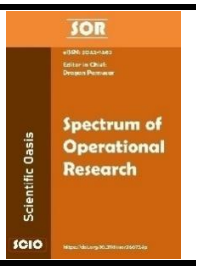
SCIENTIFIC OASIS

Spectrum of Operational Research

Journal homepage: www.sor-journal.org
ISSN: 3042-1470



Forecasting Car Repair Shops Customers' Loyalty based on SERVQUAL Model: An Application of Machine Learning Techniques

Farshid Abdi[1], Shaghayegh Abolmakarem[1], Amir Karbassi Yazdi[2,*]

[1] Department of Industrial Engineering South-Tehran Branch, Islamic Azad University, Tehran, Iran
[2] Departamento de Ingeniería Industrial y de Sistemas, Facultad de Ingeniería, Universidad de Tarapacá, Arica 10000, Chile

ARTICLE INFO

Article history:
Received 15 August 2024
Received in revised form 20 January 2025
Accepted 21 January 2025
Available online 24 January 2025

Keywords:
SERVQUAL Model, Service Quality, Customer Satisfaction, Machine Learning, Customer Behavior Prediction.

ABSTRACT

In today's competitive world, service quality and customer satisfaction are recognized as key factors in the success of service organizations. This paper examines the impact of the SERVQUAL model and machine learning techniques on these two factors in the automotive repair and maintenance industry. With increasing customer awareness of different options and increasing competition in the market, repair shops must continuously improve their service quality to attract and retain customers. Satisfied customers are more likely to return to repair shops and share their positive experiences with others, which helps attract new customers. The research aims to predict customers' willingness to return to a repair shop based on existing features and service quality indicators according to the SERVQUAL model. The results show that using machine learning methods to analyze data can more effectively identify complex patterns and predict customer behaviour. This paper also discusses the limitations of structural equation models (SEM) in predicting future customer behaviour and emphasizes that machine learning methods can provide more accurate predictions. Finally, this research emphasizes the importance of paying attention to service quality and creating a positive customer experience so that repair shops can succeed in a competitive market.

1. Introduction

With increasing competition in today's markets, establishing and maintaining long-term relationships with customers is essential for organizations' survival and profitability. As a result, organizations must meet customer needs to achieve this goal by increasing their satisfaction. Therefore, efforts to improve the quality of services and pay attention to solutions that increase customer loyalty should be a priority in organizations' plans [1].

In addition, customers are becoming more aware of competitors, products, and services every day and have different options [2]. Under these circumstances, it can be safely claimed that today, no business except monopolistic organizations can survive without having loyal customers. Given

[*] *Corresponding author.*
E-mail address: akarbassiy@academicos.uta.cl

https://doi.org/10.31181/sor21202517

that customer expectations are constantly increasing, organizations must go beyond their basic needs, meet customer expectations, and shift their focus from mere customer satisfaction to building loyalty and trust by creating long-term, mutual, and profitable relationships for both parties [3].

In today's world, competition requires that organizational managers focus on increasing customer satisfaction, reducing production costs, and providing high-quality services. Therefore, examining and understanding customer satisfaction indicators and measuring customer satisfaction is important because, ultimately, the level of customer satisfaction determines the success or failure of organizations. Satisfied customers are the source of profit for companies. Companies that cannot satisfy customers will not remain in the market long term [4]. The automotive repair and maintenance service industry is one of the most competitive industries, and this competition exists not only between different car brands but also between repair shops and service centers. In this competitive environment, repair shops must continuously improve the quality of their services and attract customers by providing better services and competitive prices. The entry of new competitors and the threat of substitute products and services puts more pressure on repair shops to maintain their market share. Customer satisfaction plays a vital role in this industry. Satisfied customers are considered a source of revenue and profit for repair shops [5]. When customers are happy with the services of a repair shop, the likelihood of them returning to use the services again increases significantly. This return not only means increased revenue for the repair shop but can also create a loyal customer base that helps grow the business over time. Also, satisfied customers share their positive experiences with others, which helps attract new customers. As a result, repair shops must focus on creating a positive experience for customers to succeed in this competitive market. Providing high-quality service, paying attention to the needs and wants of customers, and establishing effective communication with them are among the measures that can increase customer satisfaction and ultimately lead to greater profitability for repair shops.

On the other hand, in today’s world, collecting customer data has become a common trend in most industries. This data includes purchase history, online interactions, and customer reviews, which can help analyze customer behavior. Using advanced data analysis methods, businesses can predict their customers' future behavior. These predictions can include customers’ willingness to return and use repair shop services again, which is especially important in the automotive industry.

The main objective of this research in this section is to predict the customers’ willingness to revisit the repair shop based on their available features and some service quality indicators according to the SERVQUAL model. The SERVQUAL model examines service quality and its impact on customer satisfaction and can help identify the strengths and weaknesses of repair shops. Modelling the prediction of customers’ willingness to revisit the repair shop has been done using machine learning algorithms. Finally, if-then rules have been extracted to determine the level of customer loyalty to the repair shop.

In previous research, SEM have been commonly used to measure customer satisfaction and analyze relationships between variables. These methods are very popular in social sciences and marketing due to their capabilities in modelling the complexities of relationships between variables and assessing direct and indirect effects. However, these approaches may have limitations in predicting future customer behavior. Instead of relying on structural equation models, this paper examines machine learning methods. Machine learning methods can predict future customer behavior more remarkably by processing big data and identifying complex patterns. This method automatically learns from data and does not require prior assumptions.

The rest of the paper is organized as follows. In Section 2, a literature review on the subject of data-mining techniques in the insurance domain is presented. In Section 3, the theoretical

background is presented. In Section 4, the proposed research methodology of the paper is explained. In Section 5, experimental results are represented. In Section 6, conclusions are discussed

2. Literature review

In one research study, Le et al. [6] examined how machine learning and deep learning algorithms can predict customer satisfaction in the e-commerce industry. This study used customer purchase behaviour, reviews, and rating data to build a comprehensive model for analyzing and predicting customer satisfaction. The model includes several different algorithms that help identify complex patterns in the data and can help companies make optimal decisions to improve customer experience. This study identified key variables such as service quality, pricing, user experience, and customer interactions with online platforms. The researchers analyzed the data using deep learning techniques, such as convolutional neural networks (CNN) and recurrent neural networks (RNN). They were able to increase the accuracy of predicting customer satisfaction. The results of this study show that using predictive models based on machine learning and deep learning can help companies better understand customer needs and wants and improve their services. In their article, Zaghloul et al. [6] conducted a comparative study between machine learning and deep learning in predicting customer satisfaction in the e-commerce industry. Using customer feedback data, the researchers developed different models to evaluate the impact of each method on the accuracy of predicting customer satisfaction. The results show that both approaches have their capabilities. Still, deep learning usually provides higher accuracy in predicting customer satisfaction due to its ability to process complex data and identify nonlinear patterns. This research identified key variables such as service quality, user experience, and pricing affecting customer satisfaction. The researchers also discussed the challenges and benefits of both approaches and emphasized that combining these two methods can help improve the accuracy of predictions. The article concludes that using advanced machine learning and deep learning techniques is necessary to optimize customer experience and increase satisfaction. The authors used machine learning techniques, including decision trees, support vector machines (SVMs), and regression algorithms, and compared the results with deep learning models, such as neural networks. The findings show that deep learning models perform significantly better than traditional models and can predict customer satisfaction more accurately. The paper concludes that combining these two approaches can help improve the accuracy of predictions and provide valuable insights for optimizing the customer experience.. Ong et al. [7] investigated the factors influencing customer satisfaction at the National Electric Company. This study used a combination of SERVQUAL and Expectation-Confirmation Theory approaches. The researchers analyzed key variables by collecting data from 529 participants who responded to a questionnaire consisting of 49 questions. SEM was used to examine the relationships between variables, which showed that Tangibility, Empathy, and Responsiveness were positively related to service quality and that service quality led to customer expectations, energy consumption, and perceived performance (PE). The results of this study indicate that higher perceived performance leads to expectation confirmation, which ultimately leads to customer satisfaction. This research showed that integrating SERVQUAL and expectation confirmation theory can comprehensively measure customer satisfaction among electricity service providers. Overall, the paper concludes that paying attention to service quality factors and their impact on customer expectations and confirmation is crucial to enhancing customer satisfaction during similar crises. Ong et al. investigated the factors that affect customer satisfaction with water services provided by Maynilad Water Company. In this study, the researchers used an online questionnaire to collect data from 725 Maynilad customers and analyzed the data using snowball sampling. This study integrated the SERVQUAL dimensions and Expectation Confirmation Theory (ECT) to identify ten latent variables: Assurance, Tangibles, Empathy, Expectations,

Confirmation, Performance, and Water Consumption. SEM and Deep Learning Neural Networks (DLNN) were used to analyze the data. The results of this study indicate that factors such as providing water services at reasonable prices, issuing accurate water invoices, completing repairs and installations on time, and having professional staff have a positive effect on customer satisfaction. This study is the first research evaluating customer satisfaction with Maynilad services in the Philippines, and its results can help the company's managers improve their service quality and design effective policies to increase customer satisfaction. Also, this research can be a good reference for other service companies in improving service quality Ruiz et al [8] examine the quality of services public transport systems provide in developing countries. This study aims to assess and predict passenger satisfaction by collecting data from passengers and using machine learning techniques to identify factors affecting customer satisfaction. Researchers used machine learning methods, including decision trees, support vector machines (SVMs), and neural networks, to model and predict passenger satisfaction. Data analysis was also performed using statistical techniques and multivariate analysis to identify relationships between variables. The results of this study show that key factors such as service quality, pricing, service accessibility, and user experience significantly impact passenger satisfaction. Also, the findings indicate that using machine learning models can help improve the accuracy of predictions and provide valuable information for public transport system managers to develop better strategies to increase customer satisfaction. This research emphasizes the importance of paying attention to passenger needs and improving service quality to succeed in the competitive environment. Zibarzani et al. [9]investigated the effects of the coronavirus outbreak on customer satisfaction with restaurant services. In this study, researchers used a hybrid approach, including machine learning and text analytics, to examine online customer reviews. In the first stage, customer review data was collected from social networks and analyzed using clustering and supervised learning techniques. Then, thematic analysis (LDA) was used to identify customer satisfaction measures. In the second stage, the relationships between restaurant service quality factors and customer satisfaction were examined using the partial linear regression (PLS) modelling technique. The results of this study show that compliance with safety protocols during the COVID-19 era significantly impacts service quality and customer satisfaction. The findings also indicate that customers pay special attention to food quality, staff service, and hygiene. This study emphasizes the importance of service innovation and continuous quality improvement so that restaurants can succeed in critical situations such as the COVID-19 pandemic. The results of this study can help restaurant managers design effective strategies to increase customer satisfaction and maintain their loyalty. Akhavan et al. investigate machine learning and process analytics techniques to improve customer experience in online insurance services. In this study, researchers used a hybrid approach, including machine learning models and process data analytics, to identify and predict factors affecting customer satisfaction. The data was collected from various sources and included information about customer interactions with online platforms, customer reviews, and variables related to service quality. Using machine learning algorithms, such as decision trees and neural networks, the researchers analyzed the data and evaluated the prediction accuracy. The results of this study show that combining machine learning with process analytics can significantly increase the accuracy of customer experience prediction. Also, identifying key factors such as service quality, response time, and accessibility of online services positively impacts customer satisfaction. This study recommends insurance industry managers use advanced analytical approaches to improve customer experience and design effective strategies to attract and retain customers. Investigates how to predict passenger satisfaction with airline services using sentiment analysis and machine learning models. In this study, researchers used the VADER (Valence Aware Dictionary and Sentiment Reasoner) model to analyze passengers' text comments to identify positive and negative sentiments.

Data related to airline service ratings were also collected. Then, various machine learning algorithms, including decision trees, logistic regression, and random forests, were used to predict passenger satisfaction levels. The results of this study show that sentiment analysis can serve as an effective tool in predicting customer satisfaction, and combining it with machine learning techniques increases the accuracy of predictions. Also, the key factors identified include service quality, seat comfort, and staff behavior, which significantly impact passenger satisfaction. The study recommends that airlines use these approaches to improve their services and enhance customer experience to succeed in the competitive market. Wang et al. [10]examine how the performance of various features affects customer satisfaction and use machine learning techniques to analyze this relationship. The researchers identified and evaluated customer satisfaction factors using data collected from surveys and statistical analyses. This research emphasizes the importance of the interpretability of machine learning models so that the results can be easily explained to managers and decision-makers. In this study, the researchers identified key variables such as service quality, pricing, customer experience, and product features and examined their relationships and overall customer satisfaction. The results show that the performance of features directly affects customer satisfaction, and this effect may change over time. The paper concludes that using interpretable machine learning approaches can help companies better understand customer needs and expectations and design more effective strategies to improve customer satisfaction. Rezendeh et al. [11] investigate the factors affecting customer satisfaction in the food industry and use SEM techniques to analyze the data. By collecting customer data, the researchers identified and evaluated key variables such as food quality, price, customer service, and physical environment. This research seeks to understand the relationships between these variables and their impact on overall customer satisfaction. The results of this study show that food quality and customer service are recognized as the main factors affecting customer satisfaction. Also, fair prices and a positive physical environment are essential in creating a positive customer experience. The paper concludes that to increase customer satisfaction in the food industry, paying attention to these factors and improving the quality of services and products is essential. These findings can help food industry managers design effective strategies to attract and retain customers. Muni et al. examined the impact of service quality on customer satisfaction in the rail freight industry. The researchers used SEM to analyze survey data and identified key variables such as transit time, wagon availability, and reliability. The study sought to understand better customer expectations and how these factors affect their satisfaction. The results of this study indicate that transit time and wagon availability are the two main factors influencing customer satisfaction. The findings also suggest that improvements in service quality, especially in transit time-related dimensions, can lead to increased customer satisfaction and their willingness to use rail freight services again. The paper concludes that paying attention to service quality and addressing weaknesses in this area is crucial for attracting and retaining customers in the competitive rail freight market. Munde et al. [12] investigate the factors influencing consumers' purchasing decisions regarding sustainable jewelry. The researchers used machine-learning techniques to analyze data and predict consumer purchasing behavior. The data used included surveys and information on consumer preferences regarding sustainable jewelry. The main objective of this study was to identify purchasing patterns and factors that influence consumer satisfaction with sustainable jewelry. The results of this study show that factors such as material quality, price, transparency in sourcing, and commitment to ethical production practices significantly impact consumer purchasing decisions. The findings also indicate that younger generations, especially Millennials and Gen Z, seek sustainable and ethical options and are more inclined to purchase jewelry with transparent sourcing. The paper concludes that machine learning algorithms can help brands better understand consumer needs and preferences and design effective strategies to attract them. Sun et al. [13] review advanced methods

and techniques for predicting customer lifetime value (CLV). This research analyzes historical customer data and uses machine learning algorithms to predict their future behavior. The researchers used machine learning models like XGBoost and clustering models to identify purchasing patterns and determine customer lifetime value. The paper also highlights the importance of customer relationship management (CRM) analytics in improving marketing strategies and resource allocation. This research shows that using machine learning algorithms can help companies more accurately predict their customer lifetime value and make better marketing and customer service decisions based on it. The paper also emphasizes the importance of quality data and its proper preparation for accurate results. Finally, the findings show that combining CRM analytics with machine learning can increase customer satisfaction and improve overall business performance. Tanwar et al. [14] investigate the factors that affect the quality of services provided in the public transport system of Bhopal. Using data collected from passengers, the researchers identified and analyzed key variables. This study used SEM to examine the relationships between variables and their impact on service quality. Also, factor analysis was used to identify different groups of variables and determine the main factors of service quality. The results of this study show that factors such as accessibility, waiting time, travel comfort, and staff behavior significantly affect service quality. Also, the findings indicate that improvements in these factors can lead to increased passenger satisfaction and greater use of the public transport system. The paper concludes that it is necessary to pay attention to these factors and improve service quality to meet the needs of passengers better and ultimately attract more users to the public transport system. GhorbanTanhaei et al. [15] examine the use of predictive analytics to understand customer behavior better and identify future trends in the market. The research emphasizes the importance of collecting and analyzing historical customer data, which can help determine their purchasing patterns and preferences. It is possible to predict customer behavior and identify their needs using machine learning algorithms, such as decision trees, clustering, and regression. This approach allows businesses to optimize their marketing strategies and improve the customer experience. The results show that predictive analytics can help identify customer churn risks, cross-sell and upsell opportunities, and predict customer lifetime value. Companies can also personalize their services and communicate more effectively with customers using data collected from online interactions and purchase history. This article concludes that using predictive analytics in today's competitive world helps to understand customer behavior better and leads to increased customer loyalty and revenue growth. Nilashi et al. [16] examine the factors influencing customers' purchasing decisions in smart home security systems. In this study, researchers used data collected from social network surveys and machine learning algorithms to identify patterns and trends in customer purchasing. The study analyzed customer sentiment and opinions about security systems and identified key factors such as brand trust, price, product quality, and user experience. The results of this study show that positive sentiments towards the brand and product quality significantly impact customers' purchase intention. Also, the findings indicate that appropriate pricing and providing transparent information about products can help increase customers' willingness to purchase smart home security systems. The paper concludes that social data analysis and machine learning techniques can help companies better understand customers' needs and preferences and design effective strategies to attract them. Akter et al. [17] used two different analysis methods, namely structural equation modelling (PLS-SEM) and fuzzy quality analysis (fsQCA), to analyze the data and examine the relationships between variables. The main objective of this study is to determine how system quality can act as a moderating factor in the relationship between customer satisfaction and purchase intention. The results of this study show that system quality has a positive and significant effect on customer satisfaction and acts as a moderating factor in the relationship between customer satisfaction and purchase intention. In other

words, when system quality is higher, the effect of customer satisfaction on purchase intention also increases. These findings indicate the importance of paying attention to system quality in service design and delivery so that it can help increase customer satisfaction and, consequently, increase purchase intention. The paper concludes that companies should focus on improving the quality of their systems to improve customer experience and increase their loyalty. Li et al. [1] examined how excessive or "extra service" affects customer satisfaction and their intention to return to a restaurant. The research suggests that while providing exemplary service is essential, providing excessive service can lead to negative experiences for customers, especially for those who value control over their dining experience. Key findings from the study suggest that customers who perceive excessive service may experience increased stress, which can negatively impact their overall satisfaction. This is particularly evident among customers who have a high desire for control, as they may feel uncomfortable with excessive attention from staff. The research used various methods to analyze the relationship between excessive service, customer satisfaction, and intention to return, concluding that a proper balance must be struck in service delivery to improve the customer experience. This research suggests that restaurant managers should pay attention to the level of service provided and ensure that it meets customer expectations and does not cross the line into excessive service, as this can lead to dissatisfaction and reduce the likelihood of customers returning.

3. Theoretical background

3.1. Data Preparation

Data preparation is a crucial step in the data analysis process, ensuring that the dataset is clean, relevant, and ready for further analysis. In this research this phase involves several key operations, including data cleaning, and feature subset selection. Proper data preparation enhances the accuracy of the results and improves the performance of machine learning models.

3.1.1. Data Cleaning

One of the common problems with data is its low quality. Data cleaning is the operation that leads to resolving data quality problems. The presence of missing values and duplicate data are among the problems that compromise data quality.

- *Missing values:* For some reason, some values related to some features may not be available. We call such values missing values. In this research, we used the missing data estimation method to manage missing data.

3.1.2. Feature Subset Selection (FSS)

- *Information Gain:* This FSS method assigns a rank to each feature describing a set of training records. This criterion is based on the work of Claude Shannon, a pioneer in information theory, who emphasized the value or information content of messages. To determine information gain, a criterion called entropy, which is commonly used in information theory, is used.

$$Entropy(d) = -\sum_{i=1}^{m} p_i log_2(p_i) \quad (1)$$

Where D is a set of training records, Pi represents the probability that a record in D belongs to class Ci, and m is the number of classes. In the above formula, Entropy(D) is simply the average amount of information needed to identify the class label of a record in D. If we want to partition or

partition the records in D based on a feature such as A that has v distinct values of {a1,a2,…,av}, we use the following formula.

$$info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times info(D_j) \tag{2}$$

The expression $\frac{|D_j|}{|D|}$ acts as the weight of the jth segment. Finally, the information gain measure is defined as the difference between the original information requirement (defined only based on the ratio of the categories) and the new requirement (obtained after segmentation based on feature A). So we have:

$$Gain(A) = Entropy(D) - info_A(D) \tag{3}$$

Based on this relationship, we can state that by knowing the value of A, we have obtained the amount of information necessary to classify a record. We are looking for a property that, after partitioning, minimizes the amount of information necessary to finish classifying records, that is, info A(D).

- *Gini Index:* This measure calculates the gross educational record D as follows:

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2 \tag{4}$$

where Pi represents the probability that a tuple in D belongs to class Ci. If a K-fold partition on feature A partitions the data set D into K sets D1, D2, ...Dk, then the Gini index of the data set D for such a partition is:

$$Gini_A(D) = \sum_{i=1}^{k} \frac{|D_i|}{|D|} Gini(D_i) \tag{5}$$

The reduction of impurity that is done by binary failure for the characteristic A is equal to Eq. (6).

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \tag{6}$$

The feature that maximizes this reduction in impurity (or in other words, has the minimum Gini index value) is selected as the failure feature in the given training record set.

3.2. Modeling and Classification

In this study, classification methods will be used for modelling and using them; it will be predicted whether customers will tend to visit the repair shop again or not.

- *Decision tree:* Decision trees generally follow a top-down approach and structure like a flowchart. Every internal node in Decision trees contains a question about one particular feature, and every branch shows the result of the experiment. Also, each leaf (terminal) is labelled with one of the possible classes [18]. The highest node in a tree is the node of the root. Some of the decision tree algorithms produce binary trees. There is no need for specific knowledge or setting of any parameter to produce decision trees. Therefore, for such a purpose, heuristic knowledge is appropriate. During the decision tree's production, attribute selection measurements are used. The attribute that is selected by using such measurements divides the records into separate groups in the best way [19].
- **ID3** This algorithm is appropriate for identifying and producing the Classification Tree with multiple splitting in each node and this algorithm was designed for qualitative variables, but

it can be used for a set of variables, whether qualitative or numerical [20].

3.3. Rule Extraction by Decision Tree

The result of the implementation of the decision tree algorithm is a set of logical conditions in the form of IF-THEN rules. The tree structure is used to predict a feature so that the data at the end of tree leaves are labelled by one of the target feature amounts. Due to the ease of interpretation of the results and being non-parametric and non-linear, this model does not need a presupposition of the linear relationship between independent and dependent variables [21].

3.4. Validation and Evaluation of Results

In this research, the batch validation method (10 fold) was used to validate the results. We examined modelling and used evaluation indices such as accuracy, precision, and recall to evaluate the results using model evaluation indices. These indices help determine which classification method has greater accuracy and precision than other methods. These indices are calculated according to Equations (7), (8), and (9). Table 1 represents the confusion matrix, which summarizes the results of the classification model by showing the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These values are essential for calculating evaluation metrics such as accuracy, precision, and recall, which help assess the model's effectiveness in predicting customer loyalty.

Table1
Confusion Matrix

		Prediction	
		Positive	*Negative*
Actual	*Positive*	True Negative (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

4. Proposed model

In this section, the proposed model of the research has been described. As shown in Figure 1, the proposed model comprises two stages: (1) data pre-processing and Feature selection and (2) classification and Rule Extraction.

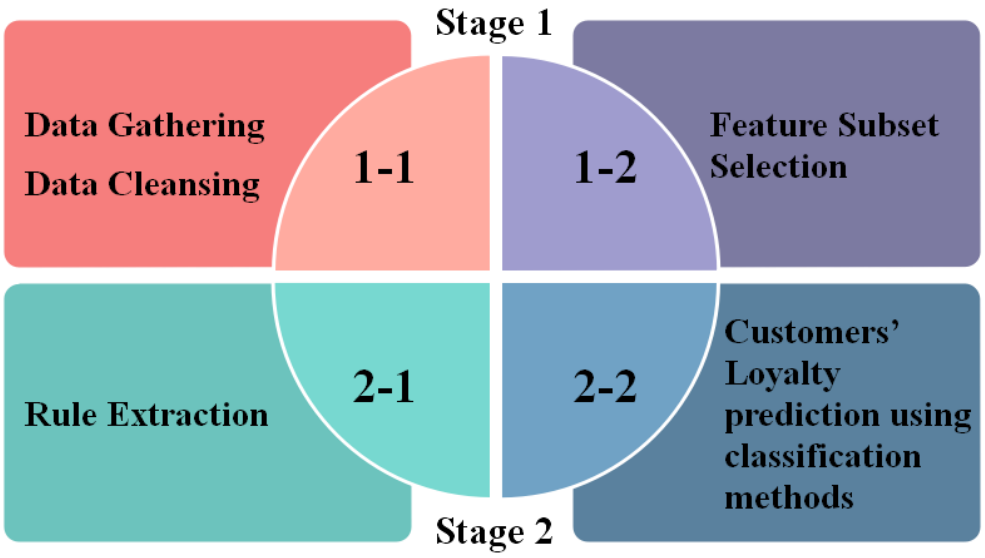


Fig. 1. Experimental Results Process

Stage 1: Data Pre-processing and Feature Selection

In the first stage, various preprocessing methods are employed to prepare the data for analysis. This includes data cleansing, which involves removing any inaccuracies or inconsistencies in the dataset to ensure that the data is reliable for further processing. Additionally, feature selection is a critical step where algorithms are utilized to identify and select the most relevant features that contribute to the predictive modeling process. Among the filtering feature selection techniques applied are Information Gain and Gini Index. These algorithms help in evaluating the significance of each feature in relation to the target variable, thereby enhancing the dataset by retaining only those features that provide substantial information for classification tasks. Once the relevant features are identified, they are input into various machine learning classification algorithms. The algorithms chosen for this study include decision trees, ID3, and Random Forest. These classifiers are designed to learn from the training data and make predictions on unseen data based on the patterns identified during training.

Stage 2: Classification and Rule Extraction

The second stage focuses on applying the selected machine learning algorithms to classify the data based on the features identified in the first stage. The performance of these classification models is evaluated using several metrics, including classification accuracy, precision, and recall. These metrics provide insights into how well each model performs in predicting customer loyalty based on service quality indicators derived from the SERVQUAL model. Furthermore, this stage also involves rule extraction, where if-then rules are derived from the trained models. These rules help in understanding the decision-making process of the classifiers and provide actionable insights into customer behavior patterns. By analyzing these rules, repair shops can better comprehend what factors influence customer loyalty, allowing them to tailor their services accordingly.

4.1. Dataset description

The data used in this section is related to information collected from repair shop customers through a questionnaire. Table 2 presents the variables used in this section of the research. In this research, the variables was collected through questionnaires administered to customers of automotive repair shops, focusing on their experiences and perceptions of service quality based on the SERVQUAL model dimensions: tangibility, reliability, responsiveness, assurance, and empathy. The collection of demographic and customer-related features in this research aims to gain a better

understanding of customer behavior and needs in automotive repair shops. These features include gender, education level, age, number of visits to the repair shop per year, main reason for visit, time distance to the repair shop, and reason for referral. To gather this information, a structured questionnaire was designed, incorporating closed and multiple-choice questions to facilitate quick and easy responses from participants. The target group for the survey consisted of customers of automotive repair shops, ensuring a diverse range of ages and educational backgrounds to enhance data variability. Demographic questions regarding gender (male, female), education level (undergraduate, diploma, postgraduate, etc.), and age (less than 20 years old, between 20 and 30 years old, etc.) help researchers identify different behavioral patterns among customers. Additionally, questions about the number of visits to the repair shop per year (less than 5 times, between 5 and 10 times, etc.) and the main reason for visit (check-ups, periodic services, etc.) provide valuable insights into customer habits. Inquiring about the time distance to the repair shop (less than 30 minutes, between 30 minutes to 1 hour, etc.) can indicate the geographical impact on customer choices. Lastly, exploring reasons for referral (proximity to work or residence, recommendations from friends and acquaintances, etc.) can assist businesses in recognizing their strengths and areas for improvement.

Table 2
The variable of the research

Category	Variable	Description	Symbol
Demographic and Customer-related features	Gender	Male, Female	V1
	Education level	Undergraduate, Diploma, Postgraduate, Bachelor's, Master's, PhD	V2
	Age	Less than 20 years old, between 20 and 30 years old, between 30 and 40 years old, between 40 and 50 years old, over 50 years old	V3
	Number of visits to the repair shop per year	Less than 5 times, between 5 and 10 times, more than 10 times	V4
	Main reason for visit	Check-ups, periodic services, electricity and wiring, suspension system, and other	V5
	Time distance to repair shop	Less than 30 minutes, 30 minutes to 1 hour, 1 to 2 hours, more than 2 hours	V6
	Reason for referral	Proximity to work or residence, recommendations from friends and acquaintances, reasonable price of services, quality of services, random visits, name and reputation of the repair shop	V7
13 Features extracted from the SERVQUAL questionnaire	Tangible	Use and benefit from modern and up-to-date repair shop equipment and tools	T1
		The condition of the repair shop is desirable in terms of appearance (including beauty and interior design, cleanliness of the environment, etc.) and amenities for customers (including suitable parking space	T2

Category	Variable	Description	Symbol
		for the car until the moment of acceptance, customer waiting room, suitable refrigeration systems, reception, etc.).	
		Communication with the customer during the provision of services (including administrative procedures, record keeping, method of accepting and delivering the car, and coordination) is regular and desirable.	T3
		The appearance and clothing of the employees are neat and clean.	T4
	Reliability	Promised services are performed on time.	Rel5
		Supervisors and repair shop employees appear professional in their work.	Rel6
		Troubleshooting is done accurately and flawlessly, and in accordance with expectations.	Rel7
		Documents related to services provided, including service descriptions and billing statements, are provided on time and without errors.	Rel8
	Responsiveness	Employees are informed clearly and accurately about old and new services to customers.	Res9
		Employees and supervisors are responsive and available to handle customer questions, criticisms, and complaints appropriately.	Res10
		Employees are held accountable for the services provided in an appropriate manner.	Res11
		Explanations about vehicle defects and guidance after repair are provided in an appropriate manner.	Res12
	Assurance	Reliability of the staff and the complex in providing services	A13
		Making the customer feel secure in interacting with the complex and staff	A14
		Polite and respectful behavior of the staff towards the customer	A15
		Sufficient knowledge and expertise of the staff to answer questions	A16
	Empathy	Showing special and appropriate attention to customers from the staff	E17
		Appropriate and flexible working hours of the complex to provide services	E18
		Understanding the specific needs of the	E19

Category	Variable	Description	Symbol
		customer from the staff	
		Requesting and providing the best suggestions in line with the interests of the customer from the complex and staff	E20
Target		Customer's willingness to return: 1 (low) 2 (medium) 3 (high)	C

As can be seen in Table 2, the variables used in this part of the study have been divided into two categories: customer-related characteristics and characteristics extracted from the SERVQUAL questionnaire. The characteristics available from customers are of the nominal discrete type, and the different categories of these variables can be seen in the table. The variables of the second category, namely the characteristics extracted from the SERVQUAL questionnaire, are also of the nominal discrete type. These variables were collected based on the Likert scale. They indicated the customer's level of desirability of the services provided at the repair shop, which ranges from 1, which shows very low satisfaction, to 5, which indicates very high customer satisfaction. Variable C is the target variable with three levels: low, medium, and high, and suggests the customer's willingness to return for future visits.

5. Experimental results

All algorithms of this research have been implemented using Rapid Miner software.

5.1. First stage: data pre-processing and Feature selection

In the first stage, the data were cleaned to detect and remove outliers using the distance-based method. Missing values were also handled by the imputation method.

- *Feature selection*

This section presents feature selection results using filter methods such as information gain and the Gini Index. These FS techniques provide a weight for each feature. The filter approach assigns weights to each feature independent of the classification algorithm [22]. To calculate the weights by Information Gain, Eq. (1) to (3) have been used and Eq. (4) to (5) have been used to calculate weights based on the Gini Index.

Table 3
Weighting variables using filtering methods

Variable	Weight by Information Gain	Weight by Gini Index
Gender	0	0
Number of visits to the repair shop per year	0.095838847	0.05922289
Education level	0.075039886	0.096876917
Time distance to repair shop	0.201215026	0.135101285
Main reason for visit	0.160155948	0.164339192
Age	0.178472928	0.221350517
T4	0.255100859	0.255992582
T2	0.291838715	0.328564694
Rel 6	0.328466165	0.334664701
Res10	0.303200868	0.347900798
Res9	0.411449339	0.36679799
A16	0.465550878	0.41663216
A15	0.292019262	0.425839041

Variable	Weight by Information Gain	Weight by Gini Index
A14	0.347861162	0.441758988
T3	0.330075859	0.443455371
E17	0.403248907	0.451069956
E19	0.49249756	0.518468672
Rel8	0.4576639	0.544229387
Res12	0.729359656	0.556838558
Res11	0.444689993	0.561009375
Rel7	0.425275838	0.568142403
T1	0.517888747	0.591701765
E18	0.495860751	0.59803125
A13	0.587887197	0.621306343
Rel5	0.450621514	0.646886577
E20	0.623348409	0.701265274
Reason for referral	1	1

As can be seen in Table 3, the variable "Gender" shows no contribution to predictive power, while factors such as "Time Distance to Repair Shop" and "Main Reason for Visit" demonstrate moderate significance. Among the key variables, "E20" emerges as the most critical, with the highest weights in both methods (0.6233 for Information Gain and 0.7013 for Gini Index), indicating its essential role in influencing customer satisfaction. Similarly, "A13" and "Rel5" also rank highly, underscoring the importance of these dimensions in customer interactions. The comparison between Information Gain and Gini [23] Index reveals a consistent ranking for many variables, although some discrepancies exist, particularly among measures of responsiveness. This suggests that while both methods provide valuable insights, they may reflect different aspects of data distribution. To enhance predictive accuracy, organizations should prioritize strategies focused on high-weight variables such as empathy and assurance. Additionally, further analysis on lower-weight variables could uncover opportunities for improvement, ultimately leading to a more refined model that better addresses customer needs and preferences. Figure 2 shows the comparison of the FSS model.

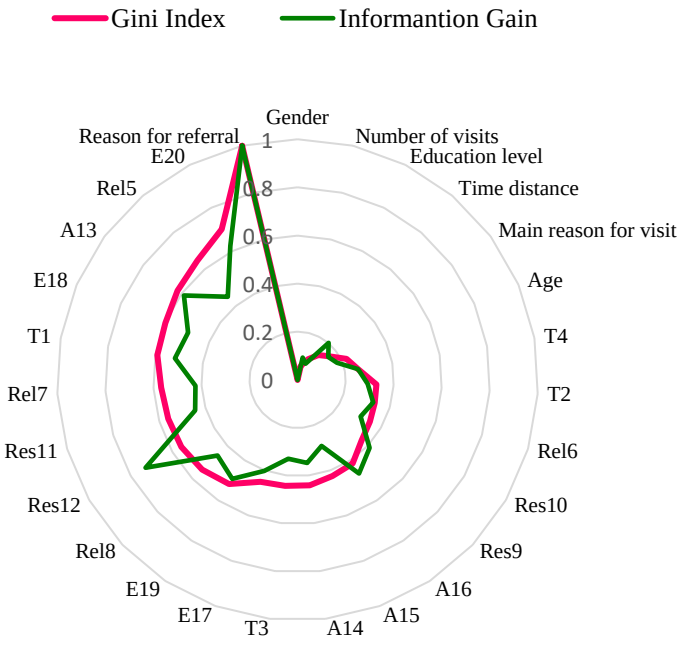


Fig. 2. Comparing FSS methods score for each feature

5.2. Results of modeling with decision tree types

In this section, the results of modeling with the decision tree algorithm are presented. Based on the weights given to each feature by the FSS algorithms in the second stage, K features with the highest weights are put into the model. In this study, K is investigated for the values of 5, 10, and 13. In each case, the accuracy, precision and recall of the model are measured. Accuracy, precision and recall have been calculated based on Eq. (7), Eq.(8), and Eq.(9) respectively. Table 4 demonstrates the modelling of the decision tree and ID3.

Table 4
The results of modeling by Decision tree and ID3

Model	Algorithm	FSS method	No. of selected Features	Accuracy	Recall	Precision
Model 1	Decision Tree	-	All	56.85% +/- 11.62%	30.10% +/- 6.63%	28.17% +/- 12.05%
Model 2		Information Gain	5	73.88% +/- 11.71%	48.03% +/- 7.86%	48.50% +/- 9.13%
Model 3		Gini Index	5	59.33% +/- 6.74%	33.56% +/- 4.41%	31.42% +/- 6.65%
Model 4		-	All	56.85% +/- 11.62%	30.10% +/- 6.63%	28.17% +/- 12.05%
Model 5		Information Gain	10	78.03% +/- 5.42%	54.11% +/- 10.84%	55.05% +/- 10.89%
Model 6		Gini Index	10	56.95% +/- 13.16%	31.32% +/- 9.06%	31.96% +/- 13.37%
Model 7		-	All	56.85% +/- 11.62%	30.10% +/- 6.63%	28.17% +/- 12.05%
Model 8		Information Gain	13	79.27% +/- 7.80%	58.42% +/- 15.68%	58.13% +/- 16.47%
Model 9		Gini Index	13	60.60% +/- 7.52%	42.51% +/- 16.95%	39.15% +/- 17.68%
Model 10	ID3	-	All	58.12% +/- 10.33%	46.15% +/- 16.23%	44.65% +/- 15.33%
Model 11		Information Gain	5	80.47% +/- 7.67% (	53.22% +/- 6.90%	53.68% +/- 5.03%
Model 12		Gini Index	5	61.35% +/- 9.79%	43.23% +/- 18.88%	43.10% +/- 19.70%
Model 13		-	All	58.12% +/- 10.33%	46.15% +/- 16.23%	44.65% +/- 15.33%
Model 14		Information Gain	10	82.58% +/- 7.77%	61.75% +/- 12.62%	60.03% +/- 12.40%
Model 15		Gini Index	10	59.28% +/- 8.01%	41.36% +/- 17.57%	41.91% +/- 17.69%
Model 16		-	All	58.12% +/- 10.33%	46.15% +/- 16.23%	44.65% +/- 15.33%
Model 17		Information Gain	13	85.47% +/- 6.00%	63.67% +/- 11.11%	62.50% +/- 10.79%
Model 18		Gini Index	13	58.05% +/- 11.31%	41.25% +/- 18.50%	40.69% +/- 16.98%

The results of modeling with the Decision Tree and ID3 algorithms reveal significant insights into the effectiveness of feature selection strategies. The baseline models that utilized all features showed relatively low performance, with accuracy rates hovering around 56-58% and recall and precision values below 30% to 46%. This indicates that without any feature selection, the models

struggle to make accurate predictions. However, when feature selection methods, specifically Information Gain and Gini Index, are applied, there is a marked improvement in model performance. Models utilizing Information Gain consistently outperformed those using the Gini Index across different configurations of selected features (K = 5, 10, and 13). For instance, Model 17, which selected 13 features based on Information Gain, achieved the highest accuracy of 85.47%, along with recall and precision values of 63.67% and 62.50%, respectively. In contrast, models based on the Gini Index tended to yield lower accuracy and predictive power. This trend underscores the importance of selecting relevant features to enhance model effectiveness. Overall, the analysis suggests that employing Information Gain for feature selection is a more effective strategy in decision tree modeling, leading to improved outcomes in predictive tasks.

5.3. Rule Extraction to Predict Customers' Loyalty

This section presents the extracted rules for predicting customer loyalty and their willingness to return to the repair shop. This prediction is based on the customers' available characteristics.

Table 5
The extracted "IF-THEN" rules

Row	IF	THEN Customer's willingness to return is:
1	the reason for referral = recommendations from friends and acquaintances and Res12 = 1 and T3 = 4	Low (1)
2	the reason for referral = recommendations from friends and acquaintances and Res12 = 3	High (3)
3	If the reason for referral = recommendations from friends and acquaintances and Res12 = 4 and E18 = 3 and Tangible3 = 3	Medium (2)
4	If the reason for referral = recommendations from friends and acquaintances and Res12 = 4 and E18 = 4	High (3)
5	If the reason for referral = recommendations from friends and acquaintances and Res12 = 5 and T3 = 3	Medium (2)
6	If the reason for referral includes recommendations from friends and acquaintances, reasonable price, quality of services, name and reputation of the repair shop, and T3 = 2	Medium (2)
7	If the reason for referral = reasonable price	High (3)
8	If the reason for referral = reasonable price and quality of services and Rel5 = 4	Medium (2)
9	If the reason for referral = reasonable price and quality of services and Rel5 = 5	High (3)
10	If the reason for referral = name and reputation of the repair shop and E18 = 2	Medium (2)
11	If the reason for referral = name and reputation of the repair shop and E18 = 3	High (3)
12	If the reason for referral = name and reputation of the repair shop and E18 = 4 and T3 = 5	Medium (2)
13	If the reason for referral = name and reputation of the repair shop and E18 = 5	High (3)
14	If the reason for referral = proximity to work or residence and A13 = 1	Low (1)
15	If the reason for referral = proximity to work or residence and A13 = 2	Medium (2)
16	If the reason for referral = proximity to work or residence and A13 = 3 and A15 = 3	Medium (2)
17	If the reason for referral = proximity to work or residence and A13 = 4	High (3)

Row	IF	THEN Customer's willingness to return is:
18	If the reason for referral includes proximity to work or residence along with recommendations from friends and acquaintances, T3 = 3, and R8 = 4	Medium (2)
19	If the reason for referral includes quality of services, E20 = 2, T3 = 2, and R7 = 2	Low (1)
20	If the reason for referral includes quality of services, E20=4, R9=4, R8=4	High (3)

As can be seen in Table 5, recommendations from friends and acquaintances are initially highlighted as a key factor in customer decision-making. For instance, Rule 1 indicates that if the reason for referral is based on recommendations and responsiveness (Res12) is low, while the quality of interaction (T3) is high, the customer's willingness to return is classified as low. Conversely, Rule 2 shows that if responsiveness is moderate (Res12 = 3), the likelihood of returning increases to high. This suggests that positive social influences can enhance customer loyalty, especially when service quality is perceived positively. Furthermore, reasonable pricing and service quality are critical factors influencing customer willingness. If the reason for visiting is solely based on reasonable pricing, the customer's likelihood of returning typically remains high. However, when reasonable pricing is combined with high service quality and a good reputation for the repair shop, this willingness can change and yield varying results depending on specific conditions. For example, a customer's willingness to return may decrease if service quality is high but staff empathy is low. The rules also indicate that the repair shop's reputation plays an important role in customer decision-making. If the reason for referral includes name recognition and credibility, but staff empathy is low (E18 = 2), the likelihood of customer return decreases (result 2). Conversely, as staff empathy increases, so does the probability of customer return. This underscores the importance of creating a positive customer experience, which can lead to increased loyalty. Proximity to work or residence is another key factor. If customers are close to the repair shop and feel a high level of assurance regarding services provided (A13 = 4), their willingness to return will increase (result 3). This suggests that easy access to repair services can create a competitive advantage for businesses. Combining various factors such as reasonable pricing, service quality, and shop reputation can significantly influence customer behavior. Many rules indicate that when multiple positive factors, such as recommendations from friends, reasonable pricing, and high service quality, are present, customers' willingness to return increases.

6. Conclusion

Building and maintaining long-term customer relationships is essential for any organization to thrive in competitive markets, especially for service organizations. Customers are increasingly aware of their options, and to do so, businesses must shift from simply paying for customer satisfaction to fostering trust and building mutually beneficial relationships.

In the competitive automotive repair industry, continuous improvement in service quality is essential for attracting and retaining customers, as satisfied clients are more likely to return and share their positive experiences. The automotive repair and maintenance service industry is not exempt from this rule and is one of the most competitive industries. This competition exists not only between different car brands but also between repair shops and service centres. In today's world, collecting customer data has become a common practice across industries, encompassing purchase history, online interactions, and customer reviews, which aid in analyzing behaviour. By employing advanced data analysis techniques, businesses can predict future customer behaviours, including their willingness to return to repair shops, a crucial aspect of the automotive service industry.

This research aims to model customers' willingness to revisit repair shops based on their characteristics and service quality indicators using the SERVQUAL model and machine learning algorithms to extract "if-then" rules that indicate customer loyalty levels. For this purpose, a two-stage model is proposed. The first stage addressed data pre-processing and feature selection, and classification and rule extraction were implemented in the second stage.

This paper used features available from customers and extracted from the SERVQUAL questionnaire to model loyalty prediction. In the feature selection stage, weights were assigned to the features using filtering methods, and the features' importance in predicting loyalty was determined. Afterwards, the selected features were entered into the decision tree and ID tree classification algorithms. Considering the accuracy of the models, the feature selection methods had a good effect on increasing the prediction. After selecting the best model, if-then rules were extracted to predict the loyalty of new customers.

Using the SERVQUAL model in repair shops can significantly increase service quality and customer satisfaction. The model helps repair shops identify their strengths and weaknesses in service delivery and improve service quality based on them. By focusing on different dimensions of service quality, including reliability, responsiveness, and empathy, repair shops can create a more positive experience for their customers. This leads to increased customer satisfaction and increases the likelihood of them returning to the repair shop, ultimately helping to build a loyal customer base. In addition, using machine learning techniques to analyze customer data can help repair shops predict future customer behaviour. By analyzing data collected from customer interactions, repair shops can better understand customer needs and expectations and adjust their services accordingly. This approach not only increases service efficiency but can also reduce marketing costs, as satisfied customers tend to share their positive experiences with others, which helps attract new customers.

Funding
This study did not receive any external financial support.

Conflicts of Interest
The authors declare no conflicts of interest.

References

[1] Li, W., Xu, Y., Jiang, T., & Cheung, C. (2024). The effects of over-service on restaurant consumers' satisfaction and revisit intention. International Journal of Hospitality Management, 122, 103881. https://doi.org/10.1016/j.ijhm.2024.103881

[2] Komijan, A. R., Yazdi, A. K., Tan, Y., Ocampo, L., & Nasrollahpourniazi, F. (2024). Spherical fuzzy multicriteria decision making for evaluating healthcare service quality of hospitals during the global pandemic. International Journal of Computational Intelligence Systems, 17(1), 105. https://doi.org/10.1007/s44196-024-00487-8

[3] Yazdi, A. K. (2015). Designing a mathematical model for indicators of service quality in the tourism industry based on SERVQUAL and Rembrandt methods. International Journal of Productivity and Quality Management, 15(4), 1–15. https://doi.org/10.1504/IJPQM.2015.069640

[4] Khalili-Damghani, K., Abdi, F., & Abolmakarem, S. (2019). Solving customer insurance coverage recommendation problem using a two-stage clustering-classification model. International Journal of Management Science and Engineering Management, 14(1), 9–19. https://doi.org/10.1080/17509653.2018.1467801

[5] Mehdiabadi, A., Yazdi, A. K., Wanke, P. F., & Correa, H. L. (2023). After-sales indicators in the liquefied petroleum gas industry: A hybrid multi-criteria decision-making approach in an uncertain environment. International Journal of Oil, Gas and Coal Technology, 32(4), 392–412. https://doi.org/10.1504/IJOGCT.2023.129582

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

[7] Ong, A. K. S., et al. (2023). Determination of factors affecting customer satisfaction towards 'Maynilad' water utility company: A structural equation modeling-deep learning neural network hybrid approach. Heliyon, 9(3), e13798. https://doi.org/10.1016/j.heliyon.2023.e13798

[8] Ruiz, E., Yushimito, W. F., Aburto, L., & de la Cruz, R. (2024). Predicting passenger satisfaction in public transportation using machine learning models. Transportation Research Part A: Policy and Practice, 181, 103995. https://doi.org/10.1016/j.tra.2024.103995

[9] Zibarzani, M., et al. (2022). Customer satisfaction with restaurants service quality during COVID-19 outbreak: A two-stage methodology. Technology in Society, 70, 101977. https://doi.org/10.1016/j.techsoc.2022.101977

[10] Wang, J., Wu, J., Sun, S., & Wang, S. (2024). The relationship between attribute performance and customer satisfaction: An interpretable machine learning approach. Data Science and Management, 7(3), 164–180. https://doi.org/10.1016/j.dsm.2024.01.003

[11] Rezende, G. A., Mariano, A. M., Santos, M. R., & Constatin, A. C. C. (2024). Modeling customer satisfaction in the food industry: Insights from a structural equation approach. Procedia Computer Science, 242, 130–137. https://doi.org/10.1016/j.procs.2024.08.250

[12] Munde, A., & Kaur, J. (2024). Predictive modelling of customer sustainable jewelry purchases using machine learning algorithms. Procedia Computer Science, 235, 683–700. https://doi.org/10.1016/j.procs.2024.04.066

[13] Sun, Y., Liu, H., & Gao, Y. (2023). Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model. Heliyon, 9(2), e13384. https://doi.org/10.1016/j.heliyon.2023.e13384

[14] Tanwar, R., & Agarwal, P. K. (2024). Analysis of the determinants of service quality in the multimodal public transport system of Bhopal city using structural equation modelling (SEM) and factor analysis. Expert Systems with Applications, 256, 124931. https://doi.org/10.1016/j.eswa.2024.124931

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

[16] Nilashi, M., et al. (2022). Factors impacting customer purchase intention of smart home security systems: Social data analysis using machine learning techniques. Technology in Society, 71, 102118. https://doi.org/10.1016/j.techsoc.2022.102118

[17] Akter, T., et al. (2024). The moderating effect of system quality on the relationship between customer satisfaction and purchase intention: PLS-SEM & fsQCA approaches. Journal of Open Innovation: Technology, Market, and Complexity, 10(4), 100381. https://doi.org/10.1016/j.joitmc.2024.100381

[18] Jiao, L., Yang, H., Liu, Z., & Pan, Q. (2022). Interpretable fuzzy clustering using unsupervised fuzzy decision trees. Information Sciences, 611, 540–563. https://doi.org/10.1016/j.ins.2022.08.077

[19] Rajesh, B., Vardhan, M. V. S., & Sujihelen, L. (2020). Leaf disease detection and classification by decision tree. In 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 705–708). IEEE. https://doi.org/10.1109/ICOEI48184.2020.9142988

[20] Kim, H., & Loh, W.-Y. (2001). Classification trees with unbiased multiway splits. Journal of the American Statistical Association, 96(454), 589–604. https://doi.org/10.1198/016214501753168271

[21] Khalili-Damghani, K., Abdi, F., & Abolmakarem, S. (2018). Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: Real case of customer-centric industries. Applied Soft Computing, 73, 816–828. https://doi.org/10.1016/j.asoc.2018.09.001

[22] Maldonado, S., & Weber, R. (2009). A wrapper method for feature selection using support vector machines. Information Sciences, 179(13), 2208–2217. https://doi.org/10.1016/j.ins.2009.02.014

[23] Do, V., & Usunier, N. (2022). Optimizing generalized Gini indices for fairness in rankings. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 737–747). ACM. https://doi.org/10.1145/3477495.3532035