Pamukkale Univ Muh Bilim Derg, 31(2), 267-275, 2025

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

A decision tree approach utilizing Quadrant analysis for satisfaction in high-speed rail systems

Yüksek hızlı raylı sistemlerde memnuniyet için Quadrant analizini kullanan bir karar ağacı yaklaşımı

Güzin AKYILDIZ ALÇURA[1], Sümeyye Şeyma KUŞAKCI GÜNDOĞAR[2], Abdulsamet SARACOGLU[1]*

[1]Department of Civil Engineering, Faculty of Civil Engineering, Yıldız Technical University, İstanbul, Türkiye.
akyildiz@yildiz.edu.tr, samet.saracoglu@yahoo.com

[2]Department of Civil Engineering, Civil Engineering Faculty, İstanbul Technical University, İstanbul, Türkiye.
seymakusakci@gmail.com

Received/Geliş Tarihi: 19.12.2023 Revision/Düzeltme Tarihi: 29.04.2024 doi: 10.5505/pajes.2024.33348
Accepted/Kabul Tarihi: 20.05.2024 Research Article/Araştırma Makalesi

Abstract

Satisfaction is an essential parameter for a public transport service provider, mainly if providers compete. For this reason, decision-makers try to identify and improve the factors affecting passenger satisfaction, which is important not only for the profitability of the service but also for sustainable and environmentally friendly transportation. In this study, 803 passengers were asked to rate their expectations and perceptions about the High-Speed Rail Systems they experienced between Ankara and Eskişehir in Turkey. By using the obtained data, a methodology was established to predict factors affecting satisfaction. The Classification and Regression Tree was used for the analysis, which does not require any predefined baseline relationship between dependent and independent variables. The quadrant analysis is utilized to determine a single input to represent expectations and perceptions instead of using them separately, which allows the comparison of expectations and perceptions. Also, the "problem experiencing rate," the decisive component of the Impact Score technique, is included in the model as an input. As a result, passengers' satisfaction is estimated with an accuracy rate of 88.79%, and the factors affecting satisfaction were determined as Service Delivery, Fare Level and Type, Passenger Information, Accessibility, and Security.

Keywords: Public transportation, Service quality, Decision trees.

Öz

Memnuniyet, özellikle operatörler arasında rekabet varsa, bir toplu taşıma sistemi sağlayıcısı için önemli bir parametredir. Bu nedenle karar vericiler, sadece hizmetin karlılığı için değil, aynı zamanda sürdürülebilir ve çevre dostu ulaşım için de önemli olan yolcu memnuniyetini etkileyen faktörleri belirlemeye ve iyileştirmeye çalışmaktadır. Bu çalışmada, Türkiye'de Ankara ve Eskişehir illeri arasında Yüksek Hızlı Raylı Sistem kullanan 803 yolcudan aldıkları hizmete ilişkin beklenti ve algılarını derecelendirmeleri istenmiştir. Elde edilen veriler kullanılarak memnuniyeti etkileyen faktörleri tahmin etmek için bir metodoloji oluşturulmuştur. Analiz için, bağımlı ve bağımsız değişkenler arasında önceden tanımlanmış herhangi bir temel ilişki gerektirmeyen Sınıflandırma ve Regresyon Ağacı kullanılmıştır. Beklentileri ve algıları ayrı ayrı kullanmak yerine ikisini temsil edecek tek bir girdiyi belirlemek için Quadrant Analizi kullanılmıştır. Bu, beklentilerin ve algıların karşılaştırılmasını dahil etme şansını getirmiştir. Ayrıca, Etki Skoru tekniğinin belirleyici bileşeni olan "sorun yaşama oranı" modele girdi olarak dahil edilmiştir. Sonuç olarak, yolcuların memnuniyeti %88.79 doğruluk oranı ile tahmin edilmiş ve memnuniyeti etkileyen faktörler Hizmet Sunumu, Ücret Düzeyi ve Türü, Yolcu Bilgileri, Erişilebilirlik ve Güvenlik olarak belirlenmiştir.

Anahtar kelimeler: Toplu taşıma, Hizmet kalitesi, Karar ağaçları.

1 Introduction

Increasing the service quality of transportation systems for passenger satisfaction and loyalty draws attention due to limited resources and changing user characteristics. The expectations and perceptions of passengers differ according to many factors, such as demographic characteristics and their experiences before and during the trip. Therefore, managers collect data from passengers regarding many aspects of the system to find the most influential factors in satisfaction. Analyzing these data with new techniques makes it possible to select the most suitable features of the system to be improved. It provides details about the relationship between service quality and satisfaction.

The advantages and disadvantages of many methods used to determine service quality in public transport systems are summarized in the review study conducted by de Ona et al. [1]. Accordingly, the methods used are classified as taking performance and expectation into account or only considering performance. Another classification is based on the methods that provide a service quality index value for the whole system (aggregate) or methods giving separate results for the quality attributes (disaggregate). Considering both classifications, methods can be grouped under four main headings:

1) Aggregated performance-expectation models (e.g., Servqual),
2) Aggregated performance-only models (e.g., Servperf),
3) Disaggregated performance-expectation models (development of Zone of Tolerance (ZOT) concept for assessment of service quality-ZSQ),
4) Disaggregated performance models (e.g., Quadrant (Importance-Performance) Analysis).

The most used service quality measurement method is Servqual, developed by Parasuraman et al. [2]. The method is based on the Gap theory, in which the difference between expectation and perception measurements represents the

*Corresponding author/Yazışılan Yazar

system's service quality. The most important objection to the method involves passenger expectations. Cronin and Taylor [3], who argue that service quality should be considered an attitude, revealed that the variation in service quality is shown better with Servperf, a performance-based measurement. They state that perceptions are already the result of customers' comparison of the expected and actual service. Another method, quadrant analysis, is the primary method for disaggregating data. It has advantages such as straightforward interpretations through its graphical display, providing weight for each component, and the determination of priorities for improvements.

Due to their structure, some methods cannot be grouped according to the abovementioned classification or fall into more than one group. Impact Score Technique, one of these methods developed by Morpace, is a disaggregate method that only considers performance [4]. Each attribute is ranked according to its score in the model, calculated by the performance ratings and incidence rate (the rate determined by the answer to the question "whether or not the passengers have experienced a problem in the last 30 days" for each component).

In this study, the critical aspects of the methods widely used in studies on service quality and satisfaction measurement are brought together to benefit from the advantages of the essential key points. One of these approaches is reducing expectation and perception ratings into one dimension with the help of Quadrant analysis. In the analysis of quadrants, the axes represent expectation and satisfaction, and it is a known fact that evaluations falling into the region where expectation is high and satisfaction is low are particularly important. Taking expectation and satisfaction into account in a single dimension is important, especially for simplifying the inputs of the decision tree. In addition to the Quadrant analysis approach, an important component of the Impact Score Technique developed by Morpace, the "problem experiencing" rate statement was also included in the model as an input.

This study combines the advantages of the most used basic models of Servqual, Quadrant Analysis, and Impact Score technique. One critical method involves merging expectation and perception ratings into a single dimension using Quadrant analysis. Taking expectation and perception into account in a single dimension is important, especially for simplifying the inputs of the decision tree. In addition to the Quadrant analysis approach, an important component of the Impact Score Technique, the "problem experiencing" rate statement, was also included in the model as an input. By combining these approaches, the study creates a forecasting model that leverages proven methods in satisfaction prediction. The study, starting with the Introduction section, continues with the Literature section, which consists of general service quality studies mainly using decision trees. Then, Data Preprocessing, the Methodology, and the Results of the model are given in sections 3, 4, and 5, respectively. Finally, the Conclusion is provided in the 6th section.

2 Literature

With the global developments in the transportation sector, studies are increasingly looking for factors affecting the satisfaction and loyalty of passengers towards the systems. In Alpu's study, the service quality of the High-Speed Rail System (HSRS) was examined to predict overall customer satisfaction [5]. The results showed that staff have a significant impact on customer satisfaction. Tsafarakis et al. use a multi-criteria satisfaction analysis method to measure passenger Satisfaction from a broad set of service dimensions and identify those needing improvement. The analysis tool, a survey-based software, is used to evaluate the passenger satisfaction level of an airline company. According to the results, the model found low and high-importance sub-criteria. It is stated that customer satisfaction is a dynamic parameter of the business organization [6].

Wu et al. examined the interrelationships among the service quality dimensions, service quality, perceived value, corporate image, customer satisfaction, and behavioral intentions for Taiwan HSRS in 2011. A multiple regression analysis was developed using the results of the questionnaire of 239 passengers. A multi-level, hierarchical framework was used to understand the relationship between service quality, value, corporate image, customer satisfaction, and behavioral intentions [7]. In their study, Kuo and Tang investigated the demands and travel behaviors of elderly passengers who use Taiwan HSRS. They used 341 observations in structural equation modeling and found that satisfaction directly affects travel behaviors. They also found that variables such as the ticket window's convenience, transferability availability, navigation indexes, and broadcasting services strongly affect satisfaction. So, their study provided valuable outputs for operating the HSRS considering the elderly's usage [8]. Another research was conducted by Chou et al. in 2014 for Taiwan HSRS. They also used structural equation modeling to explain customer loyalty. They reduced the service quality attributes into four factors with factor analysis. According to the results, service quality positively affects customer satisfaction and loyalty, while customer satisfaction positively affects customer loyalty [9].

Chen et al. aimed to determine passenger demands and analyze passenger satisfaction using online review analysis for HSRS trips. They reviewed online blogs and extracted the demands of passengers by classifying them into six groups. Also, they performed an online survey to find passengers' satisfaction levels with the demands they obtained. The findings show HSRS that passengers are not satisfied with the in-cabin features. They proposed a new methodology to evaluate HSRS passenger satisfaction based on an online review analysis [10].

Zhen et al. (2016) investigated the effect of access and egress on overall HSR satisfaction using a path analysis. They divided HSR satisfaction into four parts: Satisfaction with waiting, line-haul, HSR access, and HSR egress. It is determined that HSR line-haul satisfaction dominates overall HSR satisfaction; HSR access and egress sections have an equivalent effect [11]. Another study in the same corridor conducted by Zhen et al. (2018) used multivariate regression and importance-performance analysis to identify influential attributes and service improvement priorities. Staff attitudes, convenience of ticket purchase, and ease of trip access are the most important of the 17 attributes. They also used quadrant analysis to assess performance and determine the importance of measures. They named low-performance and high-importance areas concentrated areas and suggested prioritizing toilet sanitation and seat comfort[12]. In another study on HSRS in Turkey, service quality is considered technical (the service delivered), quality, and functional (the way the service is delivered to customers) quality [13]. It is determined that there is a relationship between corporate image, customer satisfaction, customer complaints, and customer loyalty.

Safitri and Surjandari investigated the possibility of people changing their mode of travel to improve the quality of service provided in Jakarta, where public transport intention is low. In the study in which the decision tree approach was used to predict travel mode transition, the results showed that 61.8% of public transport vehicles and 57.1% of private vehicle users tend to change their current mode of travel [14]. Parasuraman et al.'s [15] decision tree algorithm, including the GAP Theory, was used by Tsami and Nathanail [16], and a decision tree was developed that links users' perceptions and expectations to service quality using 26 quality indicators selected in a survey. "Information accessibility via telephone, mail" was the most critical parameter in evaluating the overall service.

Tsami et al. used a decision tree approach to assess the design, operation, and services provided at urban interchanges. By using the survey results of 239 users of the Riga International Coach Terminal on crucial attributes, including travel information, wayfinding information, time and movement, access, comfort and convenience, station attractiveness, safety and security, emergency handling and overall satisfaction, a decision tree was used to model how the performance evaluation of the selected indicators affect overall satisfaction level of the terminal. They have collected data for 37 different indicators and evaluated them. According to the results, the "surrounding area" has been found to have the highest correlation with overall satisfaction [17].

De Ona and De Ona studied the effects of segmentation of passengers on the perception of service quality by using decision trees, in which they focused on gender. Six models and 13 attributes were used to describe the bus transit service. Since they have analyzed three different data sets for different years, it is determined that the key factors, both for men and women, change over time [18].

3 Data preprocessing

The data used in this study were determined from a questionnaire conducted with 803 users of the HSRS line between two cities: Ankara and Eskisehir (Figure 1). This line is the first high-speed rail system in Turkey and the first part of the Ankara-Istanbul railway line. This 245 km long section has four stops and 1 hour and 30 minutes. The intercity bus system is another alternative for traveling between two cities, and its duration is about 3 hours. The interviews took place from the middle of March to the end of April 2015, on weekdays and weekends. HSRS users responded to all the questionnaires during the journey on board (%50.4 in Ankara to Eskişehir direction and %49.6 in Eskişehir to Ankara direction).

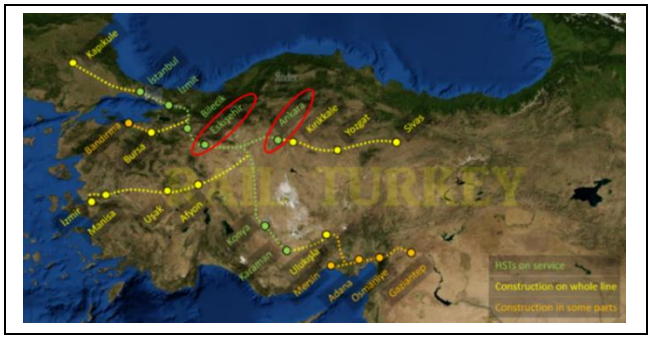


Figure 1. Turkish high-speed line network [19].

The questionnaire was structured in three sections.

1) The first section was dedicated to demographic questions, as summarized in Table 1.

Table 1. Sample characteristics.

Demographic Information		Number	%
1. Gender	Female	345	43.0
	Male	458	57.0
2. Age	18-26	386	48.1
	27-59	408	50.8
	>59	9	1.1
3. Qualification	Primary School	8	1.0
	Secondary School	11	1.4
	High School	215	26.8
	Undergraduate	425	52.9
	Postgraduate	144	17.9
4. Car Ownership	Own Car	296	36.9
	Family Car	331	41.2
	No Car	176	21.9
5. Income (monthly T.L.) The exchange rate for the American dollar, stated by the Central Bank of Turkey, was 2.6 Turkish Liras during the questionnaire.	No income	136	16.9
	1–2,500	270	33.6
	2,501–5,000	310	38.6
	5,001–7,500	65	8.1
	>7,500	22	2.8
6 . Trip Frequency	Everyday	82	10.2
	Several times a (STa) Week	143	17.8
	STa Month	368	45.8
	STa Year	210	26.2
7. HSRS Experience	First Experience	35	4.7
	Other	768	95.3
8. Trip Purpose	Work	119	14.8
	School	161	20.0
	Visiting Relatives	202	25.2
	Sightseeing	79	9.8
	Healthcare	30	3.8
	Business	185	23.0
	Private	14	1.8
	Other	13	1.6
9. Work Status	Not Working	198	24.7
	Self Employed	101	12.6
	Employee	504	62.7

2) In the second section of the questionnaire, participants rated 61 questions on a 5-point scale for evaluation of their perceptions and expectations about the dimensions of service quality. These seven dimensions adapted from TRB-Report 47, [4] and constructed to involve all the trip phases are as follows:

- Passenger information (easiness and reliability of information including the response time and accuracy of information),
- Fare level and type (cost of travel),
- Accessibility (easiness of accessing to stations),
- Station environment (lighting, cleanliness, seat availability, information boards and announcements),
- Vehicle environment (lighting, air quality, information boards, seat number, and comfort),
- Service delivery (frequency, reliable schedule), and

- Security (safety from crime, emergency precautions at stations and vehicles).
- All participants answered each question in 3 steps:
- What is their expectation about this dimension of service quality (5-point scale, "1 for not at all important" to "5 for very important")
- What is their perception of this dimension of service quality (5-point scale, "1 for not at all satisfied" to "5 for very satisfied")
- Have they recently experienced a problem with this attribute of service quality within the last 30 days? (Yes or No).

The problem experiencing rates and expectation-perception ratings of passengers determined from the questionnaire are summarized in Table 2.

Table 2. Service quality characteristics, problem experience, and evaluation ratings

Dimension of Service Quality	Experienced a Problem (% Yes)	Expectation (Mean/Std. Dev.)	Perception (Mean/Std. Dev.)
Passenger Information	55.42	4.48/0.62	3.96/0.79
Fare Level and Type	52.30	4.43/0.53	3.86/0.71
Accessibility	36.61	4.55/0.54	3.59/0.80
Station Environment	35.24	4.38/0.50	3.64/0.66
Vehicle Environment	37.86	4.42/0.54	3.65/0.72
Service Delivery	68.37	4.38/0.56	3.63/0.71
Security	52.30	4.48/0.52	3.72/0.70

3) The last section of the questionnaire rated the Satisfaction, Loyalty, Perceived Value, Image, Involvement, and Trust factors. Satisfaction is the comparison of the perceived service and expectations [20]. If the offered service meets expectations, the customer is assumed to be satisfied, or vice versa [21]. Loyalty is a customer's willingness to continue to use the service [22]. Perceived value compares what is received and what the customer gives in return [23]. Image is the total evaluation of a brand [24]. In other words, it is the picture of the product or a firm in a customer's mind [25]. Involvement represents an individual's motivation based on natural needs, judgment, and interests [26]. Trust is the behavioral intention or behavior that reflects a reliance on a partner [27]. In the study, passengers were asked to rate their perceptions about these factors on a scale from 1 to 5, and the optional answers ranged from "1, strongly dissatisfied" to "5, strongly satisfied". As a result, the average scores for Satisfaction, Loyalty, Perceived Value, Image, Involvement, and Trust were found to be 4.49, 4.01, 3.75, 3.88, 4.09, and 3.89, respectively.

The demographic characteristics were collected in the questionnaire, where the rest of the data was processed before input. These processes are summarized as follows:

1. The respondents' answers (Yes or No) for the problem experienced for each dimension of service quality within the last 30 days (for passenger information, fare level and type, accessibility, station environment, vehicle environment, service delivery, and security) are taken as input.
2. Expectations and perceptions are two fundamental parameters for analyzing Service Quality since Parasuraman et al. conducted studies from 1985 through 1988 [2], [15]. Therefore, using both performance and expectation is important for evaluation. To assess the expectations and perceptions of passengers for each service quality, passengers' ratings for service dimensions are carried to a perception-expectation axis set resembling a quadrant analysis (Figure 2). Each of the four areas was named separately, and these areas were involved in the decision tree model representing service quality expectation and perception evaluation. For example, 1st quadrant involves Low Expectation and High Perception (LE-HP). In the 2nd quadrant, both Expectation and Perception ratings are high (HE-HP). In the 3rd one, both of them are low (LE-LP). Finally, in the 4th quadrant, expectation is high while perception is low (HE-LP). The service dimensions that fall in this 4th (HE-LP) quadrant are determined as the first to be ameliorated. The 3rd and 4th quadrants (LE-LP; HE-LP) show low perception. The intersection point of the axis is set as 3 since ratings were classified as "1, 2, 3" indicating low values, and "4, 5" as high values.
3. Loyalty, Perceived Value, Image, Involvement, and Trust ratings were again grouped into two classes. Ratings 1, 2, and 3 are grouped in one class and named low; ratings 4 and 5 are grouped in another class and named high.

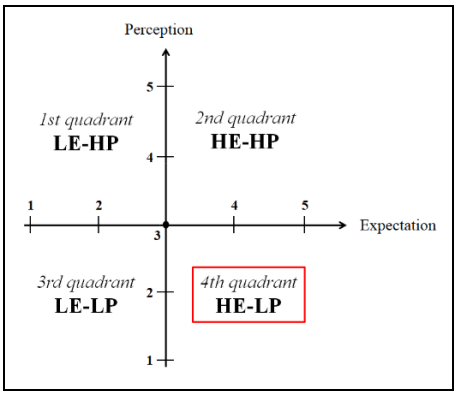


Figure 2. Quadrant analysis.

4 Methodology

One of the methods used in measuring service quality and satisfaction is Classification and Regression Trees (CART), a branch of data mining technique, or Decision Trees in general. This method has proven to be suitable for analyzing Public Transport (P.T.) service quality in the study of de Ona et al. [28]. Classification is a form of data analysis that identifies key data classes as a two-step process illustrated in Figure 3. In the first step, the data is used to build a model. Then, in the second step, the model is tested to determine whether it is suitable for classifying the data. For estimation problems, numerical estimation models such as classification (for discrete data) or regression models (for continuous data) are used [29].

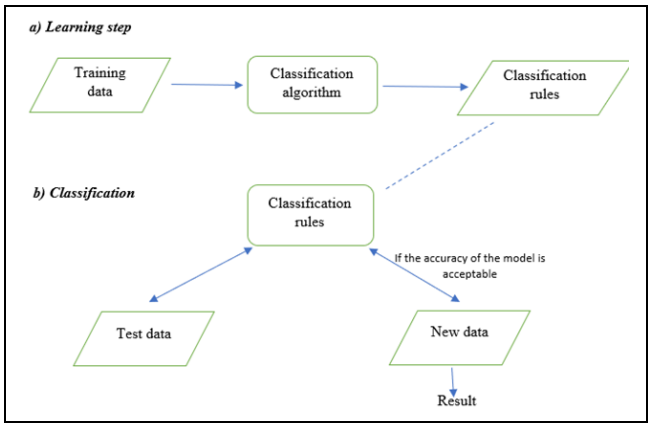
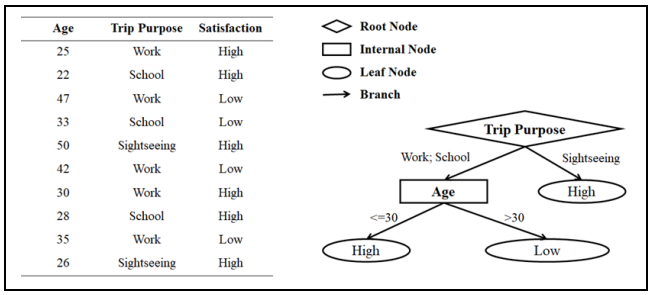


Figure 3. The data classification process [29]

The CART analysis allows many explanatory variables to be processed and makes finding the most critical variables easy. In their study, de Ona et al. aimed to model the perception of service quality among individuals by using a classification tree. They studied two phases of passenger evaluation: pre-evaluation and post-evaluation. The analysis of the two models found that passengers' perceived quality of service is practically limited to Frequency, Speed, and Punctuality in the preliminary evaluation [28].

If the dependent variable is categorical, the CART model is called a "Classification Tree"; if the variable is numerical, the model is called a "Regression Tree." In the classification process, only binary split is produced, and the 'Gini' or 'twoing' index is used as the difference criterion. The process of a CART model consists of three steps. The first step is the tree-growing step. This step aims to maximize the "purity" in the two child nodes. The second step is the "pruning" step, in which branches that contribute little to the model are eliminated. In the last step, the best tree is selected from the pruned models [30].

Decision trees consist of root nodes, internal nodes, leaf nodes, and branches. According to the example in Figure 4, two variables (age and travel purpose) and a class label (Satisfaction) are in the data. In the example, the decision tree model starts with the trip purpose variable selected as the root node. The data is then divided into two groups: travel and work/school purposes. Groups are as homogeneous as possible. Therefore, it is classified as a leaf node and labeled "High".



Age	Trip Purpose	Satisfaction
25	Work	High
22	School	High
47	Work	Low
33	School	Low
50	Sightseeing	High
42	Work	Low
30	Work	High
28	School	High
35	Work	Low
26	Sightseeing	High

Figure 4. A sample data and decision tree.

The other group's data is divided into two groups according to the age variable. Since these groups are as homogeneous as possible, each group is classified as a leaf node (labeled "High" and "Low"), and the modeling process is completed. The homogeneity of leaf nodes does not need to be 100%; it varies depending on the threshold value. In this case, the label of the leaf node is defined according to the label of the samples that make up more than 50% of the group.

The k-fold cross-validation method was used to evaluate the effectiveness of the model. This method divides the data into k sets, and k-1 of them is used as the training set for developing a model. The remaining 1 set is chosen as the testing set, and test results are obtained. This process is repeated for k iterations—each time, a different set is used for testing, and the remaining part is used for training. The test results for each iteration are obtained, and their average value is used to evaluate the developed model. For the evaluation of the effectiveness of our model, the process mentioned above was performed by taking k as 10.

The methods based on data mining have the advantage of not needing assumptions or predefined underlying relationships between dependent and independent variables, as de Ona et al. [32] stated. Their studies also mention the advantages of "If-then" rules that allow companies to decide the strategy based on their resource limitations. "If-then" rules predict a target variable (e.g., satisfaction) when a set of conditions are complied with. These rules were used by Liou et al. [33] to develop airport service quality improvement.

To summarize the results of the decision tree approach, If-then rules are used in many studies ([28],[31],[32],[34]) about service quality. The rules stated as the results of decision trees made it possible to find the most effective service quality attributes. Moreover, decision trees with effective if-then rules provide a practical model for managers and operators [32].

After preparing the data, the CART model was developed and implemented using a 10-fold sample cross-validation. The dependent variable was satisfaction, and 28 independent variables are shown in Table 3.

Table 3. Decision tree independent variables.

Demographic variables	1. Gender 2. Age 3. Qualification 4. Car Ownership 5. Income (monthly TL) 6. Trip Frequency 7. HSRS Experience 8. Trip Purpose 9. Work Status
Experiencing a problem rates (Yes or No)	1. Passenger Information 2. Fare Level and Type 3. Accessibility 4. Station Environment 5. Vehicle Environment 6. Service Delivery 7. Security
Quadrant area variables (LE-HP, HE-HP, LE-LP, HE-LP)	1. Passenger Information 2. Fare Level and Type 3. Accessibility 4. Station Environment 5. Vehicle Environment 6. Service Delivery 7. Security
Factors that Affect Satisfaction (1. 2. 3 as Low-4, 5 as High).	1. Loyalty 2. Perceived Value 3. Image 4. Involvement 5. Trust

5 Results

In this research, an estimation model to determine passengers' satisfaction is developed by using the decision tree model. The graphical results of the model in a decision tree form are given in Figure 5. The rules determined for satisfaction according to service quality characteristics, problem experience, and evaluation ratings are given in Table 4. The tree provides five levels, 20 nodes, and 11 terminal nodes. Nine parameters are used as efficient variables in the models. The effective dimensions and factors determined in the model are Involvement, Service Delivery, Perceived Value, Passenger Information, Accessibility, Fare Level and Type, Vehicle Environment, Security, and Trip Frequency. Involvement and Perceived Value contribute to the model with low or high evaluations. At the same time, Service Delivery, Passenger Information, Accessibility, Fare Level and Type, and Security are considered with expectation-perception (quadrant) evaluations (LE-HP, HE-HP, LE-LP, HE-LP). Finally, the model involves Vehicle Environment evaluations with problem-experiencing answers (Yes-No) and Trip frequency with usage per week, year, month, or day.

The first split in the tree started with Involvement, then passengers with a low perception of Involvement on the left and a high perception on the right side formed the following branches. Service Delivery is on the left branch, and Perceived Value is determined as the new nodes on the right.

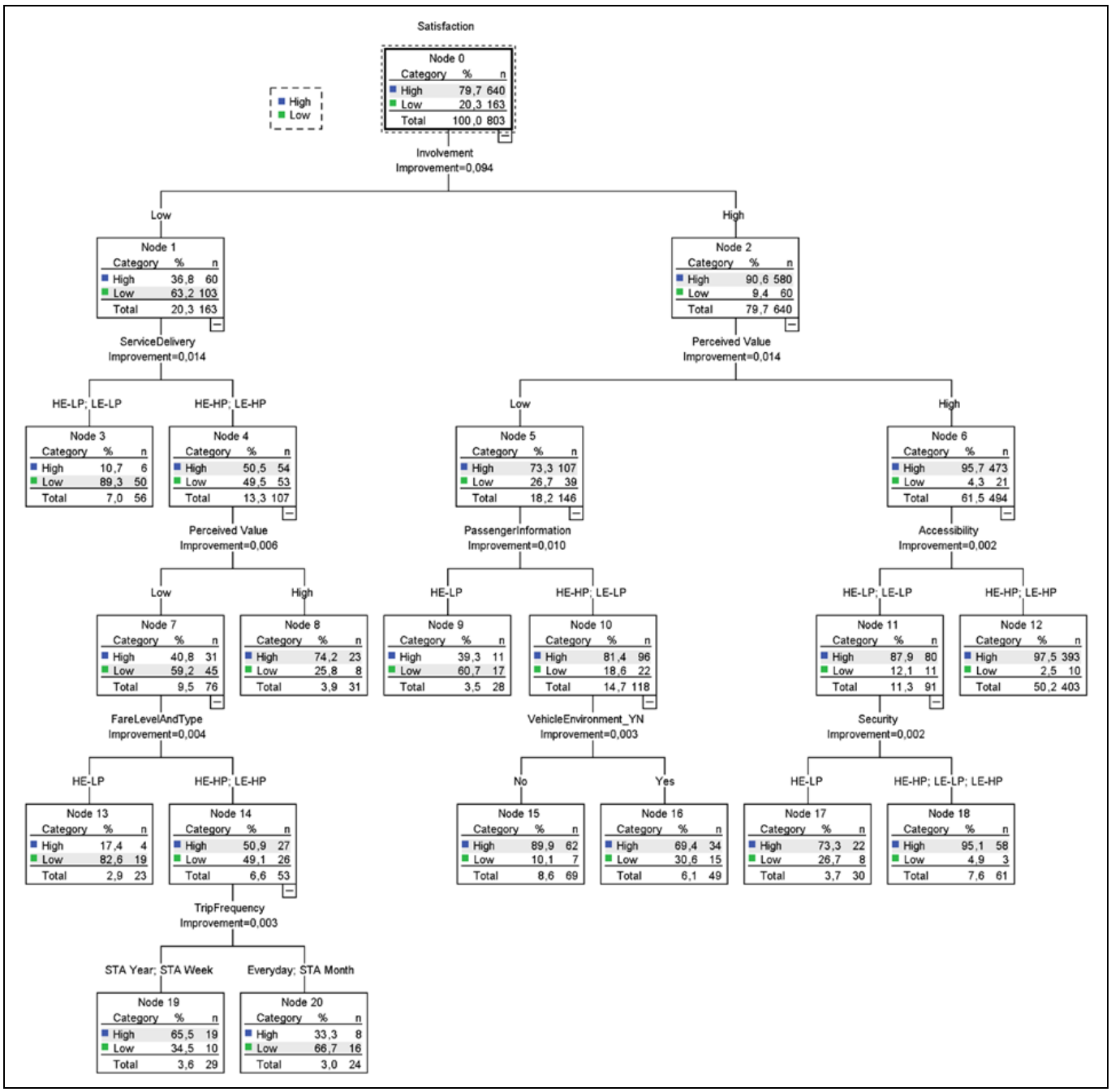


Figure 5. Developed CART model.

Table 4. Rules for satisfaction according to service quality characteristics, problem experience, and evaluation ratings.

Node	Rule		Accuracy rate (%)
	If	Then, satisfaction is rated as	
3	Involvement is rated as Low; Service Delivery is rated as HE-LP or LE-LP	"Low"	89.3
8	Involvement is rated as Low; Service Delivery is rated as HE-HP or LE-HP, and Perceived Value is rated as High	"High"	74.2
13	Involvement is rated as Low, Service Delivery is rated as HE-HP or LE-HP, Perceived Value is rated as Low, Fare Level and Type is rated as HE-LP	"Low"	82.6
20	Involvement is rated as Low; Service Delivery is rated as HE-HP or LE-HP, Perceived Value is rated as Low; Fare Level and Type is rated as HE-HP or LE-HP; Trip Frequency is rated as STA Month or every day	"Low"	66.7
19	Involvement is rated as Low, Service Delivery is rated as HE-HP or LE-HP, Perceived Value is rated as Low, Fare Level and Type is rated as HE-HP or LE-HP, Trip Frequency is rated as STA Year or STA Week	"High"	65.5
9	Involvement is rated as High, Perceived Value is rated as Low, Passenger Information is rated as HE-LP	"Low"	60.7
15	Involvement is rated as High; Perceived Value is rated as Low; Passenger Information is rated as HE-HP or LE-LP; Vehicle Environment_YN is rated as No	"High"	89.9
16	Involvement is rated as High; Perceived Value is rated as Low; Passenger Information is rated as HE-HP or LE-LP; Vehicle Environment_YN is rated as Yes	"High"	69.4
12	Involvement is rated as High, Perceived Value is rated as High, and accessibility is rated as HE-HP or LE-HP	"High"	97.5
17	Involvement is rated as High; Perceived Value is rated as High; accessibility is rated as HE-LP or LE-LP; Security is rated as HE-LP	"High"	73.3
18	Involvement is rated as High, Perceived Value is rated as High, Accessibility is rated as HE-LP or LE-LP, Security is rated as HE-HP, LE-LP or LE-HP	"High"	95.1

The first terminal node (Node 3) represents the passengers whose evaluations about Service Delivery fall on the HE-LP quadrant (low perception and high expectation). Node 4, with a high perception of Service Delivery, is split by Perceived Value, which forms the second terminal node on the left. Node 8 shows the passengers have a high perception of Perceived value and satisfaction with 74.2% accuracy. The split continued with Fare Level and Type following the node showing passengers with low Perceived value. The evaluations of HE-LP for Fare Level and Type resulted in another terminal node. This node (Node 13) shows that passengers with low perception and high expectations (HE-LP area) for Fare Level and Type, besides other evaluations mentioned previously, have low satisfaction (with an accuracy of 82.6%). For the passengers with a high perception of Fare Level and Type, splitting continued with Trip frequency. Passengers using the system several times a year or week have high Satisfaction (Node 19), while passengers traveling several times a month or every day have low Satisfaction (Node 20).

On the right side of the tree, the passengers with a high perception of Involvement were split by their perceptions of perceived value. Passengers with low Perceived Value after branching by Passenger Information created a new terminal node. Passenger information evaluations in the HE-LP area have resulted in low satisfaction. Passengers with a high perception of Passenger Information were divided by their evaluation of the Vehicle Environment. When Node 15 (passengers who did not experience a problem) and Node 16 (passengers who did experience a problem) are compared, passengers have high satisfaction. However, accuracy rates differ according to problem experiencing rates.

Passengers with high Involvement and high Perceived Value, on the right side of the tree, are evaluated according to their Accessibility expectations and perceptions. A high perception of accessibility regardless of expectation (HE-HP and LE-HP) brought high satisfaction with 97.5% accuracy (Node 12). Passengers with low perception, again, regardless of their expectations, are split by their evaluations of security. Security evaluations do not affect satisfaction level, but the accuracy rates differ for HE-LP (Node 17-accuracy rate of 73.3%) and HE-HP, LE-LP, LE-HP (Node 18-accuracy rate of 95.1%) evaluations.

Table 5 shows the model validation results. Model validation was implemented on precision, recall, and overall accuracy values. The definitions of these values are as follows:

Recall: "High" or "low" sample ratio is predicted accurately by the total of samples observed as "high" or "low."

Precision: "High" or "low" samples' ratio predicted accurately to the total of samples predicted as "high" or "low".

Overall Accuracy: The total number of samples' ratio predicted accurately to the total number of samples observed.

According to the results, 95.47% of 640 high-satisfied cases and 62.58% of 163 low-satisfied cases were predicted accurately by the CART model. When the results are evaluated from another point of view by precision values given in Table 5, it can be seen that according to the model results, precision values were 90.92% and 77.86% for cases predicted as high-satisfied and

low-satisfied, respectively. The overall accuracy value of the model is 88.79%, which can be evaluated as a high prediction value.

Table 5. Analysis results.

		Predicted			
		High	Low	Total	Recall
Observed	High	611	29	640	95.47%
	Low	61	102	163	62.58%
	Total	672	131	803	-
Precision		90.92%	77.86%	-	88.79%

6 Discussion

In this study, data for predicting the satisfaction of HSRS passengers were collected and then analyzed using the decision tree method. The relation between satisfaction and demographic information, service quality characteristics, and other parameters was analyzed. The CART model, developed to define the characteristics of passengers who are satisfied (high) or not satisfied (low), resulted in a high accuracy value of 88.79%.

According to the authors, this is the first time that expectations and perceptions are involved in the model as a single dimension. There have been many discussions about using expectations in service quality measurements. Cronin and Taylor [3] argue that measuring service quality with a performance-based method must be considered, while Iacobucci et al. [35] state that passengers' expectations must be known to provide a good level of service. Instead of involving these measurements separately, an indicator showing low/high levels of expectation and performance measures in a single input will contribute more to the service quality and satisfaction evaluations of transportation systems.

The branching of the tree determined as the result of the model provides ease in following the relations visually. For the general evaluation of the model, the service dimensions that fall on the HE-LP quadrant mainly resulted in low satisfaction, realizing that the difference between expectation and perception represents the level of satisfaction.

Involvement can be defined as being in touch with or familiar with the system. Passengers with high Involvement evaluation were found to have high satisfaction, or vice versa. Low Involvement might cause passengers to be more critical about the service.

Problem-experiencing data was found effective in terms of Vehicle Environment. Even in both situations (Yes or No), the satisfaction was high, and the accuracy rate decreased (from 89.9% for No to 69.4% for Yes) when the passengers answered problem-experiencing questions as Yes. All quadrant areas are used in the decision tree separately with equal importance. HE-LP area and LE-LP area (both having Low Perception) were found in the same tree branch two times, and LE-HP area and HE-HP area (both of them have High Perception) were found in the same branch four times, which shows that perception has a more important role than expectation.

After conducting more tests on these relations, managers would have an easy way to predict the results of changes or the most effective ways to enhance the system. After identifying the relationships, more detailed studies should be conducted for the final decisions.

EN 13816 Service Quality Management Standard in Passenger Transportation was created in 2002 to develop a quality approach in urban public transportation. It is a resource for public transportation service quality management and related standards [36]. Continuous updating of such standards through studies is among the suggestions offered.

7 Author contribution statement

Güzin AKYILDIZ ALÇURA contributed to forming the idea, literature review, data collection, and analysis of the study results. S. Şeyma KUŞAKCI GÜNDOĞAR's contribution includes a literature review, data collection, and evaluation of results. Abdulsamet SARACOĞLU contributed to the article's modeling, spelling and checking in terms of content.

8 Ethics Committee Approval and Conflict of Interest Statement

“Ethics Committee permission was obtained to collect the data used in our study and it was uploaded to the system together with the article”.

“There is no conflict of interest with any person/institution in the article prepared”.

9 References

[1] De Oña J, De Oña R. "Quality of service in public transport based on customer satisfaction surveys: a review and assessment of methodological approaches". *Transportation Science,* 49 (3), 433-719, 2015.

[2] Parasuraman A, Zeithaml VA, and Berry LL. "SERVQUAL: multiple ıtem scale for measuring consumer perceptions of service quality". *Journal of Retailing* 64(1), 12–40. 1988.

[3] Cronin JJ, Taylor SA. "Measuring service quality: a reexamination and extension". *Journal of Marketing*, 56(3), 55–68, 1992.

[4] Transportation Research Board. “A Handbook for Measuring Customer Satisfaction and Service Quality” Washington DC, USA, TCRP Report 47, 1999.

[5] Alpu O. "A methodology for evaluating satisfaction with high-speed train services: A case study in Turkey". *Transport Policy*, 44, 151–157, 2015.

[6] Tsafarakis S, Kokotas T, Pantouvakis A. "A multiple criteria approach for airline passenger satisfaction measurement and service quality improvement". *Journal of Air Transport Management*, 68, 61-75, 2017.

[7] Wu JHC, Yu-Chiang L, Fu-Sung H. "An empirical analysis of synthesizing the effects of service quality, perceived value, corporate image and customer satisfaction on behavioral intentions in the transport industry a case of Taiwan high-speed rail". *Innovative Marketing*, 7(3), 83-100, 2011.

[8] Kuo CW, Tang ML. "Relationships among service quality, corporate image, customer satisfaction, and behavioral intention for the elderly in high-speed rail services." *Journal of Advanced Transportation*, 47(5), 512-525, 2013.

[9] Chou PF, Lu CS, Chang YH. "Effects of service quality and customer satisfaction on customer loyalty in high-speed rail services in Taiwan". *Transportmetrica A: Transport Science*, 10(10), 917-945, 2014.

[10] Chen ZS, Liu XL, Chin KS, Pedrycz W, Tsui KL, Skibniewski MJ. "Online-review analysis based large-scale group decision-making for determining passenger demands and evaluating passenger satisfaction: A case study of the high-speed rail system in China". *Information Fusion*, 69, 22-39, 2021.

[11] Zhen F, Cao X, Tang J. "The role of access and egress in passenger overall satisfaction with high-speed rail." *Transportation*, 46(6), 2137-2150, 2019.
[12] Zhen F, Cao J, Tang J. "Exploring correlates of passenger satisfaction and service improvement priorities of the shanghai-nanjing high-speed rail." *Journal of Transport and Land Use*, 11(1), 559-573, 2018.
[13] Yılmaz V, Erkan A. "The effects of service quality, image, and customer satisfaction on customer complaints and loyalty in high-speed rail service in Turkey: a proposal of the structural equation model". *Transportmetrica A: Transport Science*, 13(1), 67-90, 2017.
[14] Safitri DM, Surjandari I. "Travel mode switching prediction using decision tree in jakarta greater area". *International Conference on Information Technology Systems and Innovation (ICITSI)*, Bandung, Indonesia, 23-24 October 2017.
[15] Parasuraman A, Zeithaml VA, and Berry LL. "Conceptual model of service quality and ıts ımplications for future research". *The Journal of Marketing*, 49(4), 41–50, 1985.
[16] Tsami M, Nathanail E. "Guidance provision for ıncreasing quality of service of public transport". *Procedia Engineering*, 178, 551-557, 2017.
[17] Tsami M, Adamos G, Nathanail E, Budilovich E, Yatskiv I, and Magginas V. "A decision tree approach for achieving high customer satisfaction at urban ınterchanges". *Transport and Telecommunication*, 19(3), 194-202, 2018.
[18] De Oña R, De Oña J. "Analyzing transit service quality evolution using decision trees and gender segmentation". *WIT Transactions on The Built Environment*, 130, 611-621, 2013.
[19] Turkish Railway Journal and Review, "High Speed Trains". https://railturkey.org/travel/trains/hst/ (28.03.2022).
[20] LaBarbera PA, Mazursky DA. "Longitudinal assessment of consumer satisfaction/dissatisfaction: the dynamic aspect of the cognitive process". *Journal of Marketing Research*, 20(4), 393-404, 1983.
[21] Kotler P, Keller KL. *Marketing Management*, England, Global Edition, Pearson Education Limited, 2000.
[22] Cyr D, Hassanein K, Head M, Ivanov A. "The role of social presence in establishing loyalty in e-service environments". *Interacting with Computers*, 19(1), 43-56, 2007.
[23] Hellier PK, Geursen GM, Carr RA, Rickard A. "Customer repurchase ıntention: a general structural equation model". *European Journal of Marketing*, 37(11), 1762-1800, 2003.
[24] Keller KL. "Conceptualizing, measuring, managing customer-based brand equity". *Journal of Marketing* 57(1), 1-22, 1993.
[25] Akkaya E. "Marka imajı ve bileşenleri: otomobil sektöründe bir uygulama". *4. Ulusal Pazarlama Kongresi*, Antakya, Türkiye, 18-20 Kasım 1999.
[26] Zaichkowsky JL. "Measuring the ınvolvement construct". *Journal of Consumer Research*, 12(3), 341-52, 1985.
[27] Moorman C, Zaltman G, Deshpande R. "Relationships between providers and users of market research: the dynamics of trust within and between organizations". *Journal of Marketing*, 29(3), 314-328, 1992.
[28] De Oña J, De Oña R, and Calvo FJ. "A classification tree approach to identify key factors of transit service quality". *Expert Systems with Applications*, 39 (12), 11164–11171, 2012.
[29] Han J, Kamber M, Pei J. *Data Mining Concepts and Techniques*, 3rd ed. Massachusetts, USA, Morgan Kaufmann, 2012.
[30] Saracoglu A, Ozen H. "Estimation of traffic ıncident duration: a comparative study of decision tree models". *Arabian Journal of Science and Engineering*, 45(10), 8099-8110, 2020.
[31] De Oña J, De Oña R, Eboli L, Mazzulla G. "Heterogeneity in perceptions of service quality among groups of railway passengers". *International Journal of Sustainable Transportation*, 9(8), 612-626, 2014.
[32] De Oña R, Eboli L, Mazzulla G. "Key factors affecting rail service quality in northern Italy: A Decision Tree Approach." *Transport*, 29(1), 75-83, 2014.
[33] Liou JJ, Tang HC, Yeh W, Tsai C. "A decision rules approach for improvement of airport service quality". *Expert Systems with Applications*, 38(11), 13723-13730, 2011.
[34] De Oña J, De Oña R. "Quality in public transport services for promoting a sustainable mobility: A case study." *Carreteras,* 4(191), 81-90, 2013.
[35] Iacobucci D, Grayson KA, Omstrom AL. "The calculus of service quality and customer satisfaction: theoretical and empirical differentiation and integration", *Advances in Services Marketing and Management*, 3, 1-68, 1994.
[36] Üvenç G, Kulluk S. "Toplu taşımada hizmet kalitesi yönetimi uygulaması: Hizmet kalitesinde performans ölçüm sisteminin tanımlanması". *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 26(4), 734-748, 2020.