between the applied regression models was conducted. The purpose of this analysis was to determine whether the observed differences in the performance of the LR, RF and GB models represent only descriptive differences in the values of the evaluation metrics or can be statistically confirmed as significant differences in the predictive behavior of the models. Statistical comparison was not conducted between different operators, since the datasets for the operators refer to different groups of users and, therefore, do not represent dependent samples. Instead, the analysis was conducted separately within each individual dataset: for Operator 1, Operator 2, Operator 3 and the extended resampled dataset. Such an approach is methodologically justified because within each dataset all three models generated predictions for the same test observations, which allows for a direct comparison of their predictive errors at the level of the same user profiles.

The basis for statistical comparison was the absolute prediction error, defined as the absolute difference between the actual user satisfaction rating and the value predicted by the corresponding model. In this way, three related error values were formed for each test sample, corresponding to the LR, RF and GB models. Given that three related models, i.e., three sets of errors obtained on the same test observations, are compared, the Friedman test was applied for the analysis. This non-parametric test is particularly suitable in situations when more than two related samples are compared and when it is not justified to assume a normal distribution of prediction errors. Its use is also consistent with recent comparative machine learning studies in which the Friedman test is applied to assess statistically significant differences among regression models evaluated using error-based performance measures [98].

The application of the Friedman test allows for a formal test of the null hypothesis according to which there are no statistically significant differences between the predictive errors of the observed models. The alternative hypothesis assumes that at least one of the models is statistically significantly different from the others in terms of the magnitude of the absolute prediction error. Since lower values of the absolute error indicate greater prediction accuracy, the model with the lowest median error and the lowest sum of ranks can be interpreted as the statistically most favorable within the observed data set.

Accordingly, the Friedman test was applied separately to each analyzed dataset. The results are presented in Table 9. For each dataset, the table reports the number of test observations, the median absolute error for each model, the Friedman test statistic, degrees of freedom, p -value, and the best-ranked model according to the obtained error structure. The median absolute error was included in the table because it provides a direct interpretation of the typical prediction deviation on the five-point satisfaction scale, while the Friedman statistic and p -value provide formal evidence on whether the differences among the models are statistically significant.

Table 9. Statistical comparison of absolute prediction errors among LR, RF, and GB models using the Friedman test.

Dataset	N	Median Absolute Error LR	Median Absolute Error RF	Median Absolute Error GB	Friedman Statistic (S)	df	p -Value	Best-Ranked Model
Operator 1	96	0.3557	0.2000	0.3979	42.25	2	<0.001	RF
Operator 2	64	0.2502	0.1720	0.4112	56.63	2	<0.001	RF
Operator 3	60	0.3455	0.2600	0.5360	11.43	2	0.003	RF
Extended dataset	2000	0.4749	0.0434	0.0473	1815.40	2	<0.001	RF

The results presented in Table 9 show that statistically significant differences between the LR, RF and GB models were identified in all analyzed datasets. For Operator 1, the Friedman test yielded $S = 42.25$ with $p < 0.001$, indicating that the differences in prediction errors among the three models are statistically significant. The RF model achieved the lowest median absolute error of 0.2000, while the corresponding errors for LR and GB were 0.3557 and 0.3979, respectively. This indicates that, for Operator 1, the RF model produced the most consistent and accurate predictions on the test observations.

A similar pattern was observed for Operator 2. The Friedman test result, $S = 56.63$ and $p < 0.001$, confirms statistically significant differences among the models. In this dataset, RF again achieved the lowest median absolute error, with a value of 0.1720. The LR model showed a higher median error of 0.2502, while the GB model had the highest median error of 0.4112. These results confirm that the RF model was not only descriptively better according to standard performance metrics, but also statistically more favorable when prediction errors were analyzed across the same test observations.

For Operator 3, the Friedman test also indicated statistically significant differences among the models, although the magnitude of the test statistic was lower than in the first two operator datasets. The obtained value was $S = 11.43$, with $p = 0.003$. This result is still below the conventional significance threshold of 0.05, which means that the null hypothesis of equal model performance can be rejected. The RF model again achieved the lowest median absolute error, equal to 0.2600. The LR model followed with a median absolute error of 0.3455, while the GB model showed the weakest performance, with a median error of 0.5360.

The most pronounced statistical difference was obtained for the extended resampled dataset, where the Friedman statistic reached $S = 1815.40$ with $p < 0.001$. This result is expected given the substantially larger number of test observations and the more balanced distribution of satisfaction ratings. In this dataset, both ensemble-based models achieved markedly lower errors than LR. The RF model had the lowest median absolute error, 0.0434, followed closely by GB with 0.0473. Although the difference between RF and GB was small in absolute terms, RF still retained the most favorable rank structure and was therefore identified as the best-ranked model. In contrast, LR showed a substantially higher median absolute error of 0.4749, confirming its weaker ability to capture nonlinear relationships in the extended dataset.

Overall, the statistical analysis confirms the findings obtained from MAE, MSE and R^2 metrics. The RF model achieved the lowest median absolute prediction error in all four analyzed datasets and was consistently identified as the best-ranked model. This consistency is important from both a methodological and decision-support perspective. It indicates that RF provides stable predictive performance across different empirical datasets as well as under more balanced target-variable conditions in the extended dataset. Therefore, the selection of RF for the final operator selection stage is not based only on descriptive model performance indicators, but is also supported by a formal statistical comparison of prediction errors.

Detailed outputs of the Friedman test for Operator 1, Operator 2, Operator 3, and the extended dataset are provided in Supplementary Files S1–S4.

4.5.2. Cross-Validated Analysis of Predictor Importance

To identify which customer and service-related attributes contributed most to the predictive performance of the Random Forest model, an additional predictor-importance analysis was conducted. The purpose of this analysis was to determine whether the model relied more strongly on particular characteristics of customer profiles, shipment-related variables, or service preferences when estimating customer satisfaction. Such information is important because overall model accuracy does not reveal which parts of the input data

carry the greatest predictive value. A clearer understanding of these relationships also supports a more meaningful interpretation of the model results and provides a basis for identifying service dimensions that may deserve greater managerial attention. Predictor importance was therefore examined using a cross-validated permutation procedure, and the main results are presented in Table 10.

Table 10. Most relevant predictors identified using cross-validated Random Forest permutation importance.

Dataset	Predictor	Mean Importance	95% Confidence Interval	Positive Fold Share	Interpretation
Operator 1	Customer category	0.0179	0.0153–0.0205	1.00	Stable dominant predictor
Operator 1	Shipment type	0.0035	0.0009–0.0060	0.64	Positive but weaker predictor
Operator 2	No stable dominant predictor	-	-	-	No variable showed stable positive importance
Operator 3	Expected delivery time	0.0097	−0.0027–0.0222	0.60	Highest relative importance, but not fully stable
Extended dataset	Expected delivery time	0.0481	0.0292–0.0670	1.00	Stable predictor under balanced conditions
Extended dataset	Customer category	0.0103	0.0051–0.0155	1.00	Stable predictor under balanced conditions

Predictor importance was assessed using a cross-validated permutation procedure. For each of the three original operator datasets, five-fold cross-validation was repeated five times, resulting in 25 external validation folds. Within each training fold, the Random Forest hyperparameters were selected through an internal five-fold GridSearchCV procedure. Each original predictor was then permuted 20 times in the corresponding validation fold, and the resulting increase in MAE was used as the measure of importance. This approach makes it possible to examine whether a predictor contributes consistently across different partitions of the data rather than relying on a single train–test split.

The interpretation of predictor importance was based on the indicators reported in Table 10: mean importance, the 95% confidence interval, and positive fold share. Mean importance expresses the average increase in prediction error after the values of a particular attribute were permuted. A higher positive value means that the model relied more strongly on that attribute when estimating customer satisfaction. The 95% confidence interval shows how stable this contribution was across the validation folds. When the entire interval remains above zero, the attribute can be considered a stable predictor. Positive fold share indicates the proportion of validation folds in which permutation of the attribute increased the prediction error. A value of 1.00 therefore means that the attribute contributed positively in every fold, while lower values indicate that its predictive contribution was less consistent across different data partitions.

The results indicate that the predictive relevance of customer and service attributes differed across operators. For Operator 1, customer category was the clearest and most stable predictor. Its mean importance was 0.0179, the 95% confidence interval ranged from 0.0153 to 0.0205, and the positive fold share was 1.00. This shows that the distinction between individual customers, business entities, and institutional users consistently contributed to the model's ability to differentiate satisfaction levels for this operator. Shipment type also contributed positively, with a mean importance of 0.0035, a confidence interval from 0.0009 to 0.0060, and a positive fold share of 0.64. Its contribution was therefore clearly weaker and less consistent than that of customer category.

For Operator 2, no single attribute showed a sufficiently stable positive contribution to be identified as a dominant predictor. The importance values varied across the validation folds, indicating that the model did not rely consistently on one customer characteristic. This suggests that satisfaction prediction for Operator 2 may depend on combinations of several attributes rather than on one clearly distinguishable variable. Complete results for all attributes are provided in the Supplementary Materials.

For Operator 3, expected delivery time had the highest mean importance and therefore represented the most prominent predictor in this dataset. Its mean importance was 0.0097, the 95% confidence interval ranged from -0.0027 to 0.0222 , and the positive fold share was 0.60. Although expected delivery time carried more predictive information than the remaining attributes, the confidence interval included zero and the importance was positive in only 60% of the validation folds. Its contribution was therefore not equally stable across all data partitions.

The extended dataset showed a more clearly defined importance structure. Expected delivery time was a stable predictor, with a mean importance of 0.0481, a 95% confidence interval from 0.0292 to 0.0670, and a positive fold share of 1.00. Customer category also showed a stable contribution, with a mean importance of 0.0103, a confidence interval from 0.0051 to 0.0155, and positive importance in every validation fold. Although shipment type and product category had higher mean values than customer category, their confidence intervals included zero, which indicates that their contributions varied more strongly across the grouped validation folds.

Taken together, the results do not point to one universal attribute that was equally important for all operators. Instead, the predictive structure was operator-specific. Customer category was the most relevant attribute for Operator 1, while expected delivery time was the most prominent attribute for Operator 3 and remained stable in the extended dataset. For Operator 2, no dominant predictor was identified. This distinction is important for practical application because it indicates that logistics operators should not rely on the same customer-segmentation or service-adjustment strategy. Operator 1 may benefit more from differentiating services according to customer category, whereas Operator 3 should give greater attention to the delivery-time expectations of its users. Given the breadth of the operator-specific results, the present study focuses on the most stable and practically relevant predictors, while a separate analysis could provide a more detailed examination of the complete importance profiles, rank variations, and fold-level patterns across all operators.

Detailed cross-validated permutation-importance results for all customer and service attributes across the three operator datasets and the extended dataset are provided in Table S5 in the Supplementary Materials.

4.6. Phase V: Selection of an Appropriate Operator According to Customer Preferences

After evaluating and verifying model performance, it was established that the RF method achieved the best results in predicting customer satisfaction. Therefore, it was applied in the final phase to predict satisfaction ratings for identical customer profiles, separately for each operator, in order to select the operator that best fits each customer profile.

For demonstration purposes, three different customer profiles were selected.

The first customer, Customer 1, represents a female (F) individual (C2 or 2C), aged between 18 and 25 years (Category I), with secondary education, who uses a postal and logistics service for sending or receiving a parcel of clothing. The pricing preference is per item, and delivery speed is the primary criterion for operator selection, with an expected service completion time of two days ($D + 1$).

The second customer, Customer 2, is a male (M) individual (C2 or 2C), aged 31 to 40 (Category III), with secondary education and the same pricing preference. He uses

the service for sending or receiving a parcel of car parts, with price as the main selection criterion and an expected service time of three days ($D + 2$).

The third customer, Customer 3, represents a company (C), a legal entity (B2 or 2B), operating for 6–10 years (Category II). The company representative authorized to complete the service request holds a university degree. The company requests a parcel pick-up or delivery service for cosmetics, with price as the main criterion and an expected service time of three days ($D + 2$).

The RF model was applied individually to each of these customer profiles for Operators 1, 2, and 3. The predicted satisfaction ratings generated by the RF model, together with the corresponding operator recommendations, are shown in Table 11.

Table 11. Results of the RF model in predicting satisfaction ratings for identical customer profiles across Operators 1, 2, and 3, with recommendations for best operator selection.

Customer Profile	RF Model			Recommended Operator
	Operator 1	Operator 2	Operator 3	
Customer 1	4.13	3.74	3.75	Operator 1
Customer 2	4.27	3.98	3.62	Operator 1
Customer 3	4.65	4.78	4.03	Operator 2

A detailed analysis of the results indicates that for Customer 1, the RF model predicted the highest satisfaction rating for Operator 1 (4.13), compared to Operator 2 (3.74) and Operator 3 (3.75), making Operator 1 the recommended choice. Using the same selection criteria, Customer 2 is also recommended for Operator 1, with the highest predicted satisfaction rating (4.27). Conversely, for Customer 3, the RF model predicted the highest satisfaction rating for Operator 2 (4.78), making it the most suitable choice according to customer preferences and requirements.

In the selection process, only models with predicted satisfaction ratings within the range of 3.6 to 5.5 were considered, corresponding to categories of satisfied and highly satisfied customers. Operators with lower predicted values indicating neutral or dissatisfied customers were excluded from the final recommendation. Although presented through illustrative customer profiles, this selection mechanism demonstrates how structured predictive evaluation can reduce arbitrary or purely perception-based operator choice.

5. Discussion

Postal and logistics systems, while traditionally assessed through market efficiency and competitiveness, operate in contemporary cities under spatial and environmental constraints that extend beyond purely business logic [1–3,18,99]. Service organization therefore influences not only business performance, but also the structure and intensity of urban freight movements [47–51]. At the same time, operators must respond to increasingly individualized customer expectations while maintaining operational and financial stability [41–44].

Customer satisfaction is not treated merely as a short-term service outcome. It is considered a variable shaping loyalty, demand distribution and operator positioning [45,46], and therefore indirectly influencing the organization of freight flows in urban areas. Understanding this linkage becomes particularly relevant in dense cities where small variations in demand allocation may translate into changes in delivery frequency and traffic patterns. Accordingly, the aim of this paper was to develop a methodological framework that enables formalized customer profiling and quantitative prediction of their satisfaction using machine learning methods, with the aim of supporting the process of selecting a postal and logistics operator while simultaneously respecting the interests of both customers and operators. This modelling direction is also consistent with the research agenda outlined in

our earlier review of machine learning applications in postal and logistics systems, where the development of predictive models for customer satisfaction and operator-oriented decision support was identified as a relevant avenue for future research [100].

However, within the urban context previously discussed [1–3,18,47–51] this objective is not limited to business alignment alone. By structuring operator selection through predictive modeling, the framework is positioned within broader urban freight dynamics, where individual service choices may influence transport intensity, delivery organization and spatial distribution of flows in cities. The review of existing literature has shown that the characteristics of customer profiles in postal and logistics systems have been partially analyzed. Some works have focused on demographic and geodemographic characteristics of customers such as gender, age, income, education, urban or rural environment of service use, and the like [26–35]. Another group of works has partially treated customer requirements in terms of service provision, where criteria for on-time delivery of parcels, security, price, and the use of ICT are listed [7,9,36–38]. Although these works contribute to the characteristics of customer behavior, none of these studies provides a formalized description of the customer profile that would be used in quantitative and/or predictive research models. In relation to the aforementioned research, the first scientific contribution of this work is reflected in the integration and systematized approach to customer profiling. Instead of observing individual characteristics in isolation, the customer profile is mathematically defined as a structured set of interconnected characteristics, which can be directly used as an input to further statistical and predictive models. In this way, customer-related variables are not treated as exogenous descriptors, but as analytically structured components capable of being linked with operational and potentially environmental dimensions of urban service systems.

Traditional evaluation frameworks, such as SERVQUAL and Kansei Engineering, as well as recent fuzzy multi-criteria approaches for logistics service quality assessment, successfully identify determinants of satisfaction such as reliability, security, staff professionalism and service availability [61,62,101], but are limited in terms of their application in predictive decision-making. On the other hand, machine learning applications in the postal and logistics sector are predominantly focused on the operator perspective, including customer satisfaction evaluation [55], customer churn prediction [55], price and demand forecasting [56], cargo classification [57] or product grouping [57]. Certain contributions address customer-oriented dimensions more directly, for example through satisfaction evaluation using hybrid approaches [59,60] or through socio perceptual analysis based on social media content [71]; however, these approaches are generally not embedded within structured operator selection or broader urban logistics frameworks.

Although some works combine the customer and operator perspectives, for example through analyses of last-mile delivery [63], optimization of the postal center network [64], or delivery scheduling and route optimization [102], customer satisfaction in these cases is mostly considered indirectly, through operational metrics of time, price and reliability. In [63], a combination of machine learning methods is used to predict the arrival time of a shipment, providing customers with more informative delivery estimates and enabling operators to allocate resources more efficiently, while in [64] a spatial-analytical approach is applied to optimize the locations of postal centers with the aim of improving network efficiency and service availability. The integrated framework presented in [65], as well as the concept of sustainable delivery based on AI optimization and electric vehicles [66,103], already move beyond purely operational efficiency by incorporating environmental and cost dimensions into system design. Recent intelligent logistics decision-support systems also combine machine learning with multi-criteria and uncertainty-based approaches to support circular and sustainable supply-chain decisions [104]. However, even within

these approaches, customer satisfaction remains positioned as a secondary or indirect outcome, rather than as an explicitly modeled variable influencing system configuration. This observation becomes particularly relevant when considering the broader evolution of machine learning research in logistics. In recent years, the focus has gradually expanded from efficiency-oriented optimization towards sustainability related questions. ML-based studies increasingly address emission modelling, fuel consumption reduction, routing under environmental constraints and the spatial organization of urban freight systems [2,67–70,102]. In parallel, certain contributions examine public perception of freight transport impacts in urban environments [71]. In this sense, the identified gaps in the purpose of existing models are filled by this work by developing a framework that simultaneously encompasses the context of the customer and the operator, with customer satisfaction being explicitly modeled as a measurable output variable and used as a connecting element for personalized recommendation of the postal and logistics operator and future system performance analysis. The original methodological contribution of this paper is therefore not reflected only in the application of machine learning algorithms to customer satisfaction prediction, but in the way in which the prediction problem is formulated and connected with operator selection. Unlike studies that analyze customer satisfaction, delivery performance or operator efficiency as separate research problems, the proposed framework formalizes customer profiling as a structured input layer, models satisfaction as a measurable predictive output, and uses the predicted satisfaction values as a decision-support basis for comparing postal and logistics operators. In this way, the paper introduces an integrated behavioral-predictive framework that links customer characteristics, service perceptions and operator selection within a single analytical procedure. Accordingly, the methodological novelty does not arise from modifying the internal structure of the LR, RF, or GB algorithms, which are applied in their established forms, but from their integration into a newly structured decision-support procedure that transforms a formalized customer profile into predicted satisfaction values and subsequently into an operator-selection recommendation.

The observed differences among the models can be interpreted in relation to the structure of the customer data and the learning mechanisms of the applied algorithms. Customer satisfaction is unlikely to depend on one attribute in isolation or to follow a strictly linear pattern. The relevance of a customer characteristic may change according to shipment type, service expectations, or other elements of the customer profile. Linear Regression is less suited to such relationships because it represents predictor effects mainly through an additive linear form and cannot capture nonlinearities or interaction effects unless they are specified in advance. Random Forest, by contrast, can identify nonlinear relationships, threshold effects, and interactions directly from the data without imposing a predefined functional form. Its bootstrap aggregation and random feature selection reduce the influence of individual observations and limit dependence on a single tree structure, which is particularly relevant for the relatively small, heterogeneous, and imbalanced operator-specific datasets used in this study. Gradient Boosting is capable of representing similar complexity, but its sequential learning process, in which each new tree attempts to correct the residual errors of the preceding model, may make it more sensitive to sample structure, target imbalance, and parameter tuning. The improvement of Gradient Boosting under more balanced data conditions suggests that its weaker results on the operator-specific datasets were partly associated with sample size and target imbalance. Random Forest, by contrast, showed greater robustness across different data structures, which supports its use in the operator-selection stage.

The predictor-importance analysis provides a basis for interpreting how customer satisfaction is formed differently across operator contexts. For Operator 1, the dominant role of customer category suggests that individual customers, business users, and institutional

customers evaluate the same service through different usage conditions and expectations. Their satisfaction may therefore reflect differences in shipment frequency, communication requirements, contractual flexibility, delivery reliability, and the degree of service customization expected from the operator. The additional contribution of shipment type indicates that customer evaluation also depends on the operational context in which the service is used. For Operator 3, expected delivery time carried the greatest predictive value, suggesting that satisfaction is strongly influenced by the alignment between the promised delivery period and the time customers consider acceptable. This points to the importance of managing expectations, rather than focusing only on delivery speed. For Operator 2, the absence of a single stable predictor indicates a more complex and compensatory evaluation pattern, in which several customer and service characteristics jointly shape satisfaction. The extended dataset further supports the relevance of expected delivery time and customer category under more balanced target-variable conditions. Taken together, these findings show that customer behavior is not uniform across the market. Satisfaction reflects different combinations of customer type, service expectations, and shipment conditions, which means that the same segmentation and service-adjustment strategy should not be applied uniformly across operators.

Beyond predictive performance, the results highlight the analytical relevance of explicitly modelling customer satisfaction as an outcome variable within logistics research. In many existing studies, customer-related aspects are incorporated only indirectly through operational indicators such as delivery time, service price, or reliability [7,9,36–38]. In contrast, the approach adopted in this study positions customer satisfaction as a measurable and predictable component of the urban logistics system. By doing so, the model provides a structured mechanism through which customer-related demand patterns can be analyzed and linked with the broader organization of delivery services.

From a city logistics perspective, such modelling is particularly relevant because customer behavior represents an important driver of delivery demand patterns in urban areas. The ability to predict satisfaction across different operator configurations allows the identification of customer segments whose expectations are more closely aligned with particular service profiles. This introduces the possibility of a more structured allocation of demand across logistics operators, which may contribute to more balanced utilization of delivery capacities and a more stable organization of urban freight flows. Since failed deliveries and repeated delivery attempts are recognized as sources of additional transport activity, congestion, fuel consumption, and emissions [78,79,82], the proposed framework may provide a behavioral basis for future sustainability-oriented analyses, rather than direct evidence of environmental improvement.

From a theoretical perspective, the study contributes to the emerging body of research that integrates behavioral and operational perspectives within urban logistics analysis. While traditional city logistics models predominantly focus on infrastructure design, network optimization, or policy instruments [3,18], the behavioral dimension of logistics demand has received comparatively less attention. By formalizing customer profiles and modelling satisfaction outcomes through machine learning techniques, the proposed framework introduces a behavioral analytical layer that complements existing operational approaches.

In this sense, the study contributes to bridging the gap between micro-level customer preferences and the macro-level organization of freight transport activities in cities. The predictive modelling of satisfaction provides a quantitative representation of demand-side dynamics that can be integrated into broader urban logistics planning frameworks. Such integration is consistent with recent research trends that emphasize the need for data-driven approaches to support sustainable and adaptive urban freight systems [66,70,103].

These operator-specific findings can be translated into differentiated operational responses. For Operator 1, communication channels, delivery options, contractual arrangements, and service packages could be adjusted according to customer category and further refined by shipment type. For Operator 3, operational attention should focus on aligning promised delivery periods with customer expectations through clearer communication, more realistic service commitments, and alternative delivery-time options. For Operator 2, managerial decisions should rely on combined customer profiles and service conditions rather than on a single segmentation criterion. More generally, predictive satisfaction modelling can help operators identify customer profiles associated with lower expected satisfaction and introduce targeted service adjustments before dissatisfaction results in complaints, customer loss, or a change in operator. In practical terms, these findings can be incorporated into a simple decision sequence: first, operators identify customer profiles associated with lower predicted satisfaction; second, they prioritize the service dimensions indicated by the predictor-importance results; third, they implement targeted adjustments; and finally, they monitor whether these changes improve satisfaction and related operational outcomes. Predictor importance should therefore be treated as a basis for prioritization and testing, rather than as direct evidence of causality. Where operational data are available, the effects of such adjustments can be evaluated through indicators such as first-attempt delivery success, rescheduling, parcel redirection, and repeated delivery frequency. At the same time, urban logistics planners and policy makers may benefit from a better understanding of the behavioral patterns that influence delivery demand and service distribution across the city. When combined with spatial and operational data, these insights could support the development of more adaptive and demand-responsive urban freight systems.

6. Conclusions

The rapid expansion of e-commerce and increasing demand for parcel delivery services have significantly transformed urban logistics systems, intensifying last-mile delivery activities in cities and raising new challenges related to congestion, operational efficiency and environmental pressures. In such a context, understanding behavioral demand patterns and customer expectations becomes increasingly important for improving the organization of urban parcel delivery services.

This paper addressed these challenges through the development and empirical evaluation of a machine-learning-based framework for predicting customer satisfaction with postal and logistics operators. The proposed comparative modelling approach evaluates Linear Regression, Random Forest and Gradient Boosting algorithms in order to analyze how customer characteristics and service perceptions influence satisfaction outcomes and operator preferences in urban delivery markets.

The empirical results demonstrate that machine learning techniques can effectively capture complex relationships between customer attributes and satisfaction levels across different logistics operators. Among the evaluated methods, the Random Forest model achieved the highest predictive accuracy and stability. On real datasets, RF achieved MAE = 0.264, MSE = 0.113 and $R^2 = 0.791$ for Operator 1; MAE = 0.264, MSE = 0.116 and $R^2 = 0.714$ for Operator 2; and MAE = 0.317, MSE = 0.246 and $R^2 = 0.769$ for Operator 3. On the extended dataset generated through synthetic resampling, predictive performance further improved (MAE = 0.159, MSE = 0.100, $R^2 = 0.936$), indicating improved internal predictive stability of ensemble-based learning under more balanced target-variable conditions.

Beyond predictive performance, the results indicate that behavioral customer data can provide valuable insights into demand patterns that influence the use of urban delivery services. By modelling satisfaction across different operators, the framework enables the

identification of customer segments whose expectations are more closely aligned with specific service characteristics, which may support more informed decision-making in the organization of urban parcel delivery systems. The study provides several interrelated scientific contributions. First, it introduces a structured analytical representation of customer profiles that integrates demographic, geodemographic and service-related attributes into a unified modelling framework suitable for predictive analysis. Second, it develops and empirically evaluates a comparative machine learning framework based on LR, RF and GB algorithms for predicting customer satisfaction and supporting operator selection in postal and logistics markets. Third, the study contributes to the growing body of research integrating behavioral demand analysis with logistics system modelling by explicitly treating customer satisfaction as a quantifiable outcome variable linking customer characteristics with logistics service performance. The statistical comparison of prediction errors further strengthens the methodological contribution by confirming that the observed differences among the models were not only descriptive but statistically significant. In addition, the predictor-importance analysis extends the interpretation of the model by identifying the customer and service attributes that contributed most consistently to satisfaction prediction across the analyzed operators.

From the perspective of urban logistics systems, the broader contribution of the study lies in highlighting the role of behavioral demand patterns in shaping the organization of urban freight activities. Customer preferences related to delivery speed, flexibility and service quality increasingly influence the distribution of parcel delivery demand across logistics operators in cities. By modelling customer satisfaction across multiple operators, the proposed framework provides analytical insight into how these demand patterns are structured within urban delivery markets.

Although environmental indicators were not directly embedded into the predictive models, the framework establishes a behavioral analytical layer that can support more comprehensive urban logistics analyses. In this sense, the model contributes to bridging micro-level customer behavior with the broader organization of urban freight systems, providing a foundation for future integration with operational logistics data and sustainability-oriented planning approaches.

In addition to its theoretical and methodological relevance, the model offers several practical implications for logistics operators and urban freight system stakeholders. For logistics operators, predictive modelling of customer satisfaction enables a more detailed understanding of customer segmentation, identification of service dimensions that influence satisfaction levels, and the development of delivery strategies better aligned with customer expectations. Such insights may support improved service differentiation and more adaptive delivery planning in competitive urban logistics markets.

From a broader city logistics perspective, better understanding of behavioral demand patterns may also assist urban planners and policymakers in analyzing how customer expectations influence the distribution and intensity of parcel delivery activities in urban areas. When combined with operational logistics data, such behavioral insights could contribute to the development of more data-driven and adaptive urban freight management strategies.

Despite the contributions of the study, several limitations should be acknowledged. The proposed framework is methodologically transferable, since the same modelling procedure can be applied to other countries, operators and time periods, provided that new context-specific data are collected and the models are retrained and validated accordingly. However, the empirical results obtained in this study should be interpreted within the specific national and temporal context in which the survey was conducted. Since the data were collected in one country and during one time period, the obtained model performance and identified satisfaction patterns cannot be assumed to be universally generalizable.

without additional external validation. The empirical analysis relies on survey-based data describing customer characteristics and perceptions, while operational delivery data and environmental performance indicators were not directly incorporated into the predictive modelling framework. As a result, the relationships proposed between customer satisfaction, delivery efficiency, and environmental performance remain conceptual rather than empirically validated. No logistics performance indicators, failed-delivery data, delivery success rates, transport activity, or environmental measures were included in this study, and the sustainability contribution should therefore be understood as indirect.

Future research should therefore focus on integrating behavioral prediction models with operational logistics datasets, including first-attempt delivery success, failed delivery rates, delivery frequency, travelled distance, routing characteristics, vehicle types, energy consumption, emissions, and other environmental performance indicators. Such integration would enable a more comprehensive examination of how customer-driven demand patterns influence the efficiency and environmental performance of urban freight transport systems, thereby strengthening the analytical link between behavioural modelling and city-environment interactions. In addition, future research could compare the proposed regression-based approach with ordinal regression and ordinal classification models to assess the robustness of predictive performance and operator rankings under alternative modelling formulations. Future studies should also include cross-country and longitudinal datasets in order to examine the temporal stability, external validity and broader transferability of the proposed modelling framework.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su18157786/s1>, Table S1: Regression results Operator 1; Table S2: Regression results Operator 2; Table S3: Regression results Operator 3; Table S4: Regression results Extended dataset; Table S5: Detailed cross-validated Random Forest permutation-importance results; Supplementary File S1: Detailed Friedman test results for Operator 1; Supplementary File S2: Detailed Friedman test results for Operator 2; Supplementary File S3: Detailed Friedman test results for Operator 3; Supplementary File S4: Detailed Friedman test results for the extended dataset.

Author Contributions: Conceptualization, A.M., A.K. (Amel Kosovac), E.M. and M.B.; methodology, A.M.; software, A.M. and M.B.; validation, A.K. (Amel Kosovac), E.M. and M.B.; formal analysis, A.M., A.K. (Aida Kalem), M.K. and S.T.; investigation, A.M.; resources, A.M., S.T. and M.K.; data curation, A.M. and M.B.; writing—original draft preparation, A.M., E.M. and M.B.; writing—review and editing, A.K. (Amel Kosovac), S.T. and M.K.; visualization, A.M.; supervision, A.K. (Amel Kosovac), E.M., M.B., S.T. and M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study by the Ethics Committee of the Faculty of Transport and Communications, University of Sarajevo, due to this study being an anonymous, voluntary, non-interventional survey research that posed minimal risk, and in accordance with the institutional rules and practices applicable to such studies.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All relevant data are presented within the paper or in the Supplementary Materials. The original survey data ($n = 1239$) were collected in the Bosnian language, reflecting the linguistic and market context of the target user base. Due to the contextual specificity and the nature of user-related information, the complete dataset, as well as the derived datasets for Operators 1, 2, and 3 and the extended dataset, are available from the corresponding author upon request.

Acknowledgments: Computational experiments were conducted in Google Colab (Python 3.10) using the scikit-learn, pandas, and NumPy libraries. An extended dataset was created programmatically by resampling complete observations to address target-variable imbalance and examine model behavior

under more balanced conditions. The author reviewed and validated all results and takes full responsibility for the final content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial neural network
B2B	Business-to-business
B2C	Business-to-consumer
BPMN	Business Process Model and Notation
C2C	Consumer-to-consumer
CEP	Courier, express, and parcel
CFA	Confirmatory factor analysis
CRM	Customer relationship management
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
EFA	Exploratory factor analysis
EV	Electric vehicle
GB	Gradient Boosting
GIS	Geographic information system
ICT	Information and communication technologies
KNN	K-nearest neighbors
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
ML	Machine learning
MSE	Mean Squared Error
NLP	Natural language processing
PCA	Principal component analysis
RF	Random Forest
R ²	Coefficient of determination
SEM	Structural equation modelling
SERVQUAL	Service Quality model
SVM	Support Vector Machine
SVD	Singular value decomposition
t-SNE	t-distributed Stochastic Neighbor Embedding
VADER	Valence Aware Dictionary and Sentiment Reasoner
XAI	Explainable artificial intelligence
XGBoost	Extreme Gradient Boosting

Appendix A

The tuned hyperparameters for each of the implemented models are presented in Table A1.

Table A1. Model Parameters.

Algorithm	Parameters
LR	No tuned parameters (baseline linear regression)
RF	n_estimators = {50, 100, 200}, max_depth = {None, 10, 20}, random_state = 42, scoring = 'neg_mean_squared_error', cv = 5, n_jobs = -1
GB	n_estimators = {50, 100, 200}, learning_rate = {0.01, 0.1, 0.2}, max_depth = {3, 4, 5}, random_state = 42, scoring = 'neg_mean_squared_error', cv = 5, n_jobs = -1
General settings	Data split = 80:20 (train test), cross-validation = 5 folds, evaluation metrics = MAE, MSE, R ²

References

1. Zečević, S.; Tadić, S. *City Logistika*; Univerzitet u Beogradu, Saobraćajni Fakultet: Beograd, Serbia, 2013.
2. Kosovac, A.; Medić, A.; Begović, M. Machine learning modeling for reducing greenhouse gas emissions in urban areas. In *Proceedings of the Second International Conference on Advances in Traffic and Communication Technologies (ATCT 2023): Sustainable Urban Mobility, Sarajevo, Bosnia and Herzegovina, 24–26 May 2023*; University of Sarajevo, Faculty of Traffic and Communications: Sarajevo, Bosnia and Herzegovina, 2023; pp. 131–136. [\[CrossRef\]](#)
3. Tadić, S.; Krstić, M.; Kovač, M. Assessment of city logistics initiative categories sustainability: Case of Belgrade. *Environ. Dev. Sustain.* **2023**, *25*, 1383–1419. [\[CrossRef\]](#)
4. Tadić, S.; Zečević, S. *Modelling City Logistics Concepts*, 1st ed.; University of Belgrade, Faculty of Transport and Traffic Engineering: Belgrade, Serbia, 2016.
5. Liu, J.; Yuan, C.; Hafeez, M.; Yuan, Q. The relationship between environment and logistics performance: Evidence from Asian countries. *J. Clean. Prod.* **2018**, *204*, 282–291. [\[CrossRef\]](#)
6. Tadić, S.R.; Zečević, S.M.; Krstić, M.D. City logistics initiatives aimed at improving sustainability by changing the context of urban area. *Tehnika* **2014**, *69*, 834–843. [\[CrossRef\]](#)
7. Bajec, P.; Tuljak-Suban, D. Selecting a logistics service provider: A definition of criteria that consider the requirements of an external competitive environment. *Transp. Probl.* **2017**, *12*, 157–168. [\[CrossRef\]](#)
8. Karcz, J.; Ślusarczyk, B. Criteria of quality requirements deciding on choice of the logistic operator from a perspective of his customer and the end recipient of goods. *Prod. Eng. Arch.* **2021**, *27*, 58–68. [\[CrossRef\]](#)
9. Wei Hong, L.W.; Zheng, X.P.; Pan, X. Analyzing the relationship between consumer satisfaction and fresh e-commerce logistics service using text mining techniques. *Sustainability* **2019**, *11*, 3570. [\[CrossRef\]](#)
10. Tadić, S.; Zečević, S.; Krstić, M. Assessment of the political city logistics initiatives sustainability. *Transp. Res. Procedia* **2018**, *30*, 285–294. [\[CrossRef\]](#)
11. Johansson, H. *Customer Benefits in City Logistics: Towards Viable Urban Consolidation Centres*; Linköping University Electronic Press: Linköping, Sweden, 2020; Volume 2042. [\[CrossRef\]](#)
12. De Carvalho, P.P.S.; De Araújo Kalid, R.; Rodríguez, J.L.M. Evaluation of the city logistics performance through structural equations model. *IEEE Access* **2019**, *7*, 121081–121094. [\[CrossRef\]](#)
13. Kosovac, A.; Grabovica, E.; Medić, A.; Kalem, A. Proposal of CRM conceptual model/framework to optimise relations among stakeholders of higher education. *Ekonom. Misao Praksa* **2022**, *31*, 665–685. [\[CrossRef\]](#)
14. Ngai, E.W.T.; Xiu, L.; Chau, D.C.K. Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Syst. Appl.* **2009**, *36*, 2592–2602. [\[CrossRef\]](#)
15. Jiang, L.; Mahmassani, H. City logistics. *Transp. Res. Rec.* **2014**, *2410*, 85–95. [\[CrossRef\]](#)
16. Amaral, R.R.; Aghezzaf, E.-H. City logistics and traffic management: Modelling the inner and outer urban transport flows in a two-tiered system. *Transp. Res. Procedia* **2015**, *6*, 297–312. [\[CrossRef\]](#)
17. Tadić, S.; Zečević, S.; Krstić, M. A novel hybrid MCDM model based on fuzzy DEMATEL, fuzzy ANP and fuzzy VIKOR for city logistics concept selection. *Expert Syst. Appl.* **2014**, *41*, 8112–8128. [\[CrossRef\]](#)
18. Lee, C.K.H. A GA-based optimisation model for big data analytics supporting anticipatory shipping in Retail 4.0. *Int. J. Prod. Res.* **2017**, *55*, 593–605. [\[CrossRef\]](#)
19. Tadić, S.; Veljović, M. Home delivery: A framework for structuring. *Int. J. Traffic Transp. Eng.* **2021**, *11*, 30–74. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Multaharju, S.; Hallikas, J. Logistics service capabilities of logistics service provider. *Int. J. Logist. Syst. Manag.* **2015**, *20*, 103–121. [\[CrossRef\]](#)
21. Szczucka-Lasota, B. City logistics: Influence of oversized road transport on urban development. *Sci. J. Silesian Univ. Technol. Ser. Transp.* **2017**, *97*, 157–165. [\[CrossRef\]](#)
22. Hanus, P. The business profile shaping and the logistics information systems of 2PL, 3PL, 4PL operators. *J. Econ. Manag.* **2013**, *12*, 5–21.
23. Gruchmann, T.; Melkonyan, A.; Krumme, K. Logistics business transformation for sustainability: Assessing the role of the lead sustainability service provider (6PL). *Logistics* **2018**, *2*, 25. [\[CrossRef\]](#)
24. Kharat, P.P.; Nagare, P.M.R. Business development—B2B and B2C ecommerce. *Int. J. Res. Publ. Rev.* **2021**, *2*, 999–1002.
25. Tadić, S.; Veljović, M. Home delivery: Concept and characteristics. *Int. J. Traffic Transp. Eng.* **2020**, *10*, 519–533. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Ouafthouh, S.; Zellou, A.; Idri, A. User profile model: A user dimension based classification. In *Proceedings of the 10th International Conference on Intelligent Systems: Theories and Applications (SITA)*; IEEE: Piscataway, NJ, USA, 2015; pp. 1–5. [\[CrossRef\]](#)
27. Gauch, S.; Speretta, M.; Chandramouli, A.; Micarelli, A. User profiles for personalized information access. In *The Adaptive Web; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4321, pp. 54–89. [\[CrossRef\]](#)
28. Mortimer, G.; Hasan, S.F.; Andrews, L.; Martin, J. Online grocery shopping: The impact of shopping frequency on perceived risk. *Int. Rev. Retail Distrib. Consum. Res.* **2016**, *26*, 202–223. [\[CrossRef\]](#)

29. Zhou, Y.; Wang, X. Explore the relationship between online shopping and shopping trips: An analysis with the 2009 NHTS data. *Transp. Res. Part A* **2014**, *70*, 1–9. [\[CrossRef\]](#)
30. Motte-Baumvol, B.; Belton-Chevallier, L.; Dablanc, L.; Morganti, E. Spatial dimensions of E-shopping in France. *Asian Transp. Stud.* **2017**, *4*, 585–600.
31. Weltevreden, J.W.J. Substitution or complementarity? How the Internet changes city centre shopping. *J. Retail. Consum. Serv.* **2007**, *14*, 192–207. [\[CrossRef\]](#)
32. Cao, X.; Chen, Q.; Choo, S. Geographic distribution of E-shopping application of structural equation models in the Twin Cities of Minnesota. *Transp. Res. Rec.* **2013**, *2383*, 18–26. [\[CrossRef\]](#)
33. Kirby-Hawkins, E.; Birkin, M.; Clarke, G. An investigation into the geography of corporate e-commerce sales in the UK grocery market. *Environ. Plan. B Urban Anal. City Sci.* **2019**, *46*, 1148–1164. [\[CrossRef\]](#)
34. Farag, S.; Weltevreden, J.; van Rietbergen, T.; Dijst, M.; van Oort, F. E-shopping in the Netherlands: Does geography matter? *Environ. Plan. B Plan. Des.* **2006**, *33*, 59–74. [\[CrossRef\]](#)
35. Clarke, G.; Thompson, C.; Birkin, M. The emerging geography of e-commerce in British retailing. *Reg. Stud. Reg. Sci.* **2015**, *2*, 370–390. [\[CrossRef\]](#)
36. Phuong Vu, T.; Grant, D.B.; Menachof, D.A. Exploring logistics service quality in Hai Phong, Vietnam. *Asian J. Shipp. Logist.* **2020**, *36*, 54–64. [\[CrossRef\]](#)
37. Mentzer, J.T.; Flint, D.J.; Hult, G.T.M. Logistics service quality as a segment-customized process. *J. Mark.* **2001**, *65*, 82–104. [\[CrossRef\]](#)
38. Langley, C.J., Jr.; Holcomb, M.C. Creating logistics customer value. *J. Bus. Logist.* **1992**, *13*, 2–29.
39. Akyuz, G.A.; Erkan, T.E. Supply chain performance measurement: A literature review. *Int. J. Prod. Res.* **2010**, *48*, 5137–5155. [\[CrossRef\]](#)
40. Krauth, E.; Moonen, H.; Popova, V.; Schut, M. Performance indicators in logistics service provision and warehouse management: A literature review and framework. In Proceedings of the EurOMA International Conference, Budapest, Hungary, 19–22 June 2005; pp. 19–22.
41. Frazelle, E. *Supply Chain Strategy: The Logistics of Supply Chain Management*; McGraw-Hill: New York, NY, USA, 2002.
42. Eklof, J.; Podkorytova, O.; Malova, A. Linking customer satisfaction with financial performance: An empirical study of Scandinavian banks. *Total Qual. Manag. Bus. Excell.* **2020**, *31*, 1684–1702. [\[CrossRef\]](#)
43. Erkan, B. The importance and determinants of logistics performance of selected countries. *J. Emerg. Issues Econ. Financ. Bank.* **2014**, *3*, 1237–1254.
44. Eke, C.I.; Norman, A.A.; Shuib, L.; Nweke, H.F. A survey of user profiling: State-of-the-art, challenges, and solutions. *IEEE Access* **2019**, *7*, 144907–144924. [\[CrossRef\]](#)
45. Tadić, S.R.; Zečević, S.; Krstić, M.D. City logistics initiatives aimed at improving sustainability within the existing context of urban area. *Teh.-Transp.* **2014**, *61*, 487–495. [\[CrossRef\]](#)
46. Huma, S.; Ahmed, W.; Ikram, M.; Khawaja, M.I. The effect of logistics service quality on customer loyalty: Case of logistics service industry. *South Asian J. Bus. Stud.* **2019**, *9*, 43–61. [\[CrossRef\]](#)
47. Hallowell, R. The relationships of customer satisfaction, customer loyalty, and profitability: An empirical study. *Int. J. Serv. Ind. Manag.* **1996**, *7*, 27–42. [\[CrossRef\]](#)
48. Mariano, E.B.; Gobbo, J.A., Jr.; de Castro Camiato, F.; do Nascimento Rebelatto, D.A. CO2 emissions and logistics performance: A composite index proposal. *J. Clean. Prod.* **2017**, *163*, 166–178. [\[CrossRef\]](#)
49. Zaman, K.; Shamsuddin, S. Green logistics and national scale economic indicators: Evidence from a panel of selected European countries. *J. Clean. Prod.* **2017**, *143*, 51–63. [\[CrossRef\]](#)
50. Mkasimongwa, S.W.; Livesley, S.J.; Ryan, R.G.; Schofield, R. Air pollution exceedance events in Melbourne and Sydney, Australia, between 2000 and 2024. *City Environ. Interact.* **2025**, *28*, 100239. [\[CrossRef\]](#)
51. Bogaert, P.; Huvelle, N.; Briffault, A.; Basseur, O. Modeling nitrogen dioxide concentrations using citizen science data: The case of the Brussels-Capital Region. *City Environ. Interact.* **2025**, *28*, 100236. [\[CrossRef\]](#)
52. DeGregory, K.W.; Kuiper, P.; DeSilvio, T.; Pleuss, J.D.; Miller, R.; Roginski, J.W.; Fisher, C.B.; Thomas, D.M. A review of machine learning in obesity. *Obes. Rev.* **2018**, *19*, 668–685. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Bishop, C.M.; Nasrabadi, N.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
54. Shekar, B.H.; Dagnew, G. Grid search-based hyperparameter tuning and classification of microarray cancer data. In Proceedings of the 2nd International Conference on Advanced Computational and Communication Paradigms (ICACCP); IEEE: Piscataway, NJ, USA, 2019; pp. 1–8. [\[CrossRef\]](#)
55. Matuszelański, K.; Kopczewska, K. Customer churn in retail e-commerce business: Spatial and machine learning approach. *J. Theor. Appl. Electron. Commer. Res.* **2022**, *17*, 165–198. [\[CrossRef\]](#)
56. Mahoto, N.A.; Iftikhar, R.; Shaikh, A.; Asiri, Y.; Alghamdi, A.; Rajab, K. An intelligent business model for product price prediction using machine learning approach. *Intell. Autom. Soft Comput.* **2021**, *30*, 147–159. [\[CrossRef\]](#)

57. Bakhtyar, S.; Henesey, L. Freight transport prediction using electronic waybills and machine learning. In *Proceedings of the International Conference on Informatics and Cybernetics for Computational Social Systems (ICCSS)*; IEEE: Piscataway, NJ, USA, 2014; pp. 128–133. [\[CrossRef\]](#)
58. Yaxu, Y. Comprehensive evaluation of logistics enterprise competitiveness based on SEM model. *J. Intell. Fuzzy Syst.* **2021**, *40*, 6469–6479. [\[CrossRef\]](#)
59. Tian, Z.; Zhong, R.Y.; Vatankhah Barenji, A.; Wang, Y.T.; Li, Z.; Rong, Y. A blockchain-based evaluation approach for customer delivery satisfaction in sustainable urban logistics. *Int. J. Prod. Res.* **2021**, *59*, 2229–2249. [\[CrossRef\]](#)
60. Tamayo, S.; Combes, F.; Gaudron, A. Unsupervised machine learning to analyze city logistics through Twitter. *Transp. Res. Procedia* **2020**, *46*, 220–228. [\[CrossRef\]](#)
61. Pavlović, M.; Bojičić, R.R.; Ratković, M.C. Customer satisfaction with postal services in Serbia. *Management* **2018**, *23*, 15–33. [\[CrossRef\]](#)
62. Masudin, I.; Ramadhani, A.; Restuputri, D.P.; Amrina, E. Modified Kansei engineering for the quality of logistics services during the COVID-19 pandemic: Evidence from Indonesia. *Cogent Eng.* **2022**, *9*, 2050404. [\[CrossRef\]](#)
63. Gutierrez-Franco, E.; Mejia-Argueta, C.; Rabelo, L. Data-driven methodology to support long-lasting logistics and decision making for urban last-mile operations. *Sustainability* **2021**, *13*, 6230. [\[CrossRef\]](#)
64. De Araujo, A.C.; Etemad, A. End-to-end prediction of parcel delivery time with deep learning for smart-city applications. *IEEE Internet Things J.* **2021**, *8*, 17043–17056. [\[CrossRef\]](#)
65. Kosovac, A.; Muharemović, E.; Begović, M.; Šimić, E. Determining the location of postal centers in B&H using machine learning clustering method and GIS. In *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*; IEEE: Piscataway, NJ, USA, 2020; pp. 1318–1322. [\[CrossRef\]](#)
66. Ferreira, J.C.; Esperança, M. Enhancing sustainable last-mile delivery: The impact of electric vehicles and AI optimization on urban logistics. *World Electr. Veh. J.* **2025**, *16*, 242. [\[CrossRef\]](#)
67. Tadić, S.; Krstić, M.; Radovanović, L. Assessing strategies to overcome barriers for drone usage in last-mile logistics: A novel hybrid fuzzy MCDM model. *Mathematics* **2024**, *12*, 367. [\[CrossRef\]](#)
68. Albadrani, A.; Alghayadh, F.; Zohdy, M.A.; Aloufi, E.; Olawoyin, R. Performance and predicting of inbound logistics processes using machine learning. In *Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, 27–30 January 2021; pp. 790–795. [\[CrossRef\]](#)
69. El Ouadi, J.; Errouso, H.; Benhadou, S.; Medromi, H.; Malhene, N. A machine-learning based approach for zoning urban area in consolidation schemes context. In *Proceedings of the 2020 IEEE 13th International Colloquium of Logistics and Supply Chain Management (LOGISTQUA)*, Fez, Morocco, 2–4 December 2020; pp. 1–7. [\[CrossRef\]](#)
70. Zhao, S.; Sheng, Y.; Dong, Y.; Chang, E.I.; Xu, Y. Maskflownet: Asymmetric feature matching with learnable occlusion mask. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 13–19 June 2020; pp. 6278–6287. [\[CrossRef\]](#)
71. Savic, M.; Lukic, M.; Danilovic, D.; Bodroski, Z.; Bajovic, D.; Mezei, I.; Vukobratovic, D.; Skrbic, S.; Jakovetic, D. Deep learning anomaly detection for cellular IoT with applications in smart logistics. *IEEE Access* **2021**, *9*, 59406–59419. [\[CrossRef\]](#)
72. Gupta, A.; Singh, R.K.; Mathiyazhagan, K.; Suri, P.K.; Dwivedi, Y.K. Exploring relationships between service quality dimensions and customers satisfaction: Empirical study in context to Indian logistics service providers. *Int. J. Logist. Manag.* **2023**, *34*, 1858–1889.
73. Lin, X.; Mamun, A.A.; Yang, Q.; Masukujjaman, M. Examining the effect of logistics service quality on customer satisfaction and re-use intention. *PLoS ONE* **2023**, *18*, e0286382. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Tee, Y.-Y.; Ting, M.-S.; Alexander, J.A.; Tan, K.-Y.; Ooi, T.-K. Customer satisfaction toward logistic service quality: A comprehensive analysis of key factors and strategic implications. *Int. J. Acad. Res. Econ. Manag. Sci.* **2025**, *14*, 461–478. [\[CrossRef\]](#)
75. Aziz, Z.A.; Hussin, N.S.N.; Mukhtar, D.; Nordin, N.; Nordin, N. Factors influencing the service quality for customer satisfaction on parcel service safety in the postal and courier industry: Analysis using structural equation modelling. *Pertanika J. Soc. Sci. Humanit.* **2025**, *33*, 587–617. [\[CrossRef\]](#)
76. Xie, S.; Sharma, S.; Mehra, A.; Aziz, A. Strategic expectation setting of delivery time on marketplaces. *Inf. Syst. Res.* **2024**, *35*, 1965–1980. [\[CrossRef\]](#)
77. Voigt, S.; Kuhn, H. Crowdsourced logistics: The pickup and delivery problem with transshipments and occasional drivers. *Networks* **2022**, *79*, 403–426.
78. Filani, O.M.; Olajide, J.O.; Osho, G.O. A machine learning-driven approach to reducing product delivery failures in urban transport systems. *Gyanshauryam Int. Sci. Ref. Res. J.* **2023**, *6*, 119.
79. Liu, H.; Hao, P.; Tanvir, S.; Pande, A.; Barth, M. Vehicle miles traveled and environmental impacts from on-demand delivery: A literature review. In *Proceedings of the International Conference on Transportation and Development 2024*, Atlanta, GA, USA, 10–13 June 2024; pp. 37–48.

80. Zosu, S.; Amaghionyeodiwe, C.; Oyetunji, E.; Yussouf, A. Last-mile delivery optimization: Balancing cost efficiency and environmental sustainability. *Int. J. Emerg. Trends Eng. Res.* **2024**, *12*, 153–166. [\[CrossRef\]](#)
81. Ojha, R.; Agarwal, A. Examining environmental sustainability in last-mile delivery: A quality function deployment analysis. *TQM J.* **2025**, *38*, 1573–1599. [\[CrossRef\]](#)
82. McKinnon, A. Environmentally sustainable city logistics: Minimising urban freight emissions. In *Handbook on City Logistics and Urban Freight*; Edward Elgar Publishing: Cheltenham, UK, 2023; pp. 463–482.
83. Kavitha, S.; Varuna, S.; Ramya, R. A comparative analysis on linear regression and support vector regression. In *Proceedings of the Online International Conference on Green Engineering and Technologies (IC-GET)*; IEEE: Piscataway, NJ, USA, 2016; pp. 1–5. [\[CrossRef\]](#)
84. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
85. Natekin, A.; Knoll, A. Gradient boosting machines: A tutorial. *Front. Neurobot.* **2013**, *7*, 21. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Kosovac, A.; Medić, A.; Kurešepi, A.; Neimarlija, A. Modeling of the impact of flight delays on passenger satisfaction using logistic regression. In *Proceedings of the International Conference “New Technologies, Development and Applications”*; Springer: Cham, Switzerland, 2025; pp. 253–265. [\[CrossRef\]](#)
87. Kosovac, A.; Medić, A.; Omerašević, N.; Keško, A. Statistical modeling of delivery delays due to weather and traffic factors. In *Proceedings of the 24th International Symposium INFOTEH-JAHORINA (INFOTEH)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6. [\[CrossRef\]](#)
88. Cherkassky, V.; Mulier, F.M. *Learning from Data: Concepts, Theory, and Methods*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
89. Brown-Trinh, R. Effects of winsorization: The cases of forecasting non-GAAP and GAAP earnings. *J. Bus. Financ. Account.* **2018**, *46*, 105–135. [\[CrossRef\]](#)
90. Caiafa, C.F.; Sun, Z.; Tanaka, T.; Marti-Puig, P. Machine learning methods with noisy, incomplete or small datasets. *Appl. Sci.* **2021**, *11*, 4. [\[CrossRef\]](#)
91. Thabtah, F.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data imbalance in classification: Experimental evaluation. *Inf. Sci.* **2020**, *513*, 429–441. [\[CrossRef\]](#)
92. Provost, F. Machine learning from imbalanced data sets 101. In *Proceedings of the AAAI’2000 Workshop on Imbalanced Data Sets*; AAAI Press: Menlo Park, CA, USA, 2000; pp. 1–3.
93. Hanskunatai, A. A new hybrid sampling approach for classification of imbalanced datasets. In *Proceedings of the 3rd International Conference on Computer Communication Systems (ICCCS)*; IEEE: Piscataway, NJ, USA, 2018; pp. 278–281. [\[CrossRef\]](#)
94. Shah, S.; Gandhi, D.; Kothari, J. Machine learning based synthetic data generation using iterative regression analysis. In *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*; IEEE: Piscataway, NJ, USA, 2020; pp. 1093–1100. [\[CrossRef\]](#)
95. Tianfeng, C.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [\[CrossRef\]](#)
96. Khan, M.A.; Khan, R.; Algarni, F.; Kumar, I.; Choudhary, A.; Srivastava, A. Performance evaluation of regression models for COVID-19: A statistical and predictive perspective. *Ain Shams Eng. J.* **2022**, *13*, 101574. [\[CrossRef\]](#)
97. Frost, J. *Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models*; Statistics by Jim Publishing: State College, PA, USA, 2019.
98. Stumpfe, S.F.; Shongwe, S.C. Comparative Analysis and Optimisation of Machine Learning Models for Regression and Classification on Structured Tabular Datasets. *Mathematics* **2026**, *14*, 473. [\[CrossRef\]](#)
99. Dobrodolac, M.; Lazarević, D.; Jovčić, S. Postal Operators as Providers of Shared Mobility Concept—An Analysis by Combined SWOT and AHP. *Decis. Mak. Adv.* **2024**, *2*, 213–221. [\[CrossRef\]](#)
100. Medić, A.; Kosovac, A.; Muharemović, E.; Begović, M. Machine learning application for improving customer and postal logistics operator satisfaction in urban areas—A review. *Promet Traffic Transp.* **2025**, *37*, 287–300. [\[CrossRef\]](#)
101. Li, D.; Wan, G.; Rong, Y. An Enhanced Spherical Cubic Fuzzy WASPAS Method and Its Application for the Assessment of Service Quality of Crowdsourcing Logistics Platform. *Spectr. Decis. Mak. Appl.* **2026**, *3*, 100–123. [\[CrossRef\]](#)
102. Badi, I.; Demir, G.; Bouraima, M.B.; Kimtai, M.A. Vendor Managed Inventory in Practice: Efficient Scheduling and Delivery Optimization. *Spectr. Decis. Mak. Appl.* **2025**, *2*, 157–165. [\[CrossRef\]](#)
103. Dobrodolac, M.; Bošković, S.; Jovčić, S.; Lazarević, D. Sustainable Delivery Model Selection using AROMAN Approach. *Decis. Mak. Adv.* **2024**, *2*, 73–82. [\[CrossRef\]](#)
104. Tahir, M.; Shahid, M.I. An Intelligent Industry 5.0 Logistics Decision System: Circular Supply Chain Management with Fermatean Neutrosophic Hypersoft Sets And Machine Learning. *J. Contemp. Decis. Sci.* **2026**, *2*, 260–286. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.