Ekonomicko-manazerske spektrum
2024, Volume 18, Issue 2, pp. 85-98

CUSTOMER SATISFACTION PREDICTION: A CASE STUDY FOR ELECTRO-BIKE CUSTOMERS

Lucia Svabova[1,a], Veronika Labosova[1,b,*], Martina Borovcova[2,c] and Natalia Jarabicova[3,d]

[1]Department of Economics, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitna 1, 010 26 Zilina, Slovakia
[2]Full Department of Finance, Faculty of Economics, VSB-TU Ostrava, 17. listopadu 2172/15, Ostrava-Poruba, Czech Republic
[3]Graduate at the Department of Economics, Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitna 1, 010 26 Zilina, Slovakia
[a]lucia.svabova@uniza.sk, [b]labosova1@stud.uniza.sk, [c]martina.borovcova@vsb.cz, [d]jarabicovanatalia@gmail.com
[*]Corresponding author

Cite as: *Svabova, L., Labosova, V., Borovcova, M., Jarabicova, N. (2024). Customer satisfaction prediction: A case study for electro-bike customers, Ekonomicko-manazerske spektrum, 18(2), 85-98.*

Available at: *10.26552/ems.2024.2.85-98*

Received: 8 September 2024; Received in revised form: 27 October 2024; Accepted: 29 November 2024; Available online: 31 December 2024

Abstract:
Research background: Customer satisfaction is one of the key factors influencing businesses' success and long-term profitability. Understanding customer needs and preferences is particularly important in the e-bike segment, which is experiencing dynamic growth. Satisfied customers make repeat purchases, and their recommendations help build a positive brand image, which is essential in a competitive market. Identifying the factors that most influence satisfaction enables companies to adapt their products and services to meet customer expectations better and thus secure a stable position in the market.
Purpose of the article: The article aims to analyse customer satisfaction with e-bikes and identify its key factors. Based on the findings, it offers recommendations for dealers while also proposing a prediction model to understand customer needs better.
Methods: Customer satisfaction with e-bikes was measured through an online survey in the form of a questionnaire. Data collection was conducted from January 2024 to the end of February 2024 and involved 388 respondents. The data collected will be used to create predictive models of customer satisfaction, the accuracy and quality of which will be subsequently compared. The analysis results will also be presented graphically, allowing for a better interpretation of the relationships and trends identified.
Findings & Value added: Five models were created to predict customer satisfaction. The logistic regression method was the most effective based on overall accuracy and sensitivity. This method identified engine brand, battery capacity, and product purchase location as significant variables. This model allows sellers to tailor their offerings better and provide goods significantly contributing to higher customer satisfaction.

Keywords: customer satisfaction; satisfaction prediction; prediction model; e-bikes

JEL Classification: C00; C02; M2

1. Introduction

Paying attention to customer satisfaction is a very key factor nowadays. A good customer experience increases the probability of repeat purchases and overall brand loyalty. A bad customer experience can very quickly damage the reputation of an entire company or brand. If customers are disappointed with products or services, they make it known through social media. Social networks and media have become a firm part of many people's daily lives. However, delivering a good experience and quality customer service is a pragmatic balance between reducing operational costs and investing additional efforts to improve the experience and drive revenue. Customer satisfaction surveys are a great way to understand how customers feel and think and discover what new customers might like about a product. A high satisfaction rate is an important indicator of retaining and winning new customers.

Customer satisfaction refers to the extent to which customers are satisfied with their experience with a given product, service or store. It is a key business metric because it directly affects their success and long-term vitality. If a customer's expectations are met or even exceeded, they are likely to be satisfied (Antonides and Hovestadt, 2021). Happy and satisfied customers mostly become repeat customers, and their recommendations contribute positively to the company's reputation. On the contrary, if the product does not meet the customer's expectations and needs, it brings disappointment, and the customer will probably not continue to purchase that business's products (Knapcikova et al., 2021). Factors that may cause customer dissatisfaction are, for example, poor quality, sub-standard service or unmet expectations. Dissatisfied customers show their dissatisfaction through negative reviews, complaints or by ceasing to be a business customer. It is very important to avoid these complaints because, in today's connected world, these complaints are quickly shared on social media, which can affect the business's overall reputation (Park et al., 2021).

Customer satisfaction results in loyalty, which is often referred to as loyalty. It is what drives a person to repeat purchases and encourages existing customers. Customer care solutions researcher Leonie Brown says: "People who have had a bad experience with a brand, but the brand has made it right, are more loyal than customers who never had a problem in the first place. This is because it involves trust."

Customers are loyal for a variety of reasons that can be grouped into the following loyalty categories (Oracle, n.d.):

- satisfied customers – this category of customers like the products or services, have shopped many times and never complained,
- price-loyal customers – this category of customers is satisfied with the low price of the product or service; if they can save money elsewhere, they will leave,
- loyal to the loyalty program – customers are not loyal to the company but to the loyalty program offered by the company,
- loyal for the benefits – the brand does not attract customers, but by the benefits offered by the company (free internet connection), they buy only sporadically,
- truly loyal customers – customers who make repeat purchases and talk about a great experience with the company.

If an organisation cannot resolve the issue of its customer and its complaints, the business can use ISO 10003 standards, which provide guidelines for both the organisation and external dispute service providers to investigate the complaints further and come to a satisfactory conclusion (Ang and Buttle, 2012).

Measuring customer satisfaction is key for businesses to understand how well they meet customer expectations and identify improvement areas. If a business measures customer satisfaction, it can grow in the marketplace, improve its brand image, uncover its strengths and weaknesses, and gather new inspiration. Customer review data can be collected from various sources, including online platforms and social networking services. Although customer reviews vary in many data points, they generally provide information about overall satisfaction ratings, customer profiles and content. The basic measure is the overall rating, which can be further used to assess service quality, identify customer complaints and conduct detailed statistical research (Kim and Lim, 2021).

The tools that can be used to measure satisfaction are divided by Hayes (2008) into online surveys, email surveys, customer feedback forms, online reviews and ratings, customer interviews and others. The experiment is one of the most valuable scientifically based research to uncover relationships, causes and consequences of influences. However, using this method is time-consuming. A more widely used method is observation. It is the process of learning about and recording facts without interfering with them. In other words, it helps us discover what is happening. Observation is often combined with enquiry, one of the most widely used and widely used methods. It is the collection of information based on the responses of the persons under investigation using questions. This method includes questionnaires. Data collection can be carried out via face-to-face, telephone, or written questionnaires. The advantage of the online questionnaire solution is that it does not require such a high cost and burden as questionnaires sent by post or in person. It is also possible to reach a segment of the population that is difficult to contact from the company's side. From the customer's side, the advantage is that it is mostly anonymous. (Mateides, 2001a; Mateides, 2001b, Islam et al., 2021)

The aim of the research in this paper is to develop a prediction model of customer satisfaction with e-bikes. Satisfaction includes aspects such as purchase, company attitude and customer care while also considering the customer's initial preference when purchasing the product. Based on the developed model, recommendations can be formulated for manufacturers and sellers of the type of goods.

2. Literature review

Nowadays, every individual is in the role of a customer almost every day. Understanding customers' value and expectations of their treatment is the basis for customer satisfaction analysis. The term "customer satisfaction" can be defined as an emotional state resulting from a positive evaluation of a customer's experience with a company (Agag et al., 2024). In the dynamic modern environment, customers are exposed to a wide range of products and services, with their choices determined by individual preferences and needs. According to Ashworth and Bourassa (2020), customer satisfaction depends on the extent to which their needs and expectations are met. If a product or service fails to meet these needs, it may lead to dissatisfaction and a reduced likelihood of repeat purchases. The issue of customer dissatisfaction and complaints has also been addressed by Kim and Lim (2021), who emphasised the importance of effective service quality management in their research. These authors describe how using customer ratings and statistical process analysis enables businesses to ascertain customer satisfaction systematically and prevent service failure through early identification of serious problems. The combination of these methods supports the ability of businesses to respond effectively to customer complaints at optimised costs. Besides customer satisfaction, understanding consumer behaviour also significantly influences the customer's final decision. With the growing importance of a holistic approach, companies are increasingly focusing on creating emotional experiences. Vrtana and Krizanova (2023) investigated the impact of emotionally attuned advertising appeals on the purchase behaviour of Slovak consumers. Their results showed a complex relationship between advertising with emotional appeal and purchase behaviour, with negative effects outweighing positive ones in certain age groups.

Protecting the environment is a key global priority, with companies and individuals increasingly seeking sustainable solutions for everyday life. E-bikes have become an important alternative to traditional means of transport as they combine eco-friendliness, flexibility and low cost. Their use contributes to reducing air pollution and promotes the sustainable development of urban areas. Studies by Xia et al. (2022) and Li et al. (2020) investigated the satisfaction of e-bike users in China and identified service quality as a critical factor influencing consumer choice. They found that high service quality positively impacts customer satisfaction, which is particularly crucial when introducing new products such as e-bikes. Nematchoua et al. (2020) focused their analysis on the factors limiting e-bike use, listing lack of safe cycle lanes, high procurement costs, inadequate parking infrastructure and concerns about theft as the main barriers. Despite these barriers, e-bikes scored better in several categories than conventional bicycles, cars or public transport. The e-bike sector is a dynamically growing segment that has the potential to make a significant contribution to the transition to greener cities. However, this development requires an analysis not only of the bicycle production itself but also of the organisation of the whole supply chain. Mina et al. (2024) focused on identifying factors influencing consumer preferences for e-bikes. They aimed to propose strategies to increase the attractiveness of these means of transport, thereby promoting their wider adoption in urban environments.

Customer satisfaction can be measured using various statistical methods, one of the most widely used being the questionnaire survey. According to Terek (2019), the questionnaire survey process includes questionnaire design, data collection and data analysis. In his publication, the author describes in detail the steps required to effectively conduct surveys, from the design of questions to the use of different data collection methods to the statistical processing of the results. Similarly, Patra (2019) defines a questionnaire as a systematic data collection instrument in which respondents answer a series of questions. He emphasises that the design of the questionnaire must be carefully thought out, emphasising the relevance and simplicity of the questions. According to the author, the survey's success is directly dependent on the quality of the questionnaire design, as incorrectly worded questions can skew the results. However, modern technological advances bring new challenges in data collection through questionnaires. Lebrun et al. (2024) point out the risk of automatic completion of questionnaires by artificial intelligence, which can negatively affect the quality of the data collected. Although various platforms for creating affordable and user-friendly questionnaires exist, their dependence on AI-based technologies poses potential problems. The authors highlight the need to find efficient solutions that ensure that the answers in questionnaires come from real respondents, thus maintaining the validity of the data collected.

In their research, Kamalja et al. (2024) describe the use of statistical methods such as chi-square test, tests of independence, and analysis of variance (ANOVA) to identify significant relationships between variables. In addition to retrospective analysis, data can be predicted using modern methods, including machine learning. In their study, Ulrich et al. (2023) present the use of neural networks as a prediction model, and their results confirm the high efficiency of this method in prediction and the development of business strategies. Such prediction enables organisations to understand customer behaviour better and make personalised business proposals to increase customer satisfaction. Another important prediction method is decision trees, which are often used to analyse market trends. Pitka et al. (2024) emphasise the transparency and efficiency of decision trees, which operate on the principle of partitioning data into individual nodes and leading to accurate decisions. In their study, the authors applied decision trees to classify customer orders to identify patterns in customer behaviour. This method significantly contributes to organisations seeking to understand their market better and tailor their strategies to customer needs.

Measuring customer satisfaction requires a thoughtful approach to questionnaire design and analysis of the data collected. In addition to traditional statistical methods, modern machine learning tools are finding applications, providing new opportunities for understanding customer behaviour. These insights provide a valuable basis for improving customer satisfaction and optimising business strategies.

3. Methodology

In data science, solving the customer satisfaction prediction problem is a classification problem since customer satisfaction is the output variable of interest. For this purpose, we use overall customer satisfaction as the dependent variable. The customer himself determined its value on a 5-degree satisfaction scale. However, for the purpose of this study, we have further converted satisfaction into a dichotomous variable with values of satisfied and dissatisfied customers. We took this step because of the poor representativeness of some of the five categories, which would consequently affect the predictive performance of the models developed.

The satisfied customer group in the dataset includes 342 respondents, while the dissatisfied group includes 46 respondents. As explanatory variables, we included the place of purchase, battery capacity, accessories, brand and type of e-bike, price and the last factor, battery life, in the model.

3.1. Questionnaire survey

We used an online survey in the form of a questionnaire to measure customer satisfaction with the e-bike. The questionnaire was created using the Google Forms platform. It was published in January 2024. The data was collected through social networks. Data collection lasted until the end of February 2024. The target group was exclusively customers who own, know and use the product regularly.

The questionnaire contained 31 closed questions. Of these, the first six questions describe the respondent’s position in society: i.e. gender, age, education, region, social status and net monthly income. The remaining questions were product-related and depended on various factors, such as the form of product purchase in a brick-and-mortar store or online. One of the questions was about overall customer satisfaction, and respondents had a choice of five satisfaction levels: very satisfied, satisfied, neutral, dissatisfied and very dissatisfied.

The questionnaire concludes with two questions. The first asked what attributes customers would like to see to make an e-bike more attractive. In the last question, customers could formulate their own recommendations for the manufacturer. These questions are semi-closed, where the respondent could decide whether to choose one answer to the question from the options offered or to write their answer.

3.2. Methods of creating a customer satisfaction prediction model

We then used the data collected through the questionnaire survey to create a predictive customer satisfaction model. The model uses the customer’s socio-economic characteristics and preferences in purchasing and using an e-bike.

 ISSN 1337-0839 (print) / 2585-7258 (online)

To predict customer satisfaction, we created and compared several prediction models. To create them, we used four machine learning methods: neural networks, classification trees, logistic regression, and the nearest neighbour method. We characterise these methods briefly.

Neural networks are considered one of the best machine learning algorithms and have a wide range of applications. An artificial neural network is an imitation of a real human neural network. It consists of nodes - neurons interconnected by synaptic fibers, and these connections between neurons have valuations. The neurons are arranged in layers that consist of an input layer, several hidden layers and an output layer. The most important advantage of a neural network is the high accuracy of the predictions. On the other hand, the disadvantage is the computational complexity and the very difficult to impossible interpretation of the model (Taheri et al., 2021).

Decision trees are a useful machine learning algorithm. The tree model has a hierarchical structure consisting of nodes, with the base node called the root node. In the process of decision-making, these nodes gradually branch into smaller parts, resulting in a tree-like structure. Each tree node represents a decision rule according to a selected attribute of the classified object, automatically selected by the method along with a threshold value. Thus, the whole tree model is a set of classification rules. Compared to neural networks, decision trees are more flexible and easy to understand and interpret. Their disadvantage is the need for a large sample size and the prioritization of some variables in rule generation. Different techniques are used for decision tree formation; in this study, we used CHAID (Chi-squared Automated Interaction Detection) and CRT (Classification and regression trees) type trees. CHAID is a decision tree construction technique that is designed for categorical data. It can also construct non-binary trees, meaning nodes can have more than two branches. CRT creates binary trees, meaning only two branches lead from a single node (Gunduz and Al-Ajji, 2022).

Logistic regression is a regression model that is able to predict the probability of an event occurring based on a set of input variable values. This probability can then be categorized to obtain a solution to the classification problem of predicting customer satisfaction. The advantage of logistic regression compared to other methods is its relative ease of interpretation. Limitations of this method include the requirement to meet several assumptions, such as the linearity of the relationship between the dependent variable and potential explanatory variables (Dong et al., 2022).

The nearest neighbour method is used to classify cases based on their similarity to other cases. It is one of the simpler methods to solve the classification problem. The nearest neighbours are found by calculating the distances between the vectors of values of the input variables, and this model predicts the new data cases into a class of found nearest neighbours. The advantage of this model is its ease of interpretation and efficiency for large training sets. The disadvantage of the method is computational complexity because, for each data case, the algorithm must compute distances to all other cases in all variables in the data set. The method is sensitive to missing data (Settey et al., 2021).

All statistical tests are evaluated using the p-value of the test, which is compared to the significance level, which was set at 0.05 for the purpose of this study (Labudova, 2021; Machery, 2021).

3.3. Performance evaluation of prediction models

We use evaluation statistics and the ROC curve to assess the performance of the predictions for the developed models (Maternova et al., 2023). The evaluation statistics are based on the numbers of correctly and incorrectly classified data instances in the group of truly satisfied and dissatisfied customers. These counts are reported in the so-called pronoun matrix.

We can determine the overall accuracy of the model as the proportion of correctly classified customers out of all customers:

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

where TP (true positive) dissatisfied customers are correctly predicted

FP (false positive) are customers predicted by the model to be dissatisfied but are actually satisfied

FN (false negative) are customers predicted by the model to be satisfied but are actually dissatisfied

TN (true negative) satisfied customers are correctly predicted

For this label, we consider the group of interest (hit) to be dissatisfied customers, who are indicated by the value of the dependent variable $Y = 1$.

In addition to the overall accuracy, the sensitivity of the model, which represents the proportion of correctly classified positive (i.e. dissatisfied) customers, is very important:

$$sensitivity = \frac{TP}{TP + FN} \tag{2}$$

These indicators, together with the ROC (Receiver Operating Characteristic) curve, which shows the relationship between sensitivity and specificity (specificity is the proportion of satisfied customers correctly classified by the model out of all satisfied customers), provide a good picture of the performance of the models. In general, the larger the Area Under the Curve (AUC) of the ROC curve, the better the classification ability of the model. The ideal condition is when it reaches a value close to 1.

To improve the model’s predictive performance, we use sample balancing of satisfied and dissatisfied customers. We refer to unbalanced data when one category is over-represented and the other under-represented in the data. In our case, the group of satisfied customers was predominant (342), while the group of dissatisfied customers was smaller (46). This disproportion may affect the accuracy of the model. Unbalanced data can be weighted through the undersampling or oversampling technique. Undersampling means that we randomly select only enough units from the majority category to be approximately equal in size to the minority category. Conversely, oversampling means assigning a higher weight to the minority category of dissatisfied customers so that it is approximately equal in size to the majority category of satisfied customers. In this study, we used the oversampling technique.

If we want to use the model in practice, we need to test its quality on independent data, different from those used for model creation. The most common solution is randomly splitting the data into test and training sets. In this study, we use a 60:40 split of the dataset. We will use the training data to specify the model and the test data to evaluate the functionality of the model.

3.4. Data sample description

For the purpose of this study, it was possible to collect completed questionnaires from 468 adult citizens of the Slovak Republic aged 18 years and over. More men than women participated in the research. In terms of the age of respondents, the age group 26-45 years was predominant. More than half of the respondents had attained the highest secondary education level. Respondents in all regions of Slovakia were surveyed, with the Zilina Region having the largest representation with 28.8%. On the other hand, the lowest number of respondents came from the Presov and Nitra regions (6%). This research involved mainly employed respondents, whose total percentage of the total sample was 56.4%. On the other hand, the smallest group were respondents without employment (2.4%). The respondent's net monthly income was mostly in the range of €1,001-1,500 (34.6%). Respondents with no income had the smallest representation. A detailed description of the frequency distribution of respondents in the sample is presented in Table 1.

Table 1: Distribution of respondents according to their socio-economic characteristics

Variable name	Values	Relative frequency [%]
gender	women	45.3
	men	54.3
age	19-25	19.4
	26-45	47.9
	46-65	27.4
	more than 65	5.3
education	primary	2.4
	secondary	54.3
	university	43.3
region	BA	23.1
	TT	10.9
	TN	9.2
	NR	6.0
	ZA	28.8
	BB	8.1
	PO	6.0
	KE	7.9
social status	student	12.8
	unemployed	2.4
	employee	56.4
	self-employee	20.5
	pensioner	7.9
monthly net income	no income	4.3
	lower than 500 €	9.6
	501-1,000 €	23.9
	1,001-1,500 €	34.6
	more than 1,500 €	27.6

Source: own elaboration

Of all respondents, 82.9% (388) indicated owning an e-bike. The remaining 17.1% (80 respondents) do not actually own one. We will further process the survey with a total of 388 respondents.

4. Results

To predict customer satisfaction, we created a total of 5 different models using the decision tree method (CHAID and CRT type tree), neural networks, logistic regression and nearest neighbour. In addition to the models themselves, we observe their prediction performance, which we evaluate and compare using their sensitivity and overall accuracy on a test set (Table 2).

Regarding model quality, we want the sensitivity and overall accuracy values to be as high as possible. Based on these results, the neural network has the highest overall accuracy of 89.6% on the test set (90.4% on the training set). Thus, from this point of view, we would choose the neural network model as the best-performing model, but in terms of using the prediction model in practice, it is very important to find the customers who will be dissatisfied. However, the neural network model's sensitivity on the test set is only 9.1% (25.7% on the training set). The

Table 2: Prediction performance of models for customer satisfaction prediction

Method	Sensitivity [%]	Accuracy [%]
CHAID tree	78.3	78.5
CRT tree	76.1	79.2
neural network	9.1	89.6
logistic regression	81.8	82.1
nearest neighbor	18.9	85.7

Source: own elaboration

confusion matrix of the prediction model developed by the neural network method for the test data set is shown in Table 3.

Table 3: Confusion matrix for neural networks model

Actual state / Prediction	Satisfied	Dissatisfied	Total
Satisfied	94	1	95
Dissatisfied	10	1	11
Total	104	2	106

Source: own elaboration

For the model to meet both qualitative criteria at a high level, we selected logistic regression as the most appropriate method, which achieves a sensitivity on the test set of 82.1% (84.8% on the training set) and the overall accuracy of the model on the test set is 81.8% (83.6% on the training set). The confusion matrix for the logistic regression model is in Table 4.

Table 4: Confusion matrix for the logistic regression model

Actual state / Prediction	Satisfied	Dissatisfied	Total
Satisfied	78	17	95
Dissatisfied	2	9	11
Total	80	26	106

Source: own elaboration

From the neural network model, it is impossible to say exactly what role the individual input variables play in the overall customer satisfaction or dissatisfaction. The neural network itself is very complex and uninterpretable. It consists of two input neurons (satisfied and dissatisfied customers), a hidden layer of 6 neurons, and an output layer consisting of 45 output neurons. Due to the number of connections between neurons in the model, it is impossible to interpret this model in detail.

In contrast, the logistic regression model offers the possibility of interpretations of the estimated coefficients of the input variables. The model was constructed using the forward stepwise method in 6 steps. The full model is presented in Annex A.

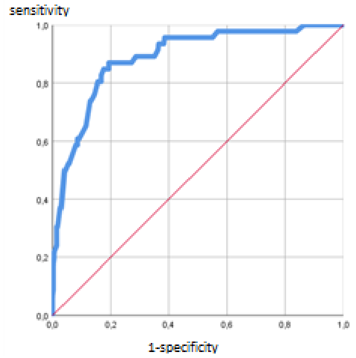
Checking the individual variables, we find that price and e-bike accessories are insignificant variables for customer satisfaction with this product. On the contrary, battery life, battery capacity, motor brand and place of e-bike purchase are significant factors. Customers who purchased an e-bike with accessory (2) (i.e., rack) are 1.7 times more likely to be dissatisfied compared to those who purchased an e-bike with accessory (5) (i.e., locks or alarm). If the battery life of the e-bike is low (denoted as battery life (1)), customers are 82.3 times more likely to be dissatisfied compared to those who purchased an e-bike with high battery life (5). The overall effect of the battery capacity variable on customer satisfaction is statistically significant. However, within each capacity category, all are statistically insignificant in the model. Thus, the only significant category is the reference category (5), i.e. the highest battery capacity. Considering the coefficients of the other battery capacities, it is clear that all lower battery capacities increase the probability of customer dissatisfaction compared to the highest battery capacity of their e-bike.

 ISSN 1337-0839 (print) / 2585-7258 (online)

Regarding the brand of electro-motor, if customers purchased the Eovolt brand (marked as (4)), they are 14 times more likely to be dissatisfied compared to those who would have purchased a brand other than all of the above (marked as (10)). If customers received an e-bike as a gift (marked as (2)), they are 13.9 times more likely to be dissatisfied compared to if they purchased the e-bike themselves through brick-and-mortar stores (marked as (3)).

In addition to the evaluation statistics listed in Table 2, we also evaluate the model's performance using the ROC curve and the size of the area under the AUC. The ROC curve is shown in Figure 1. The horizontal axis shows 1-specificity, and the vertical axis shows sensitivity.

Figure 1: ROC curve for logistic regression

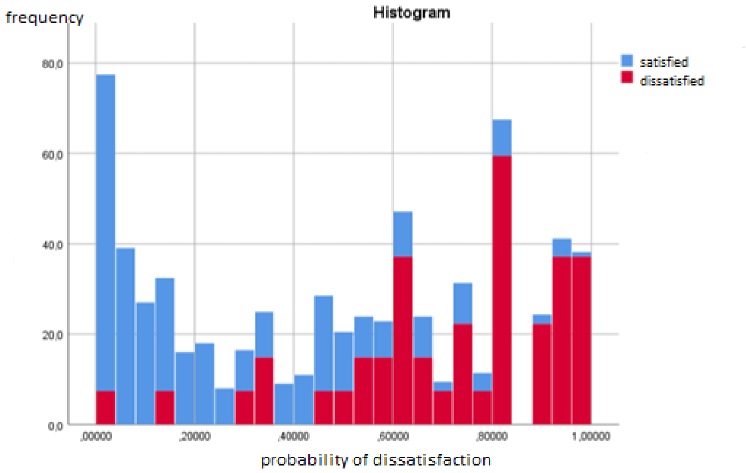


Source: own elaboration

The size of the AUC region is 0.88, which can be considered good to very good classification ability.

We can use a histogram (Figure 2) to show the frequency of occurrence of different values of the predicted probability of customer dissatisfaction.

Figure 2: Histogram of frequencies of probability of dissatisfaction



Source: own elaboration

The histogram shows that as the probability of dissatisfaction predicted by the model increases (on the horizontal axis, dissatisfaction probability values higher than 0.5), the proportion of truly dissatisfied customers increases (red). Conversely, with predicted satisfaction (probabilities lower than 0.5), truly satisfied customers predominate (blue color).

5. Discussion

Predicting customer satisfaction is a key task for businesses aiming to understand and improve the customer experience. Using predictive models, businesses can anticipate which factors affect customer satisfaction and take proactive steps to improve their services or products. By predicting customer satisfaction, businesses are able to improve customer relationships, as well as retain or increase customer loyalty, in addition to proactively solving problems. Ultimately, businesses can increase profitability by delivering exceptional experiences tailored to the needs and preferences of their customers.

In this study, we concluded that although a neural network is often a very accurate method for making predictions, the model produced tends to be complex, and its results cannot be accurately interpreted. The neural network model created had a very low sensitivity for finding dissatisfied customers, which is precisely the category that needs to be accurately predicted in order to take early action to avoid this dissatisfaction. A more appropriate model was stepwise logistic regression, whose model can also be interpreted to find significant factors for customer dissatisfaction.

When the predicted probabilities of customer dissatisfaction are plotted using a histogram colour-coded by their actual satisfaction, one can further consider shifting the threshold for categorising customers into satisfied or dissatisfied groups. The default value in logistic regression is always at 0.5, whereby customers with predicted probabilities below 0.5 are categorised as satisfied, and, in contrast, those with predicted probabilities above 0.5 are categorised as dissatisfied. According to the histogram shown, it would be possible to increase the achieved level of sensitivity of the model at the expense of its overall accuracy by shifting this threshold to, for example, 0.44. This would capture more potentially dissatisfied customers. The downside would be that more genuinely satisfied customers would be unnecessarily predicted as dissatisfied, which would mean that the firm would unnecessarily take action to reduce their dissatisfaction. In this case, however, it is worth considering whether it is costlier to encourage satisfied customers unnecessarily or to avoid losing more potentially dissatisfied customers.

6. Conclusions

Increasing customer satisfaction is essential to the success and growth of a business. By understanding the needs of their customers and offering quality products and services and quality service, dealers can build a base of loyal customers. In this study, using the data collected from a questionnaire survey and creating a prediction model using logistic regression, we found that the brand of motor, battery capacity, battery life and the location where the customers purchased the e-bike significantly impacted customer satisfaction with the product. Thus, the brand of the e-motor and the battery capacity of the e-bike, which came out as the most important factors for overall customer satisfaction, should be mainly considered when determining the firm's activities to eliminate customer dissatisfaction. In contrast, the price and accessories of the e-bike did not emerge as significant factors for customer satisfaction.

In this study, we did not use the socio-economic characteristics of the customer as a possible determinants of customer satisfaction in developing the prediction model. Thus, further study could be aimed at finding significant customer satisfaction factors both as product parameters, as implemented in this study, but also in combination with customer characteristics.

Author contributions: All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Funding: This research was funded by the Slovak Grant Agency for Science (VEGA) under Grant No. 1/0509/24.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

Agag, G., Shehawy, Y. M., Almoraish, A., Eid, R., Chaib Lababdi, H., Gherissi Labben, T., & Abdo, S. S. (2024). Understanding the relationship between marketing analytics, customer agility, and customer satisfaction: A longitudinal perspective. *Journal of Retailing and Consumer Services*, *77*, 103663.

Ang, L., & Buttle, F. (2012). Complaints-handling processes and organisational benefits: An ISO 10002-based investigation. *Journal of marketing management, 28*(9), 1021-1042.

Antonides, G., & Hovestadt, L. (2021). Product Attributes, Evaluability, and Consumer Satisfaction. *Sustainability*, *13*(22), 12393.

Ashworth, L., & Bourassa, M. A. (2020). Inferred respect: A critical ingredient in customer satisfaction. *European Journal of Marketing*, *54*(10), 2447–2476.

Dong, Y., Frees, E. W., Huang, F., & Hui, F. K. C. (2022). Multi-state modelling of customer churn. *ASTIN Bulletin: The Journal of the IAA*, *52*(3), 735–764.

Gunduz, M., & Al-Ajji, I. (2021). Employment of CHAID and CRT decision tree algorithms to develop bid/no-bid decision-making models for contractors. *Engineering, Construction and Architectural Management*, *29*(9), 3712–3736.

Hayes, B. E. (2008). *Measuring Customer Satisfaction and Loyalty: Survey Design, Use, and Statistical Analysis Methods*. Asq Press.

Islam, T., Islam, R., Pitafi, A. H., Xiaobei, L., Rehmani, M., Irfan, M., & Mubarak, M. S. (2021). The impact of corporate social responsibility on customer loyalty: The mediating role of corporate reputation, customer satisfaction, and trust. *Sustainable Production and Consumption*, *25*, 123–135.

Kamalja, K. K., Khangar, N. V., & Beh, E. J. (2024). A New Algorithm for the Partition of Pearson's Chi-Squared Statistic for Multiway Contingency Table. *Journal of the Indian Society for Probability and Statistics*, *25*(1), 121–149.

Kim, J., & Lim, C. (2021). Customer complaints monitoring with customer review data analytics: An integrated method of sentiment and statistical process control analyses. *Advanced Engineering Informatics*, *49*, 101304.

Knapcikova, L., Behunova, A., & Behun, M. (2023). The Strategic Impact of E-Business on Competitiveness of the Enterprise. *Mobile Networks and Applications*, *28*(1), 211–219.

Labudova, V., Pacakova, V., Sipkova, L., Soltes, E., & Vojtkova, M. (2021). *Statisticke metody pre ekonomov a manazerov.* Wolters Kluwer.

Lebrun, B., Temtsin, S., Vonasch, A., & Bartneck, C. (2024). Detecting the corruption of online questionnaires by artificial intelligence. *Frontiers in Robotics and AI*, *10*, 1277635.

Li, L., Zhu, B., Jiang, M., Cai, X., Lau, A. K. W., & Shin, G.-C. (2020). The role of service quality and perceived behavioral control in shared electric bicycle in China: Does residual effects of past behavior matters? *Environmental Science and Pollution Research*, *27*(19), 24518–24530.

Machery, E. (2021). The Alpha War. *Review of Philosophy and Psychology*, *12*(1), 75–99.

Mateides, A., (2001a). *Spokojnost zakaznika a metody jej merania.* (1. rd ed.). Epos.

Mateides, A., (2001b). *Spokojnost zakaznika a metody jej merania*. (2. rd ed.). Epos.

Maternova, A., Materna, M., David, A., Torok, A., & Svabova, L. (2023). Human Error Analysis and Fatality Prediction in Maritime Accidents. *Journal of Marine Science and Engineering*, *11*(12), 2287.

Mina, G., Bonadonna, A., Peira, G., & Beltramo, R. (2024). How to improve the attractiveness of e-bikes for consumers: Insights from a systematic review. *Journal of Cleaner Production*, *442*, 140957.

Nematchoua, M., Deuse, C., Cools, M., & Reiter, S. (2020). Evaluation of the potential of classic and electric bicycle commuting as an impetus for the transition towards environmentally sustainable cities: A case study of the university campuses in Liege, Belgium. *Renewable and Sustainable Energy Reviews*, *119*, 109544.

Park, S., Cho, J., Park, K., & Shin, H. (2021). Customer sentiment analysis with more sensibility. *Engineering Applications of Artificial Intelligence*, *104*, 104356.

Patra, S. (2019). Questionnaire Design. In R. N. Subudhi & S. Mishra (Eds.), *Methodological Issues in Management Research: Advances, Challenges, and the Way Ahead* (pp. 53–78). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78973-973-220191005

Pitka, T., Bucko, J., Krajci, S., Kridlo, O., Gunis, J., Snajder, L., Antoni, L., & Elias, P. (2024). Time analysis of online consumer behavior by decision trees, GUHA association rules, and formal concept analysis. *Journal of Marketing Analytics*.

Settey, T., Gnap, J., Benova, D., Pavlicko, M., & Blazekova, O. (2021). The Growth of E-Commerce Due to COVID-19 and the Need for Urban Logistics Centers Using Electric Vehicles: Bratislava Case Study. *Sustainability*, *13*(10), 5357.

Taheri, M., Xie, F., & Lederer, J. (2021). Statistical guarantees for regularised neural networks. *Neural Networks*, *142*, 148–161.

Terek, M., (2019). *Dotaznikove prieskumy a analyzy ziskanych dat.* Equilibria.

Ulrich, P., Ramzy, N., & Ratusny, M. (2023). Semantic Context Information Modeling With Neural Networks in Customer Order Behavior Classification. *IEEE Transactions on Semiconductor Manufacturing*, *36*(4), 570–577.

Vrtana, D., & Krizanova, A. (2023). The Power of Emotional Advertising Appeals: Examining Their Influence on Consumer Purchasing Behavior and Brand–Customer Relationship. *Sustainability*, *15*(18), 13337.

Oracle. (n. d.) What is Customer Loyalty? https://www.oracle.com/cx/marketing/customer-loyalty/what-is-customer-loyalty/

Xia, X., Jiang, H., & Wang, J. (2022). Analysis of user satisfaction of shared bicycles based on SEM. *Journal of Ambient Intelligence and Humanized Computing*, *13*(3), 1587–1601.

Annex

Table A1: Logistic regression model

Variable	B	S.E.	Wald	Deg. of freedom	Sig.	Exp. B
price			11.310	8	0.185	
price(2)	1.158	0.657	3.111	1	0.078	3.184
price(3)	1.671	0.713	5.489	1	0.019	5.320
price(4)	2.051	0.792	6.698	1	0.010	7.773
price(5)	1.749	0.818	4.576	1	0.032	5.751
price(6)	0.896	0.960	0.871	1	0.351	2.450
price(7)	-0.420	1.117	0.142	1	0.707	0.657
price(8)	-18.203	8916.528	0.000	1	0.998	0.000
price(9)	1.206	0.863	1.952	1	0.162	3.339
accessories			2.674	4	0.614	
accessories (1)	-21.368	4336.553	0.000	1	0.996	0.000
accessories(2)	0.525	0.339	2.393	1	0.122	1.691
accessories(3)	0.177	0.325	0.296	1	0.587	1.193
accessories(4)	0.509	0.638	0.637	1	0.425	1.664
battery life			36.583	4	< 0.05	
battery life(1)	4.411	0.794	30.825	1	< 0.05	82.346
battery life(2)	1.226	0.325	14.268	1	< 0.05	3.407
battery life(3)	2.083	9873.947	0.000	1	0.998	8.028
battery life(4)	1.367	0.502	7.415	1	0.006	3.925
engine brand			59.452	9	< 0.05	
engine brand(1)	1.072	0.459	5.462	1	0.019	2.921
engine brand(2)	-0.514	0.484	1.127	1	0.288	0.598
engine brand(3)	-0.034	0.771	0.002	1	0.965	0.967
engine brand(4)	2.637	1.086	5.895	1	0.015	13.976
engine brand(5)	1.872	0.626	8.929	1	0.003	6.499
engine brand(6)	2.559	0.536	22.774	1	< 0.05	12.920
engine brand(7)	-0.653	0.528	1.528	1	0.216	0.520
engine brand(8)	2.377	0.853	7.771	1	0.005	10.778
engine brand(9)	1.184	0.668	3.140	1	0.076	3.268
battery capacity			38.658	4	< 0.05	
battery capacity(1)	22.819	9722.597	0.000	1	0.998	8,131,269,445.733
battery capacity(2)	21.199	9722.597	0.000	1	0.998	1,609,065,12.518
battery capacity(3)	19.451	9722.597	0.000	1	0.998	280,297,192.576
battery capacity(4)	20.212	9722.597	0.000	1	0.998	599,494,451.920
purchase place			14.126	2	0.001	
purchase place(1)	0.924	0.307	9.047	1	0.003	2.519
purchase place(2)	2.633	0.796	10.947	1	0.001	13.917
Constant	-24.913	9722.597	.000	1	0.998	0.000

Source: own elaboration