


Interpretable machine learning models for customer satisfaction: A comparison of decision trees in the coffee shop context

Veronika Labosova^{1*} , Sebastian Tomczak²  and Klara Krupalova¹

* Correspondence: veronika.labosova@stud.uniza.sk

¹University of Zilina, Faculty of Operation and Economics of Transport and Communications, Department of Economics, Univerzitna 1, 010 26 Zilina, Slovakia; krupalova@stud.uniza.sk

²Wroclaw University of Science and Technology, Faculty of Management, Department of Operations Research and Business Intelligence, 27Wybrzeze Stanislaw Wyspianskiego st. 50-370 Wroclaw, Poland; sebastian.tomczak@pwr.edu.pl

Abstract

Research background: Customer satisfaction is a critical determinant of competitive performance in service markets, and the coffee shop segment is highly competitive, and experience driven. Companies increasingly predict satisfaction from survey data using machine learning, yet the most accurate models are often opaque, while managers need insight they can understand and act on.

Purpose of the article: The purpose is to classify overall customer satisfaction among coffee shop customers using interpretable models and to identify the service attributes that most strongly distinguish satisfied from dissatisfied customers. The study draws on the view that satisfaction reflects the perceived performance of the service rather than customer characteristics.

Methods: The study uses primary data gathered through an online questionnaire completed by 400 customers of a global coffeehouse chain between October 2025 and February 2026. Four decision tree algorithms were compared, the training set was balanced by oversampling the minority group, and the models were evaluated on a separate test set using accuracy and sensitivity. The most accurate model was interpreted through the importance of its predictors and the structure of its tree.

Findings & Value added: All four models classified satisfaction with accuracy above 93%, and the best model reached almost 95% while reliably identifying dissatisfied customers. Satisfaction was driven entirely by assessments of service attributes rather than customer characteristics, and the cleanliness and appearance of the interior was by far the strongest driver, followed by product quality. The study shows that interpretable models can achieve high accuracy while offering managers actionable guidance on satisfaction drivers.

Keywords: customer satisfaction; interpretable machine learning; decision trees; classification

JEL Classification: M31; C38; L83

Received

15 November 2026

Received in revised form

27 March 2026

Accepted

12 June 2026

Available online

30 June 2026

Cite as: Labosova, V., Tomczak, S., & Krupalova, K. (2026). Interpretable machine learning models for customer satisfaction: A comparison of decision trees in the coffee shop context, *Ekonomicko-manazerske spektrum*, 20(1), 129-142.



1. Introduction

Customer satisfaction has become one of the most closely monitored indicators of competitive performance in service and retail markets, where retaining an existing customer is considerably less costly than attracting a new one. According to the expectation disconfirmation paradigm, satisfaction arises from the comparison that customers draw between the performance they expected and the performance they perceive, so that confirmed or exceeded expectations produce satisfaction while unmet expectations produce dissatisfaction (Oliver, 1980). The managerial importance of this construct is well established, as more than a quarter of a century of empirical research links higher satisfaction to stronger firm performance and profitability (Otto et al., 2020). Research synthesising fifteen years of loyalty studies further confirmed that satisfaction remains among the most influential antecedents of customer loyalty across a wide range of service settings (So et al., 2025).

The coffee shop segment illustrates these dynamics with clarity, because it is a highly competitive market in which firms compete less on the product alone and more on the overall experience. In the speciality coffee café context, customer experience has been shown to translate directly into satisfaction, advocacy and repurchase intentions (Meeprom and Kokkhangplu, 2025), and neuromarketing evidence indicates that the experience inside the store strongly shapes how customers evaluate coffee retailers. Within this setting, satisfaction is rarely driven by a single attribute. Research based on surveys repeatedly identifies the quality of the products, the price and the perceived value, the atmosphere of the premises, the attitude and professionalism of the staff, the speed of service and the accessibility of the outlet as the leading determinants of café satisfaction (Nor et al., 2024). Because these attributes operate together, understanding which of them most strongly separates satisfied from dissatisfied customers is of direct practical value to managers who must allocate limited resources.

Beyond explaining satisfaction, firms increasingly seek to predict it from survey and transactional data to support decisions that are guided by evidence. Machine learning has proven effective for this purpose, and a recent benchmarking study reported that ensemble models such as random forests can classify satisfied customers with an accuracy exceeding ninety percent (Zaghloul et al., 2024). Much of the related work, however, has addressed the downstream problem of customer churn rather than satisfaction itself, and it has done so primarily in the telecommunications, banking and online retail sectors. Usman-Hamza et al. (2024) evaluated several classifiers based on decision trees for churn prediction in telecommunications, and Asif et al. (2025) combined machine learning with explainable artificial intelligence to predict churn, although the explanations were generated only after the model had been trained. In a similar vein, Poudel et al. (2024) explained churn predictions produced by tabular models and Peng et al. (2023) paired churn prediction with an analysis of model interpretability, while Zhang and Zhang (2024) linked churn prediction with satisfaction analysis within customer relationship management. Studies that target satisfaction or loyalty more directly remain comparatively scarce and tend to lie outside the hospitality setting, as illustrated by Azad et al. (2023), who predicted consumer purchase decisions from behavioural data, Amalia et al. (2022), who compared classifiers for airline customer satisfaction, and Hamdan and Othman (2022), who predicted hotel customer loyalty. Across this literature, the drivers of satisfaction in the coffee shop segment, where the encounter is strongly experiential, have received little attention, and predictive accuracy has frequently been prioritised over models whose reasoning is transparent.

A recurring concern across this literature is that the most accurate algorithms are frequently opaque models whose internal logic is difficult for managers to scrutinise. For decisions that inform strategy, the capacity to explain why a model reaches a conclusion is as important as predictive accuracy, and reliance on unexplained black box models in high stakes settings has been strongly criticised (Rudin, 2019). Two broad responses have emerged. One relies on post hoc techniques

that approximate the behaviour of an opaque model after it has been trained, an approach whose explanations have been questioned because they may not faithfully reflect the underlying model (Zschech et al., 2025). The other favours models that are interpretable by their own design, and recent evidence dispels the assumption that such transparency must be sacrificed for accuracy, showing that interpretable models can match the performance of opaque ones on tabular data (Kruschel et al., 2025). The marketing evidence on explanation is itself mixed. Post hoc explanations have been found to improve consumer responses to algorithmic decisions (Mourali et al., 2024; Kicova and Tislerova, 2025) and explainable approaches are increasingly used to infer consumer preferences from behavioural data (Zhou et al., 2023), yet recent work cautions that explanation is not universally beneficial and that imposing it indiscriminately may even produce unintended consequences for consumers (Mohammadi et al., 2025). For management purposes, the value of a transparent model lies less in raw accuracy than in the actionable and trustworthy insight it provides, because a model whose rules can be read directly allows managers to understand and act on the factors that shape satisfaction.

Despite the managerial appeal of transparent models and the growing number of machine learning applications in this area, comparative studies that apply several interpretable classifiers to customer satisfaction data and then interpret the resulting model for practical use remain limited, particularly in the coffee shop context. The aim of this study is therefore to classify overall customer satisfaction among coffee shop customers using interpretable models, to compare their classification performance, and to interpret the most accurate model to identify the factors that most strongly distinguish satisfied from dissatisfied customers. By combining a comparison of predictive performance with a managerial interpretation of the selected model, the study contributes to the literature on interpretable machine learning in marketing and offers practical insight into the drivers of satisfaction in a service and retail setting.

The remainder of the paper is organised as follows. Section 2 describes the data, the survey instrument and the methodological procedure, and specifies the research goal and the hypotheses. Section 3 reports the classification results and compares the developed models. Section 4 discusses the findings considering prior research, and Section 5 concludes by summarising the contribution, acknowledging the limitations of the study and outlining directions for future research.

2. Methodology

The goal of the study is to classify overall customer satisfaction among Starbucks customers as either satisfied or dissatisfied on the basis of their characteristics and their assessments of individual service attributes, to compare the classification performance of several interpretable models, and to interpret the most accurate model in order to identify the attributes that most strongly distinguish satisfied from dissatisfied customers. In line with this goal, the study tests two hypotheses. The first hypothesis (H1) holds that the customers' assessments of the service attributes are more important predictors of overall satisfaction than their demographic and socioeconomic characteristics. The second hypothesis (H2) holds that a small number of attributes accounts for most of the power to distinguish satisfied from dissatisfied customers, so that satisfaction can be explained by a parsimonious set of key drivers.

2.1. Data and variables

The study uses primary data collected through an online questionnaire created in Google Forms. The questionnaire was distributed through social media, specifically Instagram and Facebook, and respondents took part voluntarily, so the resulting sample is a convenience sample rather than a probability sample. Data collection ran from 1 October 2025 to 25 February 2026. The questionnaire was designed by the author for the purpose of this study. It comprised 23 items, of which the first was a dichotomous screening question that asked whether the respondent had ever purchased at

Starbucks, while the remaining items were closed questions in which each respondent selected a single answer from a fixed set of options, and all items were mandatory.

The target population consists of customers of Starbucks. Respondents who indicated at the screening question that they had never purchased at Starbucks were not eligible and did not continue to the remaining items. Of the 431 respondents who entered the questionnaire, 31 were screened out at this first question, which left 400 eligible respondents who had purchased at Starbucks and completed the full questionnaire. Because the questionnaire consisted of closed and mandatory questions, the data contained no missing values, and no respondents were removed beyond the screening stage, so all 400 eligible responses entered the analysis without further adjustment.

The overall satisfaction item served as the binary target variable, with the two classes corresponding to satisfied and dissatisfied customers. The remaining items served as predictors and covered two groups of information. The first group described the respondents and their relationship to the brand, and the second group captured their assessments of individual service attributes. The predictors were entered as nominal variables and the target as a binary flag. Table 1 presents the full set of variables together with their response categories.

2.2. Analytical procedure

The classification task was carried out in IBM SPSS Modeler 18.0 using four decision tree algorithms that are widely used and valued for the transparency of their output, namely the classification and regression tree CART (Breiman et al., 1984), the C5.0 algorithm (Quinlan, 1986), the chi-square automatic interaction detector CHAID (Kass, 1980) and the quick unbiased efficient statistical tree QUEST (Loh and Shih, 1997). These algorithms were selected because they produce models in the form of explicit decision rules that managers can read and act on, which is central to the aim of the study, and because they differ in the way they construct trees, which makes a comparison informative. CART builds binary trees using the Gini impurity measure, C5.0 builds trees using an entropy-based information gain criterion, CHAID uses Chi-square tests of independence and allows multiway splits, and QUEST separates the selection of the splitting variable from the selection of the split point in order to reduce the bias towards variables with many categories.

The data were partitioned into a training set and a test set in a 70/30 ratio. Because the two satisfaction classes were not equally represented, the training partition was balanced by random oversampling, which replicates cases from the minority class, in order to prevent the models from being dominated by the majority class (He and Garcia, 2009), while the test partition was left in its original distribution so that the reported performance reflects realistic conditions. The distribution of the sample across the two partitions is reported in Table 2. Each of the four algorithms was trained on the balanced training partition and evaluated on the untouched test partition.

Model performance was evaluated on the test partition using five measures. The classification of each case was summarised in a confusion matrix of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), where the dissatisfied class was treated as the positive class because the identification of dissatisfied customers is of primary managerial interest. Overall accuracy expresses the proportion of correctly classified cases and is defined as:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

Sensitivity (Recall), also referred to as the true positive rate, expresses the proportion of dissatisfied customers that the model correctly identifies and is defined as:

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (2)$$

Table 1: Variables, measurement level and response categories

Variable	Variable distribution (n, %)	Variable	Variable distribution (n, %)
Gender	Female (257; 64.3%), Male (143; 35.8%)	Product quality rating	Very poor (37; 9.3%), Poor (79; 19.8%), Good (163; 40.8%), Excellent (121; 30.3%)
Generation group	Baby boomers (1946–1964) (8; 2.0%), Generation X (1965–1980) (131; 32.8%), Generation Y (1981–1996) (82; 20.5%), Generation Z (1997–2007) (179; 44.8%)	Price perception	Unreasonably high (84; 21.0%), Too high (116; 29.0%), Higher but justified (136; 34.0%), Reasonable (64; 16.0%)
Region	Banska Bystrica (27; 6.8%), Bratislava (32; 8.0%), Kosice (26; 6.5%), Nitra (24; 6.0%), Presov (24; 6.0%), Trencin (34; 8.5%), Trnava (25; 6.3%), Zilina (208; 52.0%)	Interior cleanliness and appearance	Very negative (35; 8.8%), Rather negative (67; 16.8%), Rather positive (136; 34.0%), Very positive (162; 40.5%)
Highest level of education	Secondary education (208; 52.0%), Bachelor's degree (110; 27.5%), Master's degree (82; 20.5%)	Staff attitude	Very negative (38; 9.5%), Rather negative (49; 12.3%), Rather positive (135; 33.8%), Very positive (178; 44.5%)
Economic status	Employed (188; 47.0%), Student (147; 36.8%), Self-employed (51; 12.8%), Unemployed (8; 2.0%), Retired (6; 1.5%)	Service speed	Very slow (39; 9.8%), Rather slow (64; 16.0%), Rather fast (171; 42.8%), Very fast (126; 31.5%)
Net monthly income	Up to 500€ (84; 21.0%), 500€–1,000€ (72; 18.0%), 1,001€–1,500€ (86; 21.5%), 1,501€–2,000€ (63; 15.8%), 2,001€–2,500€ (31; 7.8%), 2,501€ and more (8; 2.0%), No income (37; 9.3%), I don't want to specify (19; 4.8%)	Product range rating	Very negative (33; 8.3%), Rather negative (94; 23.5%), Rather positive (167; 41.8%), Very positive (106; 26.5%)
First contact/source of awareness	Advertising on social media (162; 40.5%), Advertising on websites (95; 23.8%), Personal recommendation from an acquaintance (138; 34.5%), Television marketing campaign/advertising (5; 1.3%)	Quality problem experience	Yes, several times (44; 11.0%), Yes, rarely (96; 24.0%), No, hardly ever (152; 38.0%), No, never (108; 27.0%)
Visit frequency	Less than once a month (255; 63.7%), Once a month (84; 21.0%), Once a week (37; 9.3%), Several times a week (24; 6.0%)	Overall atmosphere	Very unpleasant (35; 8.8%), Rather unpleasant (69; 17.3%), Rather pleasant (176; 44.0%), Very pleasant (120; 30.0%)
Purchase channel	In person (297; 74.3%), via Bolt (61; 15.3%), via Wolt (42; 10.5%)	Store accessibility satisfaction	No outlet nearby (55; 13.8%), Too far (84; 21.0%), Available but not close (145; 36.3%), Available and close (116; 29.0%)
Most influential factor	Product quality (100; 25.0%), Product range (132; 33.0%), Product price (31; 7.8%), Store atmosphere (67; 16.8%), Store location/accessibility (44; 11.0%), Speed and quality of service (26; 6.5%)	Product availability problem	Always unavailable (16; 4.0%), Mostly unavailable (55; 13.8%), Mostly available (202; 50.5%), Always available (127; 31.8%)
		Environmental and social initiatives	Very negative (35; 8.8%), Rather negative (83; 20.8%), Rather positive (167; 41.8%), Very positive (115; 28.7%)

Source: own elaboration

Table 2: Distribution of the sample across the training and test partitions

Sample	Full		Training (oversampled)		Testing	
	absolute frequency	relative frequency [%]	absolute frequency	relative frequency [%]	absolute frequency	relative frequency [%]
Satisfied	283	70.8	192	50.0	91	70.0
Dissatisfied	117	29.2	192	50.0	39	30.0
Total	400	-	384	-	130	-

Source: own elaboration

Precision expresses the proportion of the customers predicted as dissatisfied that were indeed dissatisfied and is defined as:

$$Precision = \frac{TP}{(TP + FP)} \tag{3}$$

The F1 score combines precision and sensitivity into a single value as their harmonic mean, which is informative when the classes are imbalanced, and is defined as:

$$F1\ score = \frac{2TP}{(2TP + FP + FN)} \quad (4)$$

The area under the ROC curve (AUC) summarises the ability of a model to separate the two classes across all classification thresholds. The ROC curve plots the true positive rate against the false positive rate. The AUC takes a value of one for a perfect classifier and a value of 0.5 for a classifier that performs no better than chance (Fawcett, 2006). Accuracy, sensitivity, precision and the F1 score were interpreted in line with the established conventions for classification tasks (Svabova et al., 2024; Sokolova and Lapalme, 2009).

The model with the strongest performance on the test partition was then selected, and its tree structure together with the relative importance of its predictors was examined to interpret the rules and identify the attributes that most strongly distinguish satisfied from dissatisfied customers.

3. Results

The four decision tree algorithms were trained on the balanced training partition and evaluated on the test partition of 130 customers, of whom 91 were satisfied and 39 were dissatisfied. Table 3 reports the accuracy, precision, sensitivity, F1 score and area under the ROC curve of each model.

Table 3: Classification performance of the decision tree models

Modelling Technique	Accuracy (%)	Precision (%)	Sensitivity (%)	F1 score (%)	AUC
CART	93.85	91.89	87.18	89.47	0.90
C5.0	94.62	92.11	89.74	90.91	0.95
CHAID	93.08	91.67	84.62	88.00	0.94
QUEST	93.08	89.47	87.18	88.31	0.89

Source: own elaboration

All four models classified the customers with an accuracy above 93 percent, which clearly exceeds the baseline accuracy of 70.0 percent that would be obtained by always predicting the more frequent satisfied class, and all four reached a sensitivity above 84 percent, which indicates that the models identified the smaller dissatisfied group reliably rather than simply favouring the majority class. The C5.0 algorithm achieved the strongest performance on every measure, with an accuracy of 94.62 percent, a precision of 92.11 percent, a sensitivity of 89.74 percent, an F1 score of 90.91 percent and the highest area under the ROC curve of 0.95, correctly identifying 35 of the 39 dissatisfied customers while misclassifying only three satisfied customers. CART followed with an accuracy of 93.85 percent, an F1 score of 89.47 percent and an AUC of 0.90, while CHAID reached an accuracy of 93.08 percent, an F1 score of 88.00 percent and an AUC of 0.94, and QUEST reached an accuracy of 93.08 percent, an F1 score of 88.31 percent and an AUC of 0.89. Because the four models performed very similarly in accuracy, the model for interpretation was not selected on that basis alone. C5.0 achieved the best result on all five measures, and as one of the most widely used decision tree algorithms it was selected for interpretation.

To assess whether the results depend on the choice of algorithm, the four models were also compared on the predictors they relied on, which are summarised in Table 4.

The assessment of interior cleanliness and appearance was the most important predictor in every model, and the same small group of service attributes occupied the leading positions throughout, although their order varied slightly. The dominance of the interior was especially pronounced in C5.0, CHAID and QUEST, where it accounted for around half of the importance, and was less pronounced in CART, where importance was spread more evenly across predictors. A few socioeconomic or behavioural variables, such as region, net monthly income or the purchase

Table 4: Top five predictors of each model by importance

Rank	CART	C5.0	CHAID	QUEST
1	Interior cleanliness (0.19)	Interior cleanliness (0.50)	Interior cleanliness (0.45)	Interior cleanliness (0.53)
2	Product quality (0.15)	Product quality (0.19)	Service speed (0.17)	Product quality (0.24)
3	Price perception (0.13)	Price perception (0.13)	Price perception (0.13)	Purchase channel (0.03)
4	Staff attitude (0.12)	Service speed (0.11)	Product quality (0.09)	Net income (0.03)
5	Region (0.05)	Staff attitude (0.07)	Product range (0.05)	Price perception (0.03)

Source: own elaboration

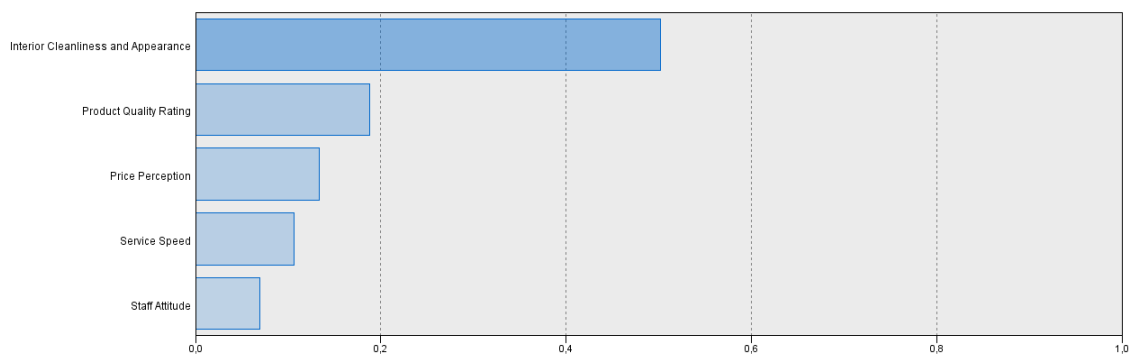
channel, entered the CART and QUEST models, but always with importance far below that of the service attributes. The importance values are normalised within each model and sum to one, so the columns should be compared by the ranking of the predictors rather than by their absolute values. The leading role of the service attributes, and of interior cleanliness, is therefore consistent across the four algorithms rather than an artefact of any single model.

Because the four models performed very similarly in accuracy and identified the same leading predictors, the model for interpretation was not selected based on the small differences in accuracy alone. C5.0 achieved the best result on all five performance measures, and as one of the most widely used decision tree algorithms it was selected for interpretation.

3.1. Interpretation of the best-performing model

The C5.0 model was interpreted through the relative importance of its predictors and through the structure of the resulting tree. Figure 1 presents the predictor importance. The model retained only five predictors, and all of them were assessments of service attributes rather than characteristics of the respondents. The assessment of interior cleanliness and appearance was by far the most important predictor, with an importance of 0.50, accounting for half of the total importance. It was followed by the product quality rating with an importance of 0.19, the price perception with 0.13, the service speed with 0.11 and the staff attitude with 0.07. None of the demographic or socioeconomic variables, such as gender, generation, education or income, entered the model.

Figure 1: Predictor importance for the C5.0 model

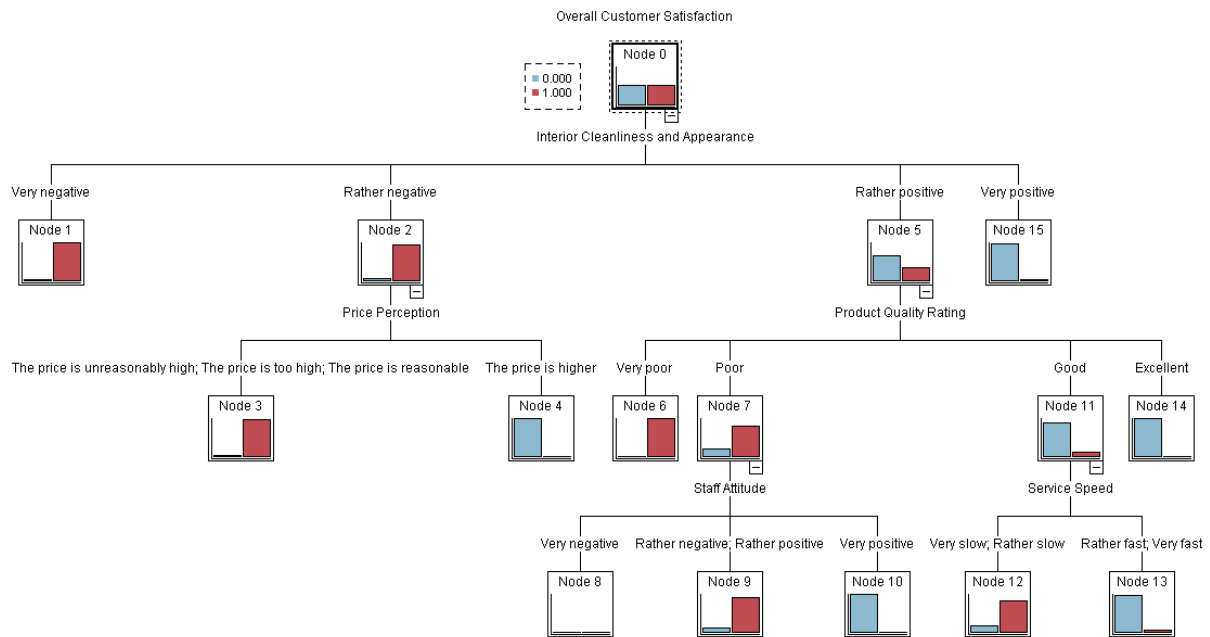


Source: own elaboration

The structure of the tree, which is presented in Figure 2, translates this ranking into a set of explicit decision rules.

The root node split the customers according to their assessment of interior cleanliness and appearance. Customers who rated the interior as very positive were classified as satisfied, and those who rated it as very negative were classified as dissatisfied, regardless of any other attribute. For the intermediate assessments, the model relied on further attributes. Where the interior was rated as rather negative, the price perception decided the outcome, so that customers were classified as satisfied only when they considered the price higher but justified, whereas those who viewed it as unreasonable, too high or merely reasonable were classified as dissatisfied. Where the

Figure 2: Decision tree of the C5.0 model



Source: own elaboration

interior was rated as rather positive, the product quality rating decided the outcome, with an excellent rating leading to satisfaction and a very poor rating leading to dissatisfaction, while intermediate ratings were resolved by the staff attitude and the service speed at the lower levels of the tree. The resulting rules are summarised in Table 5.

Table 5: Classification rules derived from the C5.0 model

Rule	Conditions	Predicted class
1	Interior very negative	Dissatisfied
2	Interior rather negative and price unreasonably high or too high or reasonable	Dissatisfied
3	Interior rather negative and price higher but justified	Satisfied
4	Interior rather positive and product quality very poor	Dissatisfied
5	Interior rather positive, product quality poor and staff attitude rather negative and positive	Dissatisfied
6	Interior rather positive, product quality poor and staff attitude positive	Satisfied
7	Interior rather positive, product quality good and service speed slow	Dissatisfied
8	Interior rather positive, product quality good and service speed fast	Satisfied
9	Interior rather positive and product quality excellent	Satisfied
10	Interior very positive	Satisfied

Source: own elaboration

In broad terms, the most satisfied customers were those who rated the interior highly, whereas dissatisfaction was concentrated among customers who assessed the interior unfavourably and were not compensated by a favourable view of the price, the product quality, the staff or the speed of service.

These findings allow the two hypotheses to be evaluated. The first hypothesis (H1) is supported because every predictor retained by the model was an assessment of a service attribute, while no demographic or socioeconomic characteristic contributed to the classification. The second hypothesis (H2) is also supported, because a small number of attributes accounted for most of the discriminatory power, with the assessment of interior cleanliness alone representing half of the total importance and the two leading attributes together representing roughly seventy percent of it.

4. Discussion

The study set out to classify customer satisfaction with interpretable models and to interpret the most accurate one. The four decision tree algorithms classified satisfaction with an accuracy above 93%, and the best model, C5.0, reached an accuracy of almost 95% and a high area under the ROC curve while remaining fully transparent. Because the satisfied and dissatisfied groups were imbalanced, the model was also evaluated on precision and the F1 score, and its strong values on both, above 90% in each case, confirm that it identified the smaller dissatisfied group reliably rather than at the expense of many false alarms. This performance is comparable to that reported for opaque ensemble models in related settings, for example the accuracy above 90% that random forests achieved in the prediction of e-commerce satisfaction (Zaghloul et al., 2024). The finding therefore supports the argument that interpretability need not be sacrificed for predictive accuracy (Kruschel et al., 2025) and that, for decisions that managers must understand and justify, models that are transparent by their own design are a credible alternative to opaque models that require explanation after the fact (Rudin, 2019; Zscheck et al., 2025).

The interpretation of the C5.0 model produced a clear and theoretically coherent picture. Every predictor retained by the model was an assessment of a service attribute, while no demographic or socioeconomic characteristic of the respondents contributed to the classification. This is consistent with the expectation disconfirmation view of satisfaction, in which satisfaction follows from the perceived performance of the product or service rather than from who the customer is (Oliver, 1980), and it reinforces the long standing evidence that perceived service performance is the principal antecedent of satisfaction and, through it, of loyalty and firm performance (Otto et al., 2020; So et al., 2025). For the coffee shop context, it suggests that satisfaction is shaped by what customers experience during the visit rather than by their background.

The most striking result is the dominance of a single attribute. The assessment of interior cleanliness and appearance alone accounted for half of the predictive importance and formed the first split of the tree, with product quality, price perception, service speed and staff attitude playing progressively smaller roles. That interior cleanliness was the most important predictor in all four algorithms, and that the same service attributes led throughout, indicates that this pattern is robust to the choice of model rather than specific to C5.0. This leading role of the physical environment aligns closely with the servicescape tradition, which holds that the physical surroundings in which a service is delivered shape how customers feel and respond (Bitner, 1992). Recent evidence reinforces this view, as a meta-analysis of servicescape research confirmed that the physical environment is a robust determinant of consumer satisfaction across service settings (Wang et al., 2024), and studies of store design and atmosphere have linked the appearance of the premises to customer evaluations and engagement (Khan et al., 2022; Nong et al., 2025). The finding is also consistent with the experiential character of coffee consumption, in which the overall encounter and its setting shape how customers evaluate the visit (Meeprom and Kokkhangplu, 2025), and with evidence that the environment of a global coffeehouse chain contributes to how it is perceived in a competitive market (Chen et al., 2022). At the same time, the result adds nuance to the determinant literature, which tends to place product quality at the centre of café satisfaction (Nor et al., 2024), because in the present data, the cleanliness and appearance of the interior outranked the quality of the products themselves. The concentration of importance in a few attributes also means that satisfaction can be explained parsimoniously rather than by a large set of equally weighted factors.

These results carry direct managerial implications. Because the drivers of satisfaction are service attributes that managers can influence rather than fixed customer characteristics, interventions aimed at raising satisfaction can be directed at the experience itself, and because a small number of attributes dominates, resources can be concentrated where they matter most. The findings suggest that maintaining a clean and attractive interior, followed by consistent product quality,

offers the greatest leverage for moving customers from the dissatisfied to the satisfied group, whereas attributes such as staff attitude, although still relevant, contributed less to the distinction. The transparency of the model is what makes this guidance possible, because the decision rules can be read directly and translated into priorities, which is the practical advantage that motivated the use of interpretable models in the first place.

5. Conclusions

The study set out to classify overall customer satisfaction among coffee shop customers using interpretable models and to interpret the most accurate model to identify the attributes that most strongly distinguish satisfied from dissatisfied customers. Four decision tree algorithms were compared on a sample of 400 customers, and all of them classified satisfaction with an accuracy above 93%. The C5.0 algorithm performed best, reaching an accuracy of almost 95%, a sensitivity of almost 90%, and a precision and an F1 score both above 90%, together with a high area under the ROC curve, while remaining fully transparent. The interpretation of this model showed that satisfaction was driven entirely by assessments of service attributes rather than by the characteristics of the respondents, and that a single attribute, the cleanliness and appearance of the interior, accounted for half of the predictive importance, followed by product quality, price perception, service speed and staff attitude. Both hypotheses were therefore supported, since service attributes outweighed demographic and socioeconomic characteristics, and a small set of drivers explained most of the distinction between satisfied and dissatisfied customers.

The study makes a methodological and a practical contribution. Methodologically, it shows that interpretable decision tree models can classify customer satisfaction with an accuracy comparable to that of opaque models while producing decision rules that managers can read and act on, which supports the use of models that are transparent by their own design in customer satisfaction analytics. In practical terms, the results indicate that the physical environment of the outlet, together with consistent product quality, offers the greatest leverage for improving satisfaction, and that because the drivers are attributes under managerial control rather than fixed customer characteristics, efforts to raise satisfaction can be directed at the experience itself and concentrated where they matter most.

The study has several limitations that should be borne in mind when interpreting its findings. The data were gathered through a convenience sample recruited on social media, so the respondents are not necessarily representative of the wider population of customers, and the sample was collected at a single point in time and was dominated by younger respondents and by one region. The study also focused on a single coffeehouse brand in one country, which limits the generalisability of the results to other brands and markets. From a methodological standpoint, single decision trees can be sensitive to small changes in the data, and the use of oversampling to balance the training partition, while it improved the recognition of the smaller dissatisfied group, can increase the risk of overfitting to replicated cases. Finally, the analysis relied on self-reported assessments collected through closed questions, which capture perceptions rather than observed behaviour.

These limitations suggest several directions for future research. The findings could be tested on a larger and more representative sample drawn through probability sampling, and across several coffeehouse brands and countries, to establish how far the dominant role of the physical environment generalises. Longitudinal data would make it possible to examine whether the drivers of satisfaction remain stable over time. Future work could also compare the interpretable models used here with ensemble models combined with post hoc explanation, to weigh the trade-off between accuracy and transparency more fully and could enrich the set of attributes with additional dimensions of the customer experience. Confirming the leading role of interior cleanliness and appearance in other settings would be a particularly valuable extension, given how strongly it emerged in the present study.

Author contributions

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

Funding

This research received no external funding.

Data Availability Statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions related to the questionnaire survey.

Acknowledgments

This research was supported by the project Integrating machine learning algorithms into the customer satisfaction evaluation process to improve business competitiveness, SAP element I-25-028-26, within the internal grant scheme of the University of Žilina.

Conflicts of Interest

The authors declare no conflict of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors declare that no generative AI or AI-assisted technologies were used in the writing or preparation of this manuscript.

References

- Amalia, S., Deborah, I., & Yulita, I. N. (2022). Comparative analysis of classification algorithm: Random Forest, SPAARC, and MLP for airlines customer satisfaction. *SINERGI*, 26(2), 213–222. <https://doi.org/10.22441/sinergi.2022.2.010>
- Asif, D., Arif, M. S., & Mukheimer, A. (2025). A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry. *Results in Engineering*, 26, 104629. <https://doi.org/10.1016/j.rineng.2025.104629>
- Azad, M. S., Khan, S. S., Hossain, R., Rahman, R., & Momen, S. (2023). Predictive modeling of consumer purchase behavior on social media: Integrating theory of planned behavior and machine learning for actionable insights. *PLoS ONE*, 18(12), e0296336. <https://doi.org/10.1371/journal.pone.0296336>
- Bitner, M. J. (1992). Servicescapes: the impact of physical surroundings on customers and employees. *Journal of Marketing*, 56(2), 57–71. <https://doi.org/10.1177/002224299205600205>
- Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (1984). *Classification and Regression Trees*. Chapman and Hall/CRC: Wadsworth, NY, USA.
- Chen, X., Lee, T. J., & Hyun, S. S. (2022). How does a global coffeehouse chain operate strategically in a traditional tea-drinking country? The influence of brand authenticity and self-enhancement. *Journal of Hospitality and Tourism Management*, 51, 176–186. <https://doi.org/10.1016/j.jhtm.2022.03.003>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Hamdan, I. Z. P., & Othman, M. (2022). Predicting customer loyalty using machine learning for hotel industry. *Journal of Soft Computing and Data Mining*, 3(2), 31–42. <https://doi.org/10.30880/jscdm.2022.03.02.004>

- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284. <https://doi.org/10.1109/TKDE.2008.239>
- Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society*, 29(2), 119–127. <https://doi.org/10.2307/2986296><https://www.jstor.org/stable/2986296>
- Khan, M. A., Vivek, Minhaj, S. M., Saifi, M. A., Alam, S., & Hasan, A. (2022). Impact of Store Design and Atmosphere on Shoppers' Purchase Decisions: An Empirical Study with Special Reference to Delhi-NCR. *Sustainability*, 15(1), 95. <https://doi.org/10.3390/su15010095>
- Kicova, E., & Tislerova, K. (2025). The Relationship Between Financial Literacy and Consumer Purchasing Behaviour in Slovakia. *Ekonomicko-manazerske spektrum*, 19(2), 87-96. <https://doi.org/10.26552/ems.2025.2.87-96>
- Kruschel, S., Hambauer, N., Weinzierl, S., Zilker, S., Kraus, M., & Zschech, P. (2025). Challenging the Performance-Interpretability Trade-Off: An evaluation of Interpretable machine learning models. *Business & Information Systems Engineering*, 68(1), 159–183. <https://doi.org/10.1007/s12599-024-00922-2>
- Loh, W. Y., & Shih, Y. S. (1997). Split selection methods for classification trees. *Statistica Sinica*, 7(4), 815–840.
- Meeprom, S., & Kokkhangplu, A. (2025). Customer experience and satisfaction in coffee consumption: an experiential marketing perspective. *Cogent Business & Management*, 12(1), 2450296. <https://doi.org/10.1080/23311975.2025.2450296>
- Mohammadi, B., Malik, N., Derdenger, T., & Srinivasan, K. (2025). Regulating explainable artificial intelligence (XAI) may harm consumers. *Marketing Science*, 44(3), 711–724. <https://doi.org/10.1287/mksc.2022.0396>
- Mourali, M., Novakowski, D., Pogacar, R., & Brigden, N. (2024). Post hoc explanations improve consumer responses to algorithmic decisions. *Journal of Business Research*, 186, 114981. <https://doi.org/10.1016/j.jbusres.2024.114981>
- Nong, S. Z., Fong, L. H. N., Albayrak, T., & Caber, M. (2025). Generation of customer engagement by hotel service atmosphere: The mediating role of flow experience. *Tourism and Hospitality Research*. <https://doi.org/10.1177/14673584251361636>
- Nor, N. M. S. N. M., Othman, M. S., Zawawi, A. N. A., & Shabani, N. A. S. (2024). Factors influencing satisfaction towards premium coffee shops in Penang. *Journal of Tourism, Hospitality & Culinary Arts*, 16(1), 102-121.
- Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460–469. <https://doi.org/10.1177/002224378001700405>
- Otto, A. S., Szymanski, D. M., & Varadarajan, R. (2020). Customer satisfaction and firm performance: insights from over a quarter century of empirical research. *Journal of the Academy of Marketing Science*, 48(3), 543–564. <https://doi.org/10.1007/s11747-019-00657-7>
- Peng, K., Peng, Y., & Li, W. (2023). Research on customer churn prediction and model interpretability analysis. *PLoS ONE*, 18(12), e0289724. <https://doi.org/10.1371/journal.pone.0289724>
- Poudel, S. S., Pokharel, S., & Timilsina, M. (2024). Explaining customer churn prediction in telecom industry using tabular machine learning models. *Machine Learning With Applications*, 17, 100567. <https://doi.org/10.1016/j.mlwa.2024.100567>
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. <https://doi.org/10.1007/bf00116251>

- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- So, K. K. F., Yang, Y., & Li, X. (2025). Fifteen years of Research on Customer Loyalty Formation: A Meta-Analytic Structural Equation Model. *Cornell Hospitality Quarterly*, 66(2), 253–272. <https://doi.org/10.1177/19389655241276506>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- 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. <https://doi.org/10.26552/ems.2024.2.85-98>
- Usman-Hamza, F. E., Balogun, A. O., Nasiru, S. K., Capretz, L. F., Mojeed, H. A., Salihu, S. A., Akintola, A. G., Mabayoje, M. A., & Awotunde, J. B. (2024). Empirical analysis of tree-based classification models for customer churn prediction. *Scientific African*, 23, e02054. <https://doi.org/10.1016/j.sciaf.2023.e02054>
- Wang, P., Wu, L., Gao, L., & Mattila, A. S. (2024). The Servicescape and its Impact on Consumer Satisfaction: A Meta-Analysis. *Psychology and Marketing*, 42(3), 799–816. <https://doi.org/10.1002/mar.22152>
- 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>
- Zhang, H., & Zhang, W. (2024). Application of GWO-attention-ConvLSTM model in customer churn prediction and satisfaction analysis in customer relationship management. *Heliyon*, 10(17), e37229. <https://doi.org/10.1016/j.heliyon.2024.e37229>
- Zhou, F., Jiang, Y., Qian, Y., Liu, Y., & Chai, Y. (2023). Product consumptions meet reviews: Inferring consumer preferences by an explainable machine learning approach. *Decision Support Systems*, 177, 114088. <https://doi.org/10.1016/j.dss.2023.114088>
- Zschech, P., Weinzierl, S., & Kraus, M. (2025). Inherently interpretable machine learning: a contrasting paradigm to post-hoc explainable AI. *Business & Information Systems Engineering*, 68(2), 445–463. <https://doi.org/10.1007/s12599-025-00964-0>