Service quality in rail systems: listen to the voice of social media

Samet Güner (sguner@sakarya.edu.tr)
Sakarya University

Kamil Taşkı̈n
Sakarya University

Halil İbrahim Cebeci
Sakarya University

Emrah Aydemir
Sakarya University

Research Article

Keywords: Rail, service quality, social media analytics, machine learning, sentiment analysis

Posted Date: August 26th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1980183/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Service quality is essential to increase and maintain users' loyalty to the railway system. In the literature, surveys have been used to measure the satisfaction of users, and statistical and mathematical methods have been applied to quantify the survey results. In recent years, user-generated contents, including comments and complaints shared via social media, have been used to measure the quality of rail services. These contents may provide important insights into the quality of the service provided with its dynamic structure. In this study, a SERVQUAL-based social media analytics approach is used to measure railway service quality, paying special emphasis on the temporal variations in a national rail system. Topic modeling was used to assign each content to the relevant service dimension and sentiment analysis was applied to measure the level of satisfaction. Importance-performance analysis was employed at the final stage to generate clear policy suggestions. Gathering more than 2.3 million social media messages posted from 2011 to 2021, we examined the temporal evolution of service quality of the Turkish rail system from users' points of view. The results revealed the most and least important services for passengers and the satisfaction level of each dimension. The differences between the priorities of conventional and high-speed rail passengers were defined and policy recommendations were presented.

1. Introduction

Rail systems are sustainable, safe, and cost-efficient alternatives compared to road and air transport. Previous researches demonstrated that an acceptable level of service has a decisive influence on user satisfaction and loyalty in the rail system (Kuo and Tang, 2013; Chou et al., 2014; Yilmaz and Ari, 2017). Therefore, the measurement of the rail service quality and customer satisfaction attracts considerable attention in the literature. As a common practice, user opinions are received by surveys or interviews in the first stage, and the service quality is measured in the second stage with statistical analysis (Cavana et al., 2007; Chou et al., 2014; De Ona et al., 2014; Alpu, 2015; Eboli and Mazzulla, 2015; Hadiuzzaman et al., 2019; Mandhani et al., 2020) and/or Multi-Criteria Decision Making techniques (Awasthi et al., 2011; Çelik et al., 2014; Khorshidi et al., 2016; Aydin et al., 2017).

In addition to structured surveys, voluntary messages left on social media reflecting user experiences contain important insights into the quality of rail services and offer significant improvement opportunities by understanding the passenger's needs and opinions (Gal-Tzur et al., 2014; Nisar and Prabhakar, 2018). Therefore, social media analysis and user-generated contents have been used in a few studies to measure the quality of rail services (Collins et al., 2013; Mogaji and Erkan, 2019; 2021; Osorio-Arjona et al., 2021; Mishra and Panda, 2022). Unlike traditional methods, these studies are based on user-generated content gathered from various social media platforms and measured the quality of rail services using various machine learning algorithms.

The unstructured big data of social media offers more penetration considering the voice of more users from a wider geographical region (Nepal et al., 2015), delivers more user views faster and at a lower cost (Rathore and Ilavarasan, 2020; Lock and Pettit 2020; Shirgaokar et al., 2021), provides real-time notifications (Pender et al., 2014; Gal-Tzur et al., 2014; Howard, 2020), may include users who are unwilling or unable to participate in structured studies (Casas and Delmelle, 2017; Howard, 2020), and offers a dynamic perspective allowing temporal evaluation (Collins et al., 2013; Hosseini et al., 2018; El-Diraby et al., 2019). Due to these advantages, social media tools are widely used in developing transport policies (Nepal et al., 2015; Grant-Muller et al., 2015; Casas and Delmelle, 2017).

This study aims to measure the quality of Turkish conventional and high-speed rail services and provide policy suggestions to improve the overall system. Unlike survey-based studies, user opinions were compiled from user-generated social media content. A social media analytics approach was then adopted, including topic modeling, machine learning, and sentiment analysis. The extended eight-dimensional SERVQUAL model proposed by Cavana et al. (2007) for rail services was adapted to social media analytics, including safety&security. More than 2.3 million social media messages related to rail services were gathered between 2011–2021 and each message was assigned to the relevant service dimension by topic modeling. Sentiment analysis was used in the second stage to examine user-generated content's polarity (negative or positive).
In recent years, priority has been given to railway policies in Turkey and significant investments have been made, including high-speed rail (HSR) systems. Despite its increasing share in the transportation budget and all the improvements, the operational efficiency of railway management is still far behind European countries (Kazancıoğlu, 2012), and HSR cannot meet its costs and is subsidized by public resources (Karakaya and Öztürk, 2021). In the national literature, it has been widely argued that low passenger satisfaction is the major source of the inefficiency of the rail system (Poyraz et al., 2004; Seçilmis et al., 2011), and service quality improvement policies will enhance loyalty in the rail system and support railways’ competition with other modes of transport (Yılmaz and Ari, 2017; Leblebicioğlu and Keskin 2021). The findings of this study are expected to contribute to improving the service quality of the Turkish rail system.

In the literature, several studies have already been conducted to analyze social media content with various machine learning algorithms, topic modeling, and sentiment analysis, and to measure rail service quality from the user's point of view. This paper differs from the previous ones in several aspects. First, the literature mostly focused on urban light rail and subway systems (Collins et al., 2013; Haghighi et al., 2018; Luo and He, 2021; Osorio-Arjona, et al., 2021), and the national rail system has only been addressed in a few studies (Yang and Anwar, 2016; Mogaji and Erkan, 2019; Mishra and Panda, 2022). The time interval in these studies ranges from twenty days to one year, but the temporal variations in passenger demands were not considered at the national level. As discussed by Osorio-Arjona et al. (2021), longer periods are more useful to capture the changes in the demand trends. Examining the trend changes allow decision-makers to assess the impact of policies implemented in the past and help to shape future policies. This is particularly important for Turkey, as it will allow monitoring user reactions to high-speed rail services launched in 2009. This being the case, a dynamic view of policymaking was provided by monitoring temporal changes in passenger satisfaction over the eleven-year period from 2011 to 2021. Second, the conventional and the high-speed rail systems were distinguished, and the differences between passengers’ expectations were defined. Service quality was measured for both conventional and high-speed rail systems and their evolution was monitored over the observation period. Thus, the necessity of adopting separate policies for both systems was revealed. Third, the two outputs of social media analytics (the importance that passengers attach to service dimensions and how they were satisfied) were combined with Importance-Performance Analysis (IPA) to provide clear policy recommendations.

Conclusions found here indicate the most and least important service dimensions and how passengers were satisfied over the observation period. Moreover, we highlighted the differences between conventional and high-speed rail systems and showed how the quality improvement policies should differ. Validation tests with manual coding have demonstrated the accuracy and reliability of classification estimates.

The rest of the paper is organized as follows. The literature review in Section 2 discusses the social media analytics implications in rail services and the historical evolution of Turkish rail policies. Section 3 details the methodology and Section 4 presents the findings. Finally, Section 5 offers a conclusion with policy suggestions.

2. Literature Review

2.1. Social Media Analytics in Rail Service Quality

Social media analytics has been used in rail services for disruption management (Pender et al., 2014; Currie and Muir, 2017; Hosseini et al. 2018; Nisar and Prabhakar, 2018; El-Diraby et al. 2019; Howard, 2020), communication strategy (Cottrill et al. 2017), demand estimation (Liu and Shi, 2019), service quality analysis (Collins et al., 2013; Yang and Anwar, 2016; Haghighi et al., 2018; Narayanaswami, 2018; Mogaji and Erkan, 2019; Myneni and Dandamudi, 2020; Luo and He, 2021; Osorio-Arjona, et al., 2021; Mishra and Panda, 2022), and user opinion capturing (Mathur et al., 2021; Chang et al., 2022) purposes. In line with the purpose and scope of this research, we will focus on the papers examining service quality and user opinion.

Collins et al. (2013) suggested social media (Twitter) as an innovative source for capturing the quality of rail services from the passengers’ perspective. The authors collected 577 tweets and analyzed the daily and hourly variation in the
commuter’s positive and negative sentiments. Similarly, Haghighi et al. (2018) analyzed transit riders’ general sentiments considering daily and hourly variations within a one-week period. Analysis of 403 tweets showed that users were more dissatisfied with transit routes with high ridership and posted more negative tweets on weekends. A word cloud of tweets was also generated to explore the sources of negative sentiments. Yang and Anwar (2016) extended the scope of the previous papers by gathering more tweets (31,008 tweets from 9,428 distinct users) and covering a wider time period (eight months). Moreover, rather than focusing on general sentiments, the authors defined five customer service criteria (reliability, safety, crowdedness, comfort, and convenience) and measured users’ satisfaction across different service dimensions.

Mogaji and Erkan (2019) gathered 2.6 million tweets from 27 train operators in the UK during the one-year period and analyzed passenger satisfaction using sentiment and thematic analysis. Tweets were then assigned to a five-dimensional SERVQUAL scale (reliability, assurance, tangibles, empathy, responsiveness) and passenger satisfaction levels on regional and national routes were determined. Mishra and Panda (2022) used the RAILQUAL scale to assess passenger satisfaction. Yang and Anwar (2016) extended the scope of the previous papers by gathering more tweets (31,008 tweets from 9,428 distinct users) and covering a wider time period (eight months). Moreover, rather than focusing on general sentiments, the authors defined five customer service criteria (reliability, safety, crowdedness, comfort, and convenience) and measured users’ satisfaction across different service dimensions.

Geolocated data provided by Twitter was also used in railway and public transport studies. Lock and Pettit (2020) applied social media analytics to assess the public transport performance (including rail and bus transit) from citizens’ point of view and compare the results with traditional surveys. Using the geolocated tweets, the authors examined the polarity of sentiments, found the relationship between delays and tweet volumes, and compared bus and train performances. Osorio-Arjona et al. (2021) also used geolocated tweets to reveal the spatial distribution of problems through the Madrid Metro network and explore the spatial sources of negative sentiments, such as population, income, density, and connections. The authors found that the punctuality and breakdowns were determined as the main problems, particularly in the early morning of weekdays and at transfer stations or central areas. Luo and He (2021) used Weibo, a Chinese social media platform, as a data source for examining the spatial and temporal aspects of user opinion about metro services in Shenzhen, China. Similar to the previous researches (Collins et al., 2013; Haghighi et al., 2018; Osorio-Arjona et al., 2021), the authors found significant temporal (on hourly and daily basis) and spatial variations among different service dimensions. More recently, Chang et al. (2022) used geotagged tweets to track the footprints of citizens and examine the impact of new transit stations on people’s activity and sentiments.

A brief literature review shows that social media is seen as an important communication tool and contributes to the improvement of customer engagement in the transport industry by creating space for passengers to share their opinions and comments (Saragih and Girsang, 2017; Howard, 2020). Social media helps to detect the problem quickly and to reach secure information for passengers with real-time information flow during planned and unplanned disruptions (Pender et al., 2014; Cottrill et al., 2017). It offers a rich resource for capturing passengers’ complaints and expectations not only in disruptions but also in routine transport operations; hence, provides opportunities for process improvement (Collins et al., 2013; Nisar and Prabhakar, 2018). In addition, social media has been praised as a low-cost (or free) and fast-accessible data source (Haghighi et al., 2018; Lock and Pettit 2020; Osorio-Arjona et al., 2021).

Another issue emphasized in the literature is temporal variations in transport demand. Time-stamp tweets provided by Twitter have been effectively used to capture temporal variations in transit transport demand and sentiments (Collins et al., 2013; Haghighi et al., 2018; Hosseini et al., 2018; El-Diraby et al., 2019; Osorio-Arjona et al., 2021; Luo and He, 2021). In studies examining temporal variation in the literature, the time interval ranges from one week to eleven months. These short-term periods are useful in examining the transport patterns of transit passengers, where homogeneity is important, but longer periods provide more insight into the long-term effects of transport policies (Osorio-Arjona et al., 2021). The current paper aims to fill this research gap by focusing on the last decade of the Turkish rail system. The following section will highlight the historical development of the Turkish railway system and policies.
2.2. Rail Policy in Turkey

The Ottoman Empire, the predecessor of the Republic of Turkey, attached great importance to the rail system since the 1850s and built a total of 8334 km of railways in many parts of the empire, from the Balkans to the Middle East and the Arabian Peninsula, as well as Anatolia (Engin, 2013). These railway investments were largely built and operated by foreign capital due to financial and technical inadequacy and political reasons (Bakırcli, 2013). In this era, railway investments were prioritized for commercial, military, and political reasons, and the highway was seen as a complementary element to the rail system (Çetin et al., 2011).

With the collapse of the empire after World War I, only 4136 km of railway remained within the borders of modern Turkey (UAB, 2021), of which approximately 2400 km were owned by foreign investors (Aydemir, 1993). In the first years of the modern Republic, the railways’ importance was maintained and adopted as the primary means of transportation. Between 1923 and 1940, an average of 190 km of railways were laid annually, the total line length reached 7381 km, and industrial and agricultural centers were connected to markets and ports (Akgünghör and Demirel, 2004; Engin, 2013). Railway lines owned by foreign investors were also purchased and nationalized (Aydemir, 1993). Between 1940 and 1950, with the effect of World War II, railway investments decreased and approximately 300 km of new railway lines could be laid (Engin, 2013; Bakırcli, 2013).

Transport policy changed drastically in the 1950s due to global political changes. Russia’s political attitude after World War II caused Turkey to get closer to Western countries. Soviet Russia’s demand for territory from Turkey in 1945, its desire to establish a military base in the Turkish Straits, and the demand for a revision of the Montreux Convention laid the groundwork for Turkey to form a full alliance with Western countries (Karpät, 1975). One of the economic consequences of this political rapprochement was to benefit from international financial aid, which prioritized road investments rather than railways (Çetin et al., 2011; Polatoğlu, 2021). After this period, investments were shifted to road transport. Many factors such as the exorbitant increase in rail transport, automotive assembly plants established in the country, foreign loans for highway construction, and the liberalization of truck imports also supported this shift (Akgünghör and Demirel, 2004).

On the one hand, many parts of the country gained road infrastructure, the number of motor vehicles increased, and citizens’ mobility opportunities were improved. On the other hand, railways remained in the background and only about 1000 km of new railways were built from 1950 to the 2000s (UAB, 2018). While the major transportation budget was allocated to the railway until 1950, this ratio decreased to 30% in the early 1960s, to 17% in the late 1960s, and remained significantly lower than the highway budget in the following years (Aydemir, 1993). Indeed, this rate has fallen below 10% since the late 1980s (İTÜ, 2005). The decrease in railway investments due to policy changes caused infrastructural deficiencies to become more evident over time, prevented competition with the highway, and caused losses since the 1960s (Akgünghör and Demirel, 2004). Today, while the average railway length per 1000 km² is 48 km in the European Union, it is only 13 km in Turkey (TCDD, 2020). This situation reflected negatively on the railway traffic. The railway rate in domestic passenger and freight transportation was 42% and 55% in 1950, decreased to 24% and 48% in 1960, and to 7.6% and 21.6% in 1970, respectively (İTÜ, 2005). This sharp decline continued in the following years. Today, the share of the railway in passenger and freight transportation is only 1.2% and 4.8%, respectively (ÇSB, 2018), while its’ share in international trade is less than 1% (TÜİK, 2021).

Rail policy has changed and privileged again after the late 2000s (Zeybek, 2018). For the first time after the 1950s, the budget allocated to the railway was more than the highway in 2010, and this privileged share has continued until today. In addition to the renewal of existing lines, the Baku-Tbilisi-Kars railway and Bosphorus Rail Tube Crossing projects were completed in this period. More importantly, high-speed train investments have started in 2003. With these investments, the capital Ankara was first connected to Eskişehir by high-speed rail in 2009, with Konya in 2011, and Istanbul in 2014. A total of 1213 km of high-speed rail lines have been built while the total railway length reached 12,803 km, including junction and station lines. The number of high-speed rail passengers increased from 942 thousand in 2009 to 8.2 million in 2019. Figure 1 presents the conventional and high-speed railways of Turkey.
The operating rights of the entire railway system was belonging to TCDD (Turkish State Railways), a state economic enterprise, since 1953, but the liberalization law came into force in 2013. With this law, the private sector was allowed to operate its own vehicles on existing lines, and the monopoly and privileged legal status of TCDD on railway transportation were ended (Göçgün, 2018; Zeybek, 2018). Thus, it is aimed to increase the competition of TCDD with other operators, enhance the service quality, and increase the competitiveness of the railway against other modes of transport (Kurt, 2017; Cebeci et al., 2021). However, the private sector’s participation in railway transportation is still far from the expected level due to the infrastructural inadequacies and the absence of a suitable competitive environment.

3. Methodology

Semi-automatic analysis of big text data is called text mining and it has been widely used in many fields such as summarizing, topical modeling, and sentiment analysis. Although the stages of the method differ in terms of subject, analyst preferences, and data characteristics, the common steps can be listed as data collection, data preprocessing, feature selection, and information extraction (Sharda et al., 2020). In this study, we followed a five-stage method. Figure 2 presents the outline of each stage. Python programming language was used in the analysis. The following subsections detail the methodology.

3.1. Data Collection

The data used in text mining is called corpus and any text document (such as article, customer comment, feedback, social media message) can be added to this structure. In this study, user-generated social media messages on Twitter were used to create a comprehensive corpus. Here, tweets related to the Turkish rail services were collected by scraping. At this stage, 52 keywords containing the derivatives of the word “train” in the Turkish language were used and 2,383,747 tweets were gathered posted between 2011 and 2021. The final dataset consists of tweet texts, numbers of quotes, likes, retweets, and the dates of posting.

3.2. Data Preprocessing

In a data analysis study, the most time is spent on data preprocessing (Sharda et al., 2020). This process aims to access high-quality data and meet the method requirements. Analyzes made with poor quality datasets with various errors and deficiencies can not achieve high performance (Cebeci, 2020). This phenomenon is called Garbage-in / Garbage-out, and five systematic steps should be performed in the text preprocessing process:

- **Data Cleaning**: Repetitive tweets and tweets containing very few characters in the corpus are removed in this step. This step should be performed before the following steps.
- **Tokenizing**: At this stage, text documents are split into their smallest parts with the help of whitespaces between words.
- **Standardization**: This process simplifies the corpus by removing some tokens that will not contribute to the analysis. First, uppercase letters in the text are converted to lowercase letters. Punctuation marks, numbers, and unnecessary spaces are also removed (Haddi et al., 2013). Apart from other text structures, social media data may contain different elements such as hashtags, special signs (@, #), and links, and they should be removed. Abbreviations and emojis can be optionally converted or cleared (Krouska et al., 2016). In this study, emojis and abbreviations were translated into expressions using a dictionary created specifically for the current research. Thus, data loss was minimized and the performance of sentiment analysis was improved.
- **Filtering**: Removing stop words and low-frequency words from the corpus are two basic filtering techniques. Words that do not contain emotion and are not expected to contribute to the analysis (conjunction, preposition, etc.) are called stop words. Using the stop word lists created for each language, unnecessary words were cleaned from the corpus. In this study, filtering was performed using the Turkish stop word list included in the “nltk” package. In addition, low-frequency
words are not expected to make a considerable contribution to the analysis. Nevertheless, low-frequency words may increase the term-document matrix size and can negatively affect the analysis performance. Hence, we controlled the size of the matrix by eliminating the words that occur ten times or less.

- **Linguistic Preprocessing**: Linguistic preprocessing is essential in studies using natural language processing approaches. First, the words are transformed to the simplest form by removing the suffixes. Here, stemming and lemmatizing are the two main methods. The tasks of the words in the sentence (part-of-speech tagging) should be determined next (Anandarajan et al., 2019).

The size of the initial corpus has reduced after the preprocessing phase and the new corpus consists of 2,015,104 high-quality cleaned tweets.

### 3.3. Digitization

The preprocessed corpus needs to be digitized to make it ready for analysis processes. Here, the texts must first be translated into a matrix structure called a vector-space model. This model is then extended with numerical selection values. Although many different methods can be used at this stage (such as GINI and Entropy), we applied TF-IDF (Term Frequency-Inverse Document Frequency), which is the most preferred method (Aizawa, 2003). TF-IDF measures how much information a word can carry. In this sense, similar to the Entropy, TF-IDF quantifies the importance of each word for both the document and the general corpus. At the end of these processes, a term-document matrix is composed, which presents terms and documents together with TF-IDF values.

### 3.4. Document Classification

Document classification is performed by assigning unstructured text format to predetermined classes. This method is a supervised learning approach and a learning dataset created by the user is needed to make class predictions. There are different forms of document classification approaches. Binary or multiple classifications can be made according to the number of classes contained in the output variable, as well as hard and soft classification approaches can be used whether a document is allowed to be included in more than one class.

In this study, firstly, document classification was used to determine whether the tweets are related to conventional rail or high-speed rail. In this sense, 1,400 tweets were read and manually labeled for high-speed rail estimation. In order to obtain a balanced learning set, half of the 1,400 tweets are labeled to high-speed rail and the other half to conventional rail. Here, binary classification is made, which allows assigning the tweet to one of the two groups, that is, hard classification. We applied 28 different machine learning methods and found that the Gradient Boosting reached the highest prediction performance of 0.9288. The rest of the corpus was estimated with Gradient Boosting and 32.75% of the tweets were labeled as high-speed rail.

Assigning the documents (tweets in this study) to predetermined topics is a common practice in text mining to better evaluate the high-dimensional corpus (Hu et al., 2014). This stage is called thematic evaluation or topic modeling. Due to the large corpus size, this assignment can not be done manually, and class prediction with supervised machine learning approaches is widely used (Pandur et al., 2020). In this paper, topic modeling approaches were used to assign tweets to each predetermined service quality dimension.

The five-dimensional SERVQUAL scale (responsiveness, reliability, tangibles, assurance, empathy) proposed by Parasuraman et al. (1988) has been widely used in rail service quality analysis (Awasthi et al., 2011; Çelik et al., 2014; Khorshidi et al., 2016; Mogaji and Erkan, 2019). Cavana et al. (2007) has extended this scale by adding three more dimensions specific to rail services: comfort, connection, and convenience. In this study, in addition to these dimensions, safety&security is also included in the model. Hence, a total of nine service quality dimensions were determined. Appendix A presents the dimensions and sub-dimensions of each service quality.
A machine learning dataset was created next by manually coding the tweets. At this stage, it was aimed to have at least 1000 tweets in the learning dataset for each service dimension. A total of 10,104 tweets were labeled but some records were deleted at the last check. Table 1 presents the final number of tweets labeled for each dimension.

**Table 1. Number of tweets in the machine learning set**

<table>
<thead>
<tr>
<th>Service Dimensions</th>
<th>Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>1221</td>
</tr>
<tr>
<td>Empathy</td>
<td>787</td>
</tr>
<tr>
<td>Reliability</td>
<td>1286</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>1005</td>
</tr>
<tr>
<td>Tangibles</td>
<td>1050</td>
</tr>
<tr>
<td>Comfort</td>
<td>1304</td>
</tr>
<tr>
<td>Connection</td>
<td>1078</td>
</tr>
<tr>
<td>Convenience</td>
<td>1137</td>
</tr>
<tr>
<td>Safety &amp; Security</td>
<td>1228</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>9797</strong></td>
</tr>
</tbody>
</table>

Topical modeling is a multi-classification-based approach. However, the increase in output classes worsens the performance of prediction models. To avoid this problem and improve the performance of prediction models, in this study, separate learning models were established for each column. In these models, a balanced learning dataset was created by adding as many unlabeled values as the number of labeled values each time. For example, for the comfort dimension, in addition to 1304 "1" labeled (comfort) values, 1304 "0" labeled (non-comfort) values were randomly selected from the rest of the learning dataset. Soft classification is achieved when the estimated values obtained for each column are combined. Because, thanks to independent learning models, a document can be labeled by more than one model. 28 different machine learning models were performed and the method with the best prediction capability was selected for each service dimension. These models and their prediction accuracies are given in Table 2.

**Table 2. Prediction performances in learning datasets**

<table>
<thead>
<tr>
<th>Service Dimensions</th>
<th>Learning Algorithms</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>Gradient Boosting</td>
<td>0.9339</td>
</tr>
<tr>
<td>Empathy</td>
<td>Bagging</td>
<td>0.9298</td>
</tr>
<tr>
<td>Reliability</td>
<td>Logistic Regression</td>
<td>0.9453</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Gradient Boosting</td>
<td>0.9644</td>
</tr>
<tr>
<td>Tangibles</td>
<td>Gradient Boosting</td>
<td>0.9559</td>
</tr>
<tr>
<td>Comfort</td>
<td>Bagging</td>
<td>0.9561</td>
</tr>
<tr>
<td>Connection</td>
<td>Gradient Boosting</td>
<td>0.9333</td>
</tr>
<tr>
<td>Convenience</td>
<td>Bagging</td>
<td>0.9576</td>
</tr>
<tr>
<td>Safety &amp; Security</td>
<td>Bagging</td>
<td>0.9763</td>
</tr>
</tbody>
</table>
As presented in Table 2, the best predictions for the nine service dimensions were obtained by Bagging, Gradient Boosting, and Logistic Regression. These models were used for prediction for the remainder of the corpus, and approximately 29% (577,414) of 2,015,104 tweets were labeled into at least one dimension. This labeled dataset was used in the following information extraction stage.

Up to this stage, the dataset has been assigned to each specified dimension. However, it has not been determined yet whether the user’s opinions are negative or positive. Sentiment analysis is the most preferred text mining method, which uses machine learning and natural language processing approaches to determine the polarity (negative or positive) of user-generated content (Poria et al., 2018), including social media messages (Pozzi et al., 2017; Yue et al., 2019; Liu, 2020). Hence, we preferred sentiment analysis in this research for the information extraction process.

BERT (Bidirectional Encoder Representations from Transformers), a natural language processing-based sentiment analysis model developed by Google has been widely used in sentiment analysis (Alaparthi and Mishra, 2021) The BERT model uses the transformer method, which is a variant of the deep feedforward neural networks, and evaluates each sentence bidirectionally (right to left and left to right). In this study, we applied TurkishBERT, which is the Turkish version of this model. Recent researches demonstrated the high predictive capability of TurkishBERT (Aras et al., 2021; Ozturk et al., 2021).

Before proceeding to the final information extraction stage, 500 tweets were manually coded and their performance was evaluated in order to ensure the validity of the classification estimates. For this purpose, accuracy, precision, Recall, and F1 values were calculated for each classification task in the validation dataset and presented in Table 3.

Table 3. Validation results of classification tasks

<table>
<thead>
<tr>
<th>Topic Modelling (SERVQUAL Dimensions)</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>0.966</td>
<td>0.871</td>
<td>0.857</td>
<td>0.864</td>
</tr>
<tr>
<td>Empathy</td>
<td>0.954</td>
<td>0.745</td>
<td>0.761</td>
<td>0.753</td>
</tr>
<tr>
<td>Reliability</td>
<td>0.916</td>
<td>0.906</td>
<td>0.913</td>
<td>0.909</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>0.970</td>
<td>0.788</td>
<td>0.765</td>
<td>0.766</td>
</tr>
<tr>
<td>Tangibles</td>
<td>0.984</td>
<td>0.857</td>
<td>0.667</td>
<td>0.750</td>
</tr>
<tr>
<td>Comfort</td>
<td>0.974</td>
<td>0.824</td>
<td>0.800</td>
<td>0.812</td>
</tr>
<tr>
<td>Connection</td>
<td>0.942</td>
<td>0.817</td>
<td>0.827</td>
<td>0.822</td>
</tr>
<tr>
<td>Convenience</td>
<td>0.962</td>
<td>0.862</td>
<td>0.820</td>
<td>0.840</td>
</tr>
<tr>
<td>Safety&amp;Security</td>
<td>0.984</td>
<td>0.897</td>
<td>0.839</td>
<td>0.867</td>
</tr>
<tr>
<td>CR / HSR Classification</td>
<td>0.950</td>
<td>0.909</td>
<td>0.927</td>
<td>0.918</td>
</tr>
<tr>
<td>Sentiment Classification</td>
<td>0.934</td>
<td>0.941</td>
<td>0.956</td>
<td>0.948</td>
</tr>
</tbody>
</table>

Table 3 shows that the accuracy of each classification model established for topic modeling, CR/HSR distinction, and sentiment polarities are quite high and the information inferences to be made with the predictions are valid.

3.5. Information Extraction

In the information extraction stage, the results of the topical modeling and the sentiment classifications were combined and the user-oriented evaluations of the SERVQUAL dimensions were evaluated with sentiment analysis. The change in the citizens’ perceptions regarding service quality over the years has been observed with the trend analysis. The findings were also supported by quotes, retweets, and likes, which reflect users’ reactions to social media posts. Finally, the importance-performance analysis was used to transform these findings into actionable policies.
4. Findings

4.1. Emphasis on Service Quality Dimensions

A total of 577,414 tweets were included in the analysis. About 67% of these tweets are related to conventional rail (CR) and 33% to high-speed rail (HSR). These values are very close to the actual CR and HSR passenger rates, showing that the machine learning-based technique used for document classification has a high estimation success\(^1\). As presented in Figure 3, the number tweets decreased between 2014 and 2017 but increased again as of 2018.

Much emphasis on a service quality dimension in social media content indicates users attached higher importance. Tables 4 and 5 show how much each service quality dimension has been emphasized in social media messages over the years.

Reliability is the most mentioned service dimension in social media messages for both CR and HSR, and its share in total mentions has been increasing over the years. Its’ quite high ratio in the total emphasis (49% and 54%, respectively) indicates that it is the most vital service dimension for citizens. This finding supports the previous study on Turkish users, which revealed that the timeliness of train movements is the most important service dimension (Seçilmiş et al., 2011). Reliability is followed by connection, where its share in HSR is slightly higher (15% and 17%, respectively). Assurance, empathy, and tangibles have similar rates in both systems. While the rate of assurance is 11%, tangibles remained at just 2%. This is also the lowest value among all service dimensions. In addition, HSR passengers gave a little more attention to comfort in their social media posts (8% and 11%, respectively).

In addition to these similarities, significant differences were also detected between the priorities of CR and HSR passengers. The most striking difference was observed in the responsiveness. Responsiveness was emphasized in only 2% of HSR-related messages, while this rate rises to 12% in CR. Similarly, only 3% of HSR passengers posted about safety&security, compared to 8% in CR. This shows that responsiveness and safety&security issues are discussed more by CR passengers. On the other hand, HSR passengers care more about convenience (12%), compared to 7% in CR.

Overall, these findings are important as they exhibit the similarities and differences of the most discussed service dimensions by CR and HSR passengers. While there are no significant differences between assurance, empathy, reliability, tangibles, connection, and comfort, CR passengers share more messages about responsiveness and safety&security. HSR passengers, however, discussed more about convenience. This finding indicates which service dimensions gain more importance in CR and HSR systems.

Table 4. Emphasis on conventional rail

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>7%</td>
<td>8%</td>
<td>11%</td>
<td>9%</td>
<td>10%</td>
<td>11%</td>
<td>11%</td>
<td>14%</td>
<td>13%</td>
<td>11%</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>Empathy</td>
<td>7%</td>
<td>5%</td>
<td>4%</td>
<td>5%</td>
<td>6%</td>
<td>8%</td>
<td>6%</td>
<td>6%</td>
<td>6%</td>
<td>4%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>Reliability</td>
<td>45%</td>
<td>48%</td>
<td>53%</td>
<td>54%</td>
<td>49%</td>
<td>46%</td>
<td>44%</td>
<td>42%</td>
<td>46%</td>
<td>50%</td>
<td>56%</td>
<td>49%</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>11%</td>
<td>10%</td>
<td>7%</td>
<td>8%</td>
<td>11%</td>
<td>18%</td>
<td>17%</td>
<td>20%</td>
<td>10%</td>
<td>10%</td>
<td>9%</td>
<td>12%</td>
</tr>
<tr>
<td>Tangibles</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>2%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>Comfort</td>
<td>10%</td>
<td>8%</td>
<td>7%</td>
<td>10%</td>
<td>9%</td>
<td>10%</td>
<td>10%</td>
<td>8%</td>
<td>10%</td>
<td>8%</td>
<td>7%</td>
<td>8%</td>
</tr>
<tr>
<td>Connection</td>
<td>19%</td>
<td>19%</td>
<td>14%</td>
<td>12%</td>
<td>13%</td>
<td>11%</td>
<td>13%</td>
<td>14%</td>
<td>16%</td>
<td>18%</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>Convenience</td>
<td>10%</td>
<td>8%</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
<td>7%</td>
<td>8%</td>
<td>7%</td>
<td>9%</td>
<td>8%</td>
<td>6%</td>
<td>7%</td>
</tr>
<tr>
<td>Safety&amp;Security</td>
<td>8%</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
<td>8%</td>
<td>5%</td>
<td>6%</td>
<td>8%</td>
<td>11%</td>
<td>8%</td>
<td>9%</td>
<td>8%</td>
</tr>
</tbody>
</table>
Note: Due to the multi-labeling approach, the sum of the ratios is more than 100%.

Table 5. Emphasis on high-speed rail

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>6%</td>
<td>7%</td>
<td>10%</td>
<td>13%</td>
<td>11%</td>
<td>12%</td>
<td>14%</td>
<td>15%</td>
<td>13%</td>
<td>10%</td>
<td>8%</td>
<td>11%</td>
</tr>
<tr>
<td>Empathy</td>
<td>5%</td>
<td>4%</td>
<td>3%</td>
<td>5%</td>
<td>5%</td>
<td>7%</td>
<td>5%</td>
<td>5%</td>
<td>6%</td>
<td>4%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>Reliability</td>
<td>53%</td>
<td>51%</td>
<td>49%</td>
<td>53%</td>
<td>49%</td>
<td>50%</td>
<td>51%</td>
<td>51%</td>
<td>53%</td>
<td>56%</td>
<td>65%</td>
<td>54%</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>3%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Tangibles</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>2%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>3%</td>
<td>3%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Comfort</td>
<td>11%</td>
<td>9%</td>
<td>9%</td>
<td>9%</td>
<td>14%</td>
<td>12%</td>
<td>11%</td>
<td>12%</td>
<td>12%</td>
<td>10%</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td>Connection</td>
<td>17%</td>
<td>22%</td>
<td>24%</td>
<td>17%</td>
<td>15%</td>
<td>15%</td>
<td>17%</td>
<td>16%</td>
<td>16%</td>
<td>18%</td>
<td>14%</td>
<td>17%</td>
</tr>
<tr>
<td>Convenience</td>
<td>13%</td>
<td>16%</td>
<td>14%</td>
<td>14%</td>
<td>13%</td>
<td>14%</td>
<td>14%</td>
<td>13%</td>
<td>11%</td>
<td>10%</td>
<td>8%</td>
<td>12%</td>
</tr>
<tr>
<td>Safety &amp; Security</td>
<td>4%</td>
<td>2%</td>
<td>2%</td>
<td>3%</td>
<td>3%</td>
<td>4%</td>
<td>3%</td>
<td>5%</td>
<td>4%</td>
<td>4%</td>
<td>3%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Note: Due to the multi-labeling approach, the sum of the ratios is more than 100%.

Users' reactions (retweets, likes, and quotes) to service dimensions support emphasis rates. Table 6 shows the ratio of total retweets, likes, and quotes for each service quality dimension. It is interesting to note that the rates of the retweet, like, and quotes received by the service quality dimensions are largely similar to the rate of their emphasis on social media contents, both for CR and HSR systems.

Overall, reliability creates the highest social impact and it is followed by connection and assurance while empathy and tangibles are the least. This means that passengers respond more to messages about timing, the accuracy of reservations, access, parking availability, and connectivity with other modes of transport, and less to empathy, the appearance of staff, and physical equipment. The most striking finding here is users' responses to responsiveness posts. Compared to HSR users, CR passengers responded significantly more to prompt service, availability and willingness of staff in handling their requests, suggestions, and complaints. HSR users are more responsive to comfort and convenience posts, while CR passengers are more concerned with safety & security. This clearly reveals the difference in expectations of CR and HSR users and the safety concerns of CR passengers.

Table 6. Social impacts of each service quality dimension
Readers should recall that the values discussed in this section do not show the users’ satisfaction with the service dimensions but how much each dimension has been discussed on social media. The level of satisfaction for each dimension will be measured by sentiment analysis in the next section.

4.2. Level of Service Quality

Tables 7 and 8 display the positivity rates of tweets for CR and HSR services, respectively. The results are quite striking and support the view that Twitter content mostly tends to be negative (Collins et al., 2013; Yang and Anwar, 2016; Haghighi et al., 2018; El-Diraby et al. 2019; Whatman, 2018). As Schweitzer (2014) discussed, Twitter comments about transport services are mostly more negative compared to other public services. HSR received much higher satisfaction rates, but what is striking is the extremely low satisfaction rate of the conventional rail. Compared to the study conducted on the British railway system (Mogaji and Erkan, 2019), we found that the user satisfaction reflected in social media is extremely low in the Turkish conventional rail system (19% on average). This finding clearly indicates that users are mostly dissatisfied with the quality of conventional rail services. On the other side, half of the tweets about HSR are positive and contain expressions of satisfaction. The superiority of HSR is an expected result as it offers services with newer, faster, and more comfortable equipment and a higher level of security at the stations. Thus, the clear segregation of the results into the CR and HSR scores demonstrates the robustness of the analysis results.

Although the satisfaction rate of HSR is much higher than CR, the other half of the social media comments regarding HSR contain negative judgments. Nevertheless, a relative interpretation of the results would provide important insights into which service dimensions are better or worse. Despite the large difference in satisfaction rates, some similarities were found between the two rail systems. Both CR and HSR received the highest rate of positive comments from the comfort dimension (28% and 56%, respectively), and comfort satisfaction tends to increase over the years. Empathy, convenience, reliability, connection, and safety&security demonstrated a higher level of service than assurance and tangibles in both systems. Tangibles and convenience satisfaction tended to increase in both CR and HSR. Empathy satisfaction increased from 14% to 22% in CR but decreased from 55% to 51% in HSR over the observation period. While HSR has improved the reliability, it has experienced a decline in assurance satisfaction. Finally, responsiveness has the lowest satisfaction rate among all service dimensions for both CR and HSR (9% and 30%, respectively). This finding strongly supports the previous studies on Turkish rail passengers, where responsiveness was the least satisfied service dimension in CR (Poyraz et al. 2004; Seçilmiş et al. 2011) and HSR (Altan and Ediz, 2016). This demonstrates the inability of staff to respond to users’ requests and the
unwillingness of existing staff to assist passengers. Fortunately, compared to the first year of the observation period, the satisfaction rate for this service dimension has improved significantly.

Table 7. Level of satisfaction in conventional rail

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>16%</td>
<td>28%</td>
<td>13%</td>
<td>15%</td>
<td>18%</td>
<td>18%</td>
<td>23%</td>
<td>12%</td>
<td>14%</td>
<td>18%</td>
<td>17%</td>
<td>17%</td>
</tr>
<tr>
<td>Empathy</td>
<td>14%</td>
<td>18%</td>
<td>23%</td>
<td>27%</td>
<td>22%</td>
<td>26%</td>
<td>24%</td>
<td>20%</td>
<td>22%</td>
<td>23%</td>
<td>22%</td>
<td>22%</td>
</tr>
<tr>
<td>Reliability</td>
<td>21%</td>
<td>22%</td>
<td>27%</td>
<td>20%</td>
<td>20%</td>
<td>24%</td>
<td>24%</td>
<td>19%</td>
<td>21%</td>
<td>21%</td>
<td>22%</td>
<td>22%</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>4%</td>
<td>8%</td>
<td>10%</td>
<td>7%</td>
<td>6%</td>
<td>4%</td>
<td>6%</td>
<td>6%</td>
<td>16%</td>
<td>14%</td>
<td>17%</td>
<td>9%</td>
</tr>
<tr>
<td>Tangibles</td>
<td>11%</td>
<td>17%</td>
<td>16%</td>
<td>11%</td>
<td>29%</td>
<td>17%</td>
<td>15%</td>
<td>14%</td>
<td>15%</td>
<td>16%</td>
<td>17%</td>
<td>16%</td>
</tr>
<tr>
<td>Comfort</td>
<td>19%</td>
<td>21%</td>
<td>33%</td>
<td>38%</td>
<td>32%</td>
<td>26%</td>
<td>30%</td>
<td>28%</td>
<td>24%</td>
<td>26%</td>
<td>27%</td>
<td>28%</td>
</tr>
<tr>
<td>Connection</td>
<td>17%</td>
<td>17%</td>
<td>20%</td>
<td>19%</td>
<td>19%</td>
<td>19%</td>
<td>22%</td>
<td>12%</td>
<td>14%</td>
<td>15%</td>
<td>18%</td>
<td>17%</td>
</tr>
<tr>
<td>Convenience</td>
<td>15%</td>
<td>17%</td>
<td>17%</td>
<td>15%</td>
<td>17%</td>
<td>16%</td>
<td>22%</td>
<td>21%</td>
<td>21%</td>
<td>17%</td>
<td>19%</td>
<td>18%</td>
</tr>
<tr>
<td>Safety&amp;Security</td>
<td>13%</td>
<td>17%</td>
<td>12%</td>
<td>15%</td>
<td>27%</td>
<td>23%</td>
<td>21%</td>
<td>15%</td>
<td>16%</td>
<td>18%</td>
<td>21%</td>
<td>18%</td>
</tr>
</tbody>
</table>

Table 8. Level of satisfaction in high-speed rail

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Assurance</td>
<td>47%</td>
<td>42%</td>
<td>46%</td>
<td>56%</td>
<td>48%</td>
<td>41%</td>
<td>46%</td>
<td>42%</td>
<td>41%</td>
<td>40%</td>
<td>40%</td>
<td>44%</td>
</tr>
<tr>
<td>Empathy</td>
<td>55%</td>
<td>49%</td>
<td>47%</td>
<td>51%</td>
<td>54%</td>
<td>69%</td>
<td>45%</td>
<td>49%</td>
<td>49%</td>
<td>54%</td>
<td>51%</td>
<td>52%</td>
</tr>
<tr>
<td>Reliability</td>
<td>49%</td>
<td>49%</td>
<td>50%</td>
<td>51%</td>
<td>53%</td>
<td>54%</td>
<td>53%</td>
<td>52%</td>
<td>47%</td>
<td>51%</td>
<td>52%</td>
<td>51%</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>27%</td>
<td>30%</td>
<td>21%</td>
<td>26%</td>
<td>29%</td>
<td>20%</td>
<td>25%</td>
<td>42%</td>
<td>34%</td>
<td>39%</td>
<td>42%</td>
<td>30%</td>
</tr>
<tr>
<td>Tangibles</td>
<td>36%</td>
<td>42%</td>
<td>44%</td>
<td>42%</td>
<td>56%</td>
<td>51%</td>
<td>40%</td>
<td>35%</td>
<td>38%</td>
<td>36%</td>
<td>42%</td>
<td>42%</td>
</tr>
<tr>
<td>Comfort</td>
<td>50%</td>
<td>50%</td>
<td>55%</td>
<td>55%</td>
<td>45%</td>
<td>67%</td>
<td>56%</td>
<td>60%</td>
<td>55%</td>
<td>58%</td>
<td>60%</td>
<td>56%</td>
</tr>
<tr>
<td>Connection</td>
<td>44%</td>
<td>54%</td>
<td>60%</td>
<td>47%</td>
<td>47%</td>
<td>46%</td>
<td>55%</td>
<td>46%</td>
<td>44%</td>
<td>44%</td>
<td>44%</td>
<td>48%</td>
</tr>
<tr>
<td>Convenience</td>
<td>47%</td>
<td>69%</td>
<td>62%</td>
<td>47%</td>
<td>50%</td>
<td>53%</td>
<td>46%</td>
<td>54%</td>
<td>45%</td>
<td>49%</td>
<td>50%</td>
<td>52%</td>
</tr>
<tr>
<td>Safety&amp;Security</td>
<td>47%</td>
<td>46%</td>
<td>41%</td>
<td>42%</td>
<td>48%</td>
<td>47%</td>
<td>44%</td>
<td>43%</td>
<td>45%</td>
<td>44%</td>
<td>47%</td>
<td>45%</td>
</tr>
</tbody>
</table>

Figure 4 presents the differences between the overall satisfaction levels of CR and HSR. While HSR provided superior services in all dimensions, it outperformed CR particularly in convenience (34%), connection (31%), and empathy (30%). This indicates that the most distinguishing feature of HSR compared to CR is the ease of booking, travel information access, boarding-alighting, and support for disabled users. It also offers better connectivity in terms of access, parking, and integration with other modes of transport. Moreover, in HSR services, user needs are clearly better understood and due attention is paid to the interests of users. On the other hand, the most similar service offered by both systems is responsiveness (21%), which indicates HSR made less difference in the staff’s approach to the user.

4.3. Service Quality Improvement
The importance-performance analysis introduced by Martilla and James (1977) is widely used to visualize the relationship between the importance attributed to service dimensions and the perceived service level, and to identify prior improvement areas. In this simple method, the perceived service level and the importance of service dimensions are displayed in a two-dimensional graph, including mean importance and performance ratings. Thus, service dimensions are divided into one of four quadrants according to its performance and perceived service quality: high importance – high satisfaction (keep up the good work), low importance – high satisfaction (possible overkill), low importance – low satisfaction (low priority), and high importance – low satisfaction (concentrate here) (Martilla and James, 1977).

In the case of rail services, these quadrants refer to the following policy recommendations for decision-makers:

- **Quadrant 1 – Keep up the good work**: Rail passengers attach great importance to these service dimensions and are satisfied with the level of service provided. This success must be maintained and improved to compete with other modes of transport and enhance passenger loyalty.

- **Quadrant 2 – Possible overkill**: Passengers are satisfied with the level of service offered, but these services are of little importance to them. Nevertheless, successful practices must be sustained in the public services.

- **Quadrant 3 – Low priority**: Passengers are not satisfied with the level of service offered, but they do not worry much about it. If resources are limited, the focus should be placed on other services at Quadrants I and IV. Since the rail service is provided by the state and is a public service in Turkey, these services should not be ignored and ways of improvement should be sought without worsening the priority services.

- **Quadrant 4 – Concentrate here**: Rail passengers attach great importance to these services, but perceived satisfaction is low. This quadrant includes the service dimensions that need the most urgent improvement, as they can make a major contribution to the user satisfaction and competitiveness of the rail system.

This study determined the importance of service dimensions according to their frequency in social media content. That is, we assumed that much emphasis on a service quality dimension indicates higher importance (see Tables 4 and 5). The level of satisfaction was measured by sentiment analysis and obtained from Tables 7 and 8. In fact, it is clear that the satisfaction levels of all services are very low and they all need improvement, especially in the CR system. Nevertheless, the importance-performance structure provides important insights to decision-makers about which services they should prioritize.

Figures 5 and 6 display the importance-satisfaction of the CR and HSR systems, respectively, and visualize the change from 2011 to 2021. In both figures, the importance value of reliability is excluded because it deviates excessively from the others. It was observed in the figures that the average satisfaction of both systems increased at the end of the observation period and the satisfaction lines moved upward. In particular, CR satisfaction exhibited a significant improvement from 14% to 20%.

In the CR system, reliability and connection took place in Quadrant I in 2011. Reliability managed to stay in the first quadrant in 2021, thanks to its high importance and increased satisfaction. Passengers care about reliability and satisfaction level is relatively high, and its’ continuity should be ensured. However, the connection shifted to Quadrant IV in 2021. Although there was a slight increase in its’ satisfaction, this increase remained below the general average. Moreover, its’ importance has also decreased. Therefore, the connection is positioned in the fourth quarter, which indicates prior improvement areas. This implies that the connection is of higher importance to users (despite the recent decrease) but the perceived performance is low. Due to its high importance, improvements in connection will make more sense for users and may contribute more to the competitiveness of conventional railways.

Satisfaction levels of empathy and comfort increased by 8% in 2021 and both service dimensions remained in Quadrant II. The operator put much effort into developing empathy and comfort, and offered a relatively higher level of service, but users
do not perceive this as an important service dimension. Convenience is positioned in Quadrant II in 2021 despite small increase in satisfaction rate.

Tangibles, assurance, responsiveness, and safety&security have fallen to Quadrant III in 2011, which refers to low satisfaction and low importance. It was observed that these four service dimensions showed great improvement in 2021. In particular, the satisfaction rate of responsiveness has made a great leap from 4% to 17%. In the same period, tangibles improved by 6% and safety&security by 8%. Thus, safety&security is positioned in Quadrant II in 2021. ASR, on the other hand, has shown a horizontal development, that is, the importance attributed to it has increased. The conclusion to be drawn from this is not that operator should ignore these services and focus on the prior ones listed in Quadrants I and IV. TCDD should be expected to maintain and improve these services as a service-oriented public institution. Indeed, the fact that the safety&security is less discussed on social media does not mean that this dimension can be neglected. But, it is clear that passengers care more about reliability and connection than the appearance of personnel, station, and trains (tangibles), prompt response, and the staff attitudes. Thus, decision-makers should ensure that resources are allocated primarily to services that are more important to users.

Similar to CR, reliability remained important in the HSR system and was positioned in Quadrant I. HSR passengers attach more importance to this service and are relatively satisfied. Convenience was included in Quadrant I in 2011, but its importance to users has decreased significantly in recent years. It shifted to Quadrant II with a horizontal movement in 2021. Due to the possible improvements made in travel information sharing, ticketing, and reservation, the issue may have been less discussed on social media recently and occupied less place on the agenda of the passengers. On the other hand, the connection is still located in Quadrant IV and has been identified as a priority dimension to focus on, although it has been less discussed on social media recently. Decreased service satisfaction of assurance and stability of safety&security caused a regression from Quadrant II to Quadrant III. Tangibles and responsiveness remained in Quadrant III, despite the rising service quality. In particular, responsiveness satisfaction achieved a significant improvement from 27% to 42%. During the analysis period, an increase in the service quality of comfort and a decrease in empathy were observed. Yet both are positioned in Quadrant II in 2021.

[1] The proportion of CR and HSR passengers was 66.7% and 33.3% in 2018, 68.2% and 31.8% in 2019, respectively (TCDD, 2020).

5. Conclusion And Suggestions

User-generated social media content can provide operators with important insights into measuring and improving the quality of rail services, while time-stamp tweets provided by Twitter highlight the temporal variations in rail demand and user sentiments. In this study, the evolution of the quality of Turkish conventional and high-speed rail services from 2011 to 2021 is measured and monitored with social media analytics. The most important service dimensions were determined, users’ satisfaction with these services was estimated, and temporal changes in passenger demands and sentiments were recorded. Priority improvement areas were also identified for conventional and high-speed rail systems and policy recommendations were presented.

The results showed that significant temporal variations in rail demand were experienced over eleven years. During this period, remarkable changes were recorded in the level of importance attributed to some service dimensions and in the perceived service quality levels. Considering these variations, a dynamic view of policymaking has been achieved.
The main findings are as follows. The satisfaction level of conventional rail services was extremely low, and as expected, HSR provided a much higher level of satisfaction to its users, particularly in convenience, connection, and empathy. Nevertheless, during the observation period, CR's service satisfaction rates improved relatively more compared to HSR, particularly in empathy, comfort, and safety/security. On the other hand, HSR's empathy and assurance (courtesy and knowledge of staff, guiding information, notification of changes in schedule and delay) satisfaction decreased remarkably in 2021. The negative comments were notable and potential dimensions open for improvement are listed. In both rail systems, the highest user satisfaction was measured in the comfort dimension (seating, air conditioning, noise isolation, smoothness of the ride, entertainment systems, meal services, etc.), which indicates that the citizens perceive the railway as a comfortable alternative for transport. This finding supports Osorio-Arjona et al. (2021), who found that comfort received the highest positive sentiment. The results also drew attention to the difference in expectations of CR and HSR passengers. Responsiveness and safety/security make more sense for CR passengers while convenience for HSR users.

Improvements in responsiveness (willingness and ability of the staff to provide prompt response to passengers' suggestions and complaints) also deserve special emphasis. While satisfaction with this service dimension was extremely low in 2011, it achieved the highest improvement during the observation period. Despite this improvement in recent years, the satisfaction rate is still relatively low and TCDD needs to develop policies, including staff training, to provide prompt responses to user demands.

Reliability (timeliness, frequency, the accuracy of reservation and ticketing, etc.) is by far the most important service dimension. More than half of social media posts are related to reliability, and this ratio increased significantly during the observation period. Despite this, the rate of positive sentiment did not increase at the same rate. While there were no significant changes in connection (integration with other modes of transport, parking facilities, access to the home, nearest station, working place, and school, etc.) and convenience (ease of reservation, ticketing, boarding-alighting, access to travel information, etc.) satisfactions, the importance of these services have decreased drastically recently. Here, readers should note that importance shows how much the topic is discussed on social media. Less discussion of the topic doesn't necessarily mean less importance, but shows decreasing priority for passengers. Thus, reliability has recently become more important than others.

Importance-performance analysis pointed to connection as the primary improvement area in both CR and HSR systems. Passengers attribute high importance to connection but perceive low satisfaction. Responsiveness, tangibles, and assurance satisfaction in both systems remained below the average and need improvement. Responsiveness and assurance offer great improvement opportunities for CR. In the HSR system, assurance should be prioritized, because passengers have placed more emphasis on this service dimension but the perceived quality has declined in recent years. Therefore, priority improvement policies for CR and HSR should address these aspects. These findings and recommendations are expected to guide TCDD in policymaking and improving the quality of rail services.

**Declarations**

**Ethical Approval**

Not applicable.

**Competing interests**

The authors declare no conflict of interest.

**Authors' contributions**

Samet Güner: Conceptualization, Writing - Original Draft, Writing - Review & Editing
Kamil Taşkın: Writing - Original Draft, Formal analysis
Halil Ibrahim Cebeci: Conceptualization, Methodology, Data Curation, Formal analysis, Writing - Original Draft, Writing - Review & Editing
Emrah Aydemir: Methodology, Data Curation, Formal analysis

Funding

No funding was received for this work.

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

References


44. Mogaji, E. and Erkan, I. (2019), Insight into consumer experience on UK train transportation services, Travel Behaviour and Society, 14, 21–33.


61. UAB (2021), *Ulaşan ve Erişen Türkiye 2021*, Ulaştırma ve Altyapı Bakanlığı Yayınları, Ankara

**Figures**

![Railways in Turkey](image_url)

*Figure 1*

Railways in Turkey
Figure 2
Methodology of the research
Figure 3
Number of tweets

Figure 4
Comparison of CR and HSR services
Figure 5

Importance-satisfaction plot for conventional rail
Figure 6

Importance-satisfaction plot for high-speed rail

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixA.docx