GSI JOURNALS

e-ISSN 2645-9116

https://dergipark.org.tr/tr/pub/atrss
DOI: 10.53353/atrss.1327615

GSI Journals Serie A: Advancements in Tourism, Recreation and Sports Sciences
(GSI Dergileri Seri A: Turizm, Rekreasyon ve Spor Bilimlerindeki Gelişmeler)

ATRSS 2024, 7 (1): 111-122 – Research Article/Araştırma Makalesi- Received/Geliş T.: 14.07.2023 - Published/Kabul T.: 21.10.2023

Comparison of Machine Learning Algorithms for Classification of Hotel Reviews: Sentiment Analysis of TripAdvisor Reviews*

Hüseyin Ertan İNAN, Ondokuz Mayıs University, Faculty of Tourism, Tourism Management, hertaninan@gmail.com, Samsun, Türkiye, ORCID: 0000-0002-6642-4813

Abstract

Sentiment analysis can help extract meaningful information from these data piles from various websites and social media and measure consumers' reactions by classifying consumers' emotions as positive, negative or neutral. The success of sentiment analysis varies according to feature selection, vector space selection and machine learning method. For this reason, determining the most successful method in sentiment analysis is still controversial and important. A limited number of studies have been conducted comparing the success of various machine learning methods in sentiment analysis of hotel reviews in English. Considering this gap, the purpose of this research is to determine the most successful machine learning algorithm for sentiment analysis of hotel reviews. For this purpose, 708 reviews for 5-star hotels in Istanbul were collected manually. Obtained data were classified as positive and negative using logistic regression, k-nearest neighbor, naive Bayes and support vector machine methods. Analysis results show that the logistic regression method was the most successful classification algorithm, with an accuracy rate of 0.92. It is followed by support vector machine (0.90), naive Bayes method (0.77) and k-nearest neighbor algorithms (0.66).

Keywords*:* *Tourism Management, Data Mining, Machine Learning, Text Mining, Sentiment Analysis*

* This study is not included in the study group that requires TR Index Ethics Committee Approval.

1. Introduction

Today, consumers share their experiences, recommendations, opinions and complaints on websites and the amount of data collected online is growing rapidly. Customers use chat rooms, newsgroups and electronic consumer forums to express their opinions about brands (Gelb & Sundaram, 2002). Increasing competition and intensifying market dynamism are driving businesses to seek new opportunities. In an unprecedentedly large and complex mass of data, businesses are in search of a way out and a solution (Altunışık, 2015). Thanks to the developing information technologies and software, the analysis of the collected data can be realized more easily and quickly. In particular, new techniques have been developed for the analysis of large data stacks that have emerged as a result of the use of the internet and social media. These techniques are generalized as data mining and can be grouped under data mining as machine learning algorithms (Balaban & Kartal, 2015). Data mining is the search for connections that will enable us to make predictions about the future from big data using computer programs (Pektaş, 2013).

Machine learning is the automation of learning tasks and learning, where observations of environmental conditions and transition-based rules are equivalent. Machine learning examines previous examples and results and learns how to redo these tasks and generalizes about new situations. Usually, a machine learning system does not use a single observer but a whole system called a training set. This set contains some machine-readable forms with sample observation codes. The discovery of knowledge in the database is achieved with the help of data mining and machine learning algorithms (Akgöbek & Çakır, 2009). Text mining is a type of data mining that uses text as a data source. Text mining aims to obtain structured data from text (Şeker, 2015). Statistical and mathematical methods form the basis of text mining methods. The areas of use of text mining vary. Sentiment analysis, opinion mining, text classification, author recognition, title extraction and keyword extraction are examples of these areas (Kılınç et al., 2016). The problems referred to as opinion or opinion mining in the literature are one of the problems in social networks. These problems are referred to as sentimental analysis in the literature (Şeker, 2016). Sentimental analysis aims to reveal the emotional expressions in texts and its most frequently used application is emotional sacredness. In other words, it is questioned whether the text content is positive, negative or neutral. According to the results of the analysis, people's attitude towards the subject under investigation is determined (Şeker, 2015; Kaynar, 2016). On the other hand, sentiment analysis studies focus on extracting opinion, mood and more complex emotions from texts, in addition to sentiment polarity (Şeker, 2015). Electronic word-of-mouth communication is a communication practice that allows many topics and comments to be shared in a virtual environment through different methods (Sarıışık & Özbay, 2012). There are many websites where customers can comment on hotel businesses and share their past experiences, suggestions, complaints and recommendations. The comments on these websites are also important sources for text mining applications for businesses. Hotel managers and tourism planners can develop new management and marketing strategies by exploring consumers' attitudes and perceptions through sentiment analysis.

The effectiveness of sentiment analysis varies depending on the choice of features, vector space, and machine learning technique. This variation makes figuring out the most effective technique for sentiment analysis still debatable and significant. There has not been much research evaluating the effectiveness of different machine learning techniques for sentiment analysis of hotel reviews. In addition, even if the same methods were used and compared in some research (Sayed et al., 2020; Tuna et al., 2021), vector space models or languages of reviews differed. It might affect the results and performance of the algorithms. There is still a gap in determining the most effective machine-learning methods for the classification of hotel reviews in sentiment analysis.

This study aims to identify the most effective machine learning method for sentiment analysis of hotel reviews in light of this gap. The study was completed in five stages. First, data on hotel reviews were collected from the Trip Advisor website. For this purpose, 708 reviews in English for 5-star hotels in Istanbul, one of the world's most important tourism destinations, were collected manually. The second stage is the pre-processing stage of the data. After this stage, attribute selection and vector space selection stages were carried out. The Term Frequency (TF) vector, which has not been used very often in previous research, was used in the current study before training and classifying hotel reviews. Finally, four different machine learning algorithms (logistic regression (LOGR), Naïve Bayes (NB), support vector machines (SVM) and K-Nearest Neighbor (KNN) were used, and their classification performances were compared.

2.Literature Review

Due to the widespread use of the Internet and technology and the ease of access to information, the analysis of online customer reviews has become important for businesses and other organizations. In recent years, sentiment analysis studies have become more important in different fields such as finance, health, travel and tourism. In general, sentiment analysis aims to determine a speaker's or an author's attitude towards a particular topic or the overall contextual sentiment polarity of a document. There are two types of methods in sentiment analysis, namely the semantic orientation approach and the statistical machine learning approach (Zeng, Wang & Gao, 2018). The semantic approach analyzes emotions based on the extraction of emotional words and expressions, while the statistical approach is based on the extracted emotional features and machine learning approach (Yang, Xu & Shi, 2012; Li, Xu, Wang & Mo, 2003).

Sentiment analysis is an interdisciplinary subject. In the studies in the literature, different data mining techniques have been applied to analyze the content created by users on various platforms. Kaynar et al. (2016) conducted a study using movie reviews as a data set. In the study, sentiment analysis of movie reviews was performed using various machine learning techniques. They compared artificial neural networks, center-based classifier, Naive Bayes and support vector machines methods. As a result of the research, they stated that artificial neural networks gave more successful results than others. Kılınç et al. (2016) compared the success of k-nearest neighbor, naive Bayes and j48 algorithms in classifying academic publications and concluded that k-nearest neighbor algorithm gave more successful results. Çoban et al. (2015) developed a feature extraction model to determine whether Turkish user messages obtained from Twitter are positive or negative and examined the effect of this model on classification success. The experimental results indicated that support vector machines, Naive Bayes, Multinomial Naive Bayes and k-nearest neighbor algorithms, Multinomial Naive Bayes gave better results. Akın and Gürsoy Şimşek (2018) compiled the messages shared on Twitter about the programs broadcast on a TV channel and labeled each message as positive, negative or neutral with a sentiment analysis technique. They examined the relationship between tweets and rating value for all programs in the broadcast stream in a certain time period. They concluded that an increase in positive and neutral sentiment tweets leads to an increase in ratings, while an increase in negative sentiment tweets decreases ratings. Bastem and Şeker (2017) developed an artificial intelligence system that performs personality analysis of people by using the posts made by people on social media. By using machine learning, data mining and data science techniques, they classified the MBTI personality types into the closest type to the person and predicted the personality by looking at the tweets of people on twitter. Random Tree, Naive Bayes and Gradiend Boosted Tree algorithms were used in the analysis and up to 54% successful results were obtained in personality prediction.

Sentiment analysis for tourism businesses and hotel reviews is another application area that has attracted the attention of researchers in recent years. When the studies on the classification of sentiments in hotel reviews are examined, Schmunk et al. (2014) stated that support vector machines and Oğul and Ercan (2016) stated that the random forest method gave more successful results. Dey et al. (2016) evaluated the performance of machine learning algorithms in the classification of hotel and movie reviews and stated that k-nearest neighbor method gave more successful results in hotel reviews and naive Bayes method in movie reviews. Polat and Ağca (2022) conducted a study for classification and sentiment analysis of Turkish and English Tripadvisor reviews. Accordingly, machine learning algorithms were found to be more effective in classification than dictionary-based analysis. However, machine learning algorithms produced more successful results in Turkish reviews. Acar and Uğur (2021) conducted a sentiment analysis study using 665 Tripadvisor reviews of hotels in Ankara. IBM Watson Tone Analyzer artificial intelligence program was used in data analysis. According to the results of the study, the emotions that guests emphasized with strong intonation regarding the services of hotel businesses were "trust" and "enjoyment". Tuna, Kaynar and Akdoğan (2021) conducted sentiment analysis of customer reviews about 164 hotels in Antalya using seven different machine learning algorithms, mainly Logistic Regression, Random Forest, CART Decision Tree. Among these methods, Logistic Regression (87.9%) was found to produce the most successful results. Support vector machines (86.84%) were determined as the second most successful application after logistic regression. Al-Smadi et al. (2018) conducted a sentiment analysis study for hotel reviews using recurrent neural network (RNN) and support vector machines (SVM) methods. According to the results of the research, it was stated that the support vector machines approach outperformed the iterative neural network (RNN) approach. Shi and Li (2011) developed a supervised machine learning approach using a unigram feature with two types of input (frequency and TF-IDF) to achieve sentiment classification of online hotel reviews. Their experimental findings demonstrate that TF-IDF information is more useful than frequency. Bagherzadeh et al. (2021) used a public dictionary-based method and a complex machine-learning algorithm and compared the accuracy metrics. According to the results, the dictionary-based method performs a machine-learning algorithm approach. In another research, Sayed et al. (2020) used nine classifiers for analyzing 6318 reviews written in different forms of the Arabic language. They found out the Ridge Classifier (RC) has the best overall performance according to accuracy, precision, recall, F1-score, and training time. SVM and LOGR followed the Ridge Classifier.

The efficiency of sentiment analysis varies based on the vector space, machine learning method, and feature selection. The efficiency of various machine learning approaches for sentiment analysis of hotel reviews has not been thoroughly investigated. The best machine-learning techniques for categorizing hotel reviews in sentiment analysis are still not fully understood. More studies are needed on this subject. The method of research and application steps were stated in the next part.

3. Methodology

Classification applications using machine learning algorithms can be performed using open-source software such as WEKA and KNIME or programming languages such as R and Python. In this study, Python programming language and Numpy, Pandas, nltk and re modules were used for the analysis. Techniques under the name of machine learning are techniques based on learning from data and enable the classification or clustering of outputs with the help of algorithms (Balaban & Kartal, 2015). K-nearest neighbor (KNN), Logistic Regression (LOGR), Naive Bayes (NB) and Support Vector Machines (SVM) algorithms from machine learning algorithms were applied for sentiment analysis of hotel reviews and their success was compared. This study is not included in the group of studies requiring ethics committee permission.

3. 1. Logistic Regression

Logistic regression calculates the effect of multiple independent variables to predict the class of one or the other of two categories of dependent variables (Balaban & Kartal, 2015). In the logistic regression approach, the probabilities of the attributes extracted from the dataset are calculated according to equation (1). As seen in the formula, a class is determined for the attributes by calculating probability values using an exponential function (Parlar, 2022).

$$P(c_i \mid f_1, \dots f_n) = \frac{\exp\left(\sum_{i=1}^{N} w_i f_i\right)}{\sum_c \exp\left(\sum_{i=1}^{N} w_i f_i\right)} \quad (1)$$

3. 2. K-Nearest Neighbor Algorithm

In cases where the independent variables are numerical, the K-Nearest Neighbor Algorithm can be used. This algorithm performs classification based on the distances between observations (Altunkaynak, 2017:

110). In the K-NN algorithm, the samples in the training set are specified with n-dimensional numerical attributes. All training samples are kept in an n-dimensional sample space so that each sample represents a point in the n-dimensional space. When an unknown example is encountered, the class label of the new example is assigned by determining the k closest examples from the training set according to the majority vote of the class labels of its k nearest neighbors (Han & Kamber, 2006; Taşçı & Onan, 2016).

3. 3. Naive Bayes Algorithm

Naive Bayes Algorithm is an algorithm often used in sentiment analysis problems. The probabilistic approach of the Naive Bayes algorithm makes strong assumptions about how the data is generated and proposes a probabilistic model that incorporates these assumptions. It then performs parameter estimation using a collection of labeled training samples (Balaban & Kartal, 2015; Vinodhini, 2012; Oğul & Ercan, 2016). In Naive Bayes (NB) approach, the computational cost of computing $P(c_i \,|\, f_1, \dots f_n)$ in datasets consisting of many attributes $(f_1, \dots f_n)$ is quite high and to eliminate this problem, class information is considered conditionally independent with the naive assumption. $P(c_i \,|\, f_1, \dots f_n)$ is calculated according to equation (2) (Han & Kamber, 2006; Parlar, 2022).

$$P(c_i \,|\, f_1, \dots f_n) = \frac{P(c_i)\mathrm{P}(f_1, \dots f_n|c_i)}{\mathrm{P}(f_1, \dots f_n)} \quad (2)$$

3. 4. Support Vector Machines

Support Vector Machines (SVMs) are one of the most commonly used linear classifiers available and include many classifiers for finding a good linear discriminator to distinguish between different classes. The basic principle of SVM is to find the linear discriminators in the search space that best discriminate different classes (Oğul & Ercan, 2016). In linearly separated problems, the goal is to find the hyperplane passing through the features. The hyperplane consists of two lines where the features belonging to the classes are the farthest apart. In a linear equation ax + b, the aim is to calculate the values of a and b for the equation that will separate the classes (Demir, Erdoğmuş & Kekeçoğlu, 2018).

3. 5. Machine Learning Performance Metrics

The confusion matrix is used to compare the performance of machine learning algorithms. In the error matrix, the prediction values for the target attribute are compared with the actual values. In this study, positive and negative comments were classified. As shown in Table 1, when the true value is negative, if the predicted value is negative, it is true negative (true negative-TN), and if the predicted value is positive, it is false positive (false-positive-FP); when the true value is positive, if the predicted value is positive, it is true positive (true positive-TP), and if the predicted value is negative, it is false negative (false negative-FN). Accuracy, precision and sensitivity values were calculated with the help of the error matrix and the performances of the machine learning algorithms were compared. Description and formulation of the accuracy, precision, sensitivity, and F-measure were given below (Balaban & Kartal, 2015; Parlar, 2022).

Table 1. The confusion matrix

	Predicted Class	
	Negative	Positive
Negative	TN	FN
Positive	FP	TP

Accuracy measures how much of all instances a classification model correctly classifies. Formula (3) is used to calculate the accuracy rate.

$$\text{Accuracy} = \frac{(\mathrm{TP+TN})}{(\mathrm{TP+TN+FP+FN})} \quad (3)$$

Precision measures how many of the samples that a classification model predicts as positive are actually positive. Formula (4) is used to calculate the precision rate.

$$\text{Precision} = \frac{\mathrm{TP}}{(\mathrm{TP+FP})} \quad (4)$$

Sensitivity (True Positive Rate) indicates how accurately a classification model can detect true positive examples. Formula (5) is used to calculate the sensitivity rate.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP+FN})} \qquad (5)$$

F Measure is calculated as the harmonic mean of the precision and sensitivity measures. Formula (6) is used for the calculation of the F Measure.

$$\text{F Measure} = 2 \text{ x } \frac{(\text{Precision x Sensitivity})}{(\text{Precision + Sensitivity})} \qquad (6)$$

3. 6. Application

The computer used in the study has 1.8 GHz Intel Core i5 processor, 8 GB RAM memory, Intel HD Graphics 6000 graphics card. Within the scope of the study, data mining application was realized in five steps. These steps are as follows:

3.6.1 Step 1: Collection of Dataset

Within the scope of the application, hotel reviews and evaluation scores of five-star hotels operating in Istanbul were collected manually from Tripadvisor.com. The sentiment polarity of the reviews was determined by considering the evaluation scores. Reviews with 1 and 2 points were included in the analysis as negative reviews, while reviews with 4 and 5 points were included in the analysis as positive reviews. In total, 708 comments, 355 positive and 353 negative, were included in the study. Examples of negative comments are given in Table 2. The word cloud of negative comments, which visually represents the frequency of words in the text data, is shown in Figure 1.

Table 2. Example of Negative Comments

Example of Negative Comments
"My son dropped I bit chocolate on floor. It can be easily cleaned. they said i have to change the carpet. 100 euro for that I hope never coming back here anymore. Be careful they bring food to your room even if you don't asked for it. That's why it dropped. Very expensive"
"Yes, the location is excellent - but we were mightily disappointed in the hotel itself. In my room, there was a dead rat. My traveling companions complained of a dirty room. My room was changed, with great apology from the staff...my friends complaints were not so well received. The view of the city at night from the rooftop terrace is spectacular. However, I would not recommend this hotel. "
"Let's make it simple. Service is not smiling at all, room are very small, noisy, bathroom with hair in the sink, pictures different than the advertising. The big plus is the terrace which is lovely but you can't drink alcohol even a beer, it is the hotel policy, many places around don't have this restriction, I won't enumerate all the places! This hotel is not terrible maybe not poor either but there are much better around"

Figure 1. Word Cloud of Negative Comments

Examples of positive comments are given in Table 3. The word cloud of positive comments, which visually represents the frequency of words in the text data, is shown in Figure 2.

Table 2. Example of Positive Comments

Example of Positive Comments
"What a most beautiful Hotel. The front desk staff were so very helpful with very good understanding of english. The dinning staff were so pleasant & interested in finding out where we were from & how we liked their country. The doormen went above helpful when we needed transport, arranging taxis and fares for us before the taxi even arrived. The room was so well laid out, most very comfortable bed, superb furniture with a wonderful shower. More than we could have hoped for."
"Has wonderful pool, good sauna fine food and spacious rooms. I liked the service, smile and hospitality. Located on the Europe side of the city and close to the new airport and also operates a shuttle"
"Here have good respect and gentle.. But here really good enjoy relaxed and you will see that what can I say different? I like it I love it I hope you like it you love it here better for alone business couples"

Figure 1. Word Cloud of Positive Comments

3.6.2. Step 2: Data preprocessing phase

In the data cleaning phase, punctuation was removed, all comments were converted to lower case and stopwords were removed. In the next stage, the roots of the words were separated. The inflectional suffixes added to the words were removed and a new list of word roots was created.

3.6.3. Step 3: Feature selection

In the feature selection stage, subsets of features are selected without compromising accuracy. By deleting irrelevant data, it is aimed to achieve high dimensionality reduction and accuracy. (Tunç & Ülger, 2016). Creating a vector with all the words in the comments may cause problems in terms of RAM, memory and processing speed. In addition, analyzing different numbers of attributes may affect algorithm success. Parlar (2022) found that algorithms were less successful in applications with a number of attributes below 2000. In this study, in the comparison of algorithms, the analysis was performed with the 2000 most used words among 708 comments.

3.6.4. Step 4: Vector space model selection

In the vector space model, each object is defined in a vector structure. The different properties that objects have constitute the axes of the vector space. In addition, each object has a certain position in the vector space according to its properties (İlhan et al., 2008). Three different methods are used to represent a text in a vector space model. These are binary vector, TF (Term Frequency) vector and TF - IDF vector (Term Frequency-Inverse Document Frequency) (Göker & Tekedere, 2017). Term Frequency (TF) vector is a form of definition that includes the number of occurrences of word roots in the data (Çalış et al., 2013). In this study, Term Frequency (TF) model was utilized.

3.6.5. Step 5: Classification

Before applying classification algorithms, a certain percentage of the data is divided into test and training data in the data separation phase. The general approach in percentage division is to separate 1/3 as test data, but different ratios can be taken (Altunkaynak, 2017: 114). In this study, the data were divided into four parts as two dependent and two independent variables as test and training. The dependent and independent variables were divided by 1/3 and 33% of them were divided as test data and 77% as training data. The separated training sets were allowed to learn from each other and Logistic regression (LOGR), K-nearest neighbor (KNN), Support Vector Machines (SVM) and Naive Bayes (NB) algorithms were applied to predict the sentiment polarity of the data in the test set. The best prediction method was determined.

4. Findings

In order to classify hotel reviews as positive and negative, 708 reviews of Istanbul hotels were divided into 234 test data and 474 training data. These data were classified by logistic regression (LOGR), K-nearest neighbor (KNN), support vector machines (SVM) and naive Bayes (NB) methods. The research test data included 115 positive comments and 119 negative comments. Accordingly, the logistic regression method correctly predicted 103 out of 115 positive comments by classifying them as positive. In addition, it classified 112 out of 119 negative comments as negative. The support vector machines method correctly classified 100 of the positive comments as positive and 111 of the negative comments as negative. The K nearest neighbor algorithm correctly assigned 68 of the positive comments and 86 of the negative comments to the correct class. In the Naive Bayes method, 96 positive comments and 85 negative comments were assigned to the correct classes. The error matrices of the research findings are shown in Table 3. In order to compare the performance of the algorithms, accuracy, precision and sensitivity values were calculated with the help of the data in the confusion matrix.

Table 3. Accuracy matrices of algorithms

LOGR	Positive	Negative	Total	**DVM**	Positive	Negative	Total
Positive	103	12	115	Positive	100	15	115
Negative	7	112	119	Negative	8	111	119
Total	110	124	234	Total	108	126	234
KNN	Positive	Negative	Total	**NB**	Positive	Negative	Total
Positive	68	47	115	Positive	96	19	115
Negative	33	86	119	Negative	34	85	119
Total	101	133	234	Total	130	104	234

The accuracy, precision, sensitivity and F-measure values showing the success rates of the algorithms are shown in Table 4. As seen in the tables, the algorithm with the highest accuracy rate is the logistic regression algorithm. The data set tested with this method was correctly classified with a rate of 92%. Support vector machines method has a success rate close to logistic regression. The support vector machines method performed 90% successful classification.

Table 4. Success Rates of Machine Learning Algorithms

LOGR	Accuracy	Precision	Sensitivity	F-Measure
	0,92	0,90	0,94	0,92
KNN	Accuracy	Precision	Sensitivity	F-Measure
	0,66	0,59	0,67	0,63
DVM	Accuracy	Precision	Sensitivity	F-Measure
	0,90	0,87	0,93	0,90
NB	Accuracy	Precision	Sensitivity	F-Measure
	0,77	0,83	0,74	0,78

Naive Bayes method was less successful compared to these two methods. Naive Bayes algorithm has an accuracy rate of 77%. K-nearest neighbor algorithm is the least successful algorithm among the four algorithms used in the study. The K-nearest neighbor method has a success rate of 66%. Figure 3 shows a graph of the accuracy rate showing the performance of logistic regression (LOGR), K-nearest neighbor (KNN), Support vector machines (SVM) and Naive Bayes (NB) algorithms.

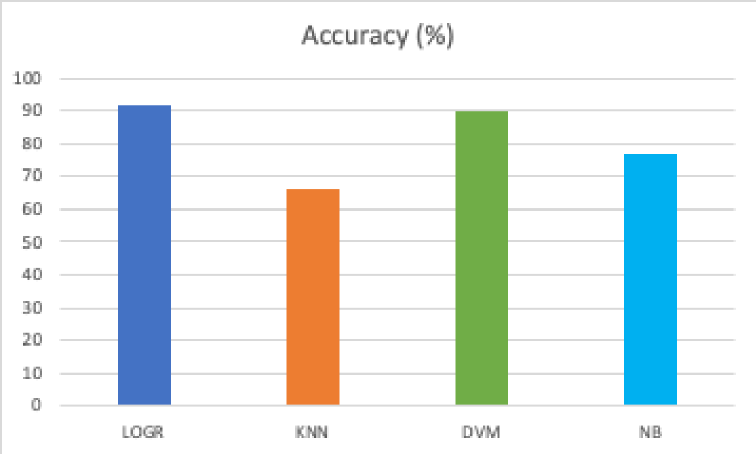

Figure 3. Algorithm Performances

When the calculated precision and sensitivity values are analyzed, the precision rate of the logarithmic regression method is 0.90 and the sensitivity rate is 0.94. The precision rate of the K nearest neighbor algorithm is 0.59 and the sensitivity rate is 0.67. After the application of the support vector machines method, the precision ratio is 0.87 and the sensitivity ratio is 0.93. When the findings of the Naive Bayes method are examined, the precision value is calculated as 0.83 and the sensitivity value is 0.74. Precision and sensitivity criteria may not be sufficient for meaningful results. For this reason, f-measure is used as an evaluation criterion that can provide more accurate results where both criteria are used together. F-measure is calculated by taking the harmonic mean of precision and sensitivity values (Haltaş et al., 2015). Considering the F-measure value showing the success of the algorithms, it was determined that the logistic regression method was the most successful classification algorithm with a rate of 0.92. It is followed by support vector machine with 0.90, naive Bayes method with 0.78 and k-nearest neighbor algorithms with 0.63. In light of the comprehensive analysis conducted in Section 4, it is evident that the insights gained from this research are pivotal for shaping the conclusions and recommendations that follow in Section 5.

5. Conclusions

The amount of data on customer reviews and customer complaints on the Internet has been increasing in recent years. These contents are important sources of information not only for customers but also for businesses. Sentiment analysis provides a number of advantages in terms of understanding people's perceptions of different services and products, discovering emotions in web, creating advertising content, developing recommendation systems and analyzing market trends (Dey et al., 2016). When examined in the literature on sentiment analysis, classification success may vary according to data type, feature selection and classification algorithm. In some studies (Schmunk et al., 2014; Oğul & Ercan, 2016; Dey et al., 2016), support vector machines, k-nearest neighbor and naive Bayes methods were used for sentiment analysis of hotel reviews. According to the research findings, it was discovered that the logistic regression method is more successful than these three methods frequently used in the literature. In parallel with the research results, Tuna et al. (2021) discovered that Logistic Regression (87.9%) was the algorithm that produced the most successful results in sentiment analysis of hotel reviews, and support vector machines (86.84%) were determined as the second most successful application after logistic regression. Another study, which is in line with the results of the research, found that logistic regression method is more successful than support vector machines and naive Bayes methods in the classification of text data (Parlar, 2022). According to the results of the study, support vector machines outperformed naive Bayes and k-nearest neighbor algorithms. This result is consistent with the study of Schmunk et al. (2014), who reached similar findings using different vector space models. The fact that the Naive Bayes method outperforms the K-nearest neighbor algorithm contradicts the results of Dey et al. (2016).

The recommendation is the basis of word-of-mouth marketing. Sharing comments and recommendations about products and brands with other consumers accelerates the purchasing decision process (Kalpaklıoğlu, 2015). Hotel managers should be aware of which content tourists attach the most importance to in their comments in order to identify the aspects of their businesses that need improvement (Bayer & Aksöz, 2015). Hotel businesses can carry out sentiment analysis studies by using the methods found to be successful as a result of this research. Thus, the strengths and weaknesses of the hotel business can be revealed objectively. By analyzing negative comments about hotels, new arrangements can be planned quickly and deficiencies can be eliminated. With the help of the proposed method, the data stored on the internet and in the hotels' own databases can be analyzed quickly.

According to the findings of the study, it is encouraged that businesses can use logistic regression and SVM for sentiment analysis of customer reviews. These algorithms can help companies gain valuable insights into customer opinions, improve products or services, and make data-driven decisions. Developing tools and platforms that incorporate these algorithms to automatically analyze and categorize customer feedback will enable quicker response times and better customer satisfaction.

Hotels can create new applications in light of the research results for more accurate sentimental analyses. Thus, hotels have many benefits from this kind of practice. For instance, sentiment analysis allows hotels to evaluate themselves against competitors, identifying their strengths and weaknesses in guest satisfaction. Insights from sentiment research can inform efforts to position the hotel competitively in the market. Moreover, hotels can employ sentiment analysis to understand what components of their services clients love the most, allowing for customized marketing initiatives. By finding areas that frequently garner positive feedback, hotels may deploy resources wisely to maintain and improve those elements.

Furthermore, identifying and fixing persistent issues can reduce operational expenses related to guest displeasure. Monitoring sentiment can assist in ensuring a consistent level of service quality across different hotel locations or businesses. In addition to these benefits, hotels can receive real-time feedback from visitor evaluations, enabling them to fix concerns swiftly and improve the guest experience during their stay. Analysis of visitor opinions can indicate areas where the hotel can strengthen its services, such as improving room cleanliness, staff attentiveness, or amenities. Additionally, sentiment analysis and rapid responses to unfavourable reviews can help hotels maintain a positive internet reputation, attracting more potential guests. The more hotels use those machine learning algorithms that ensure accurate sentiment analysis results, the more they gain those benefits above.

This study also has its own limitations. The study was limited to hotels in Istanbul and only English reviews were evaluated. In the study, the data were only categorized into two classes as positive and negative. Future studies may focus on sentiment analysis for Turkish hotel reviews. In addition, neutral comments can also be classified when analyzing sentiment polarity, and artificial neural networks, random forest and other classification algorithms that have shown successful results in sentiment analysis of texts in different fields can be used for hotel reviews and the results can be compared.

Acknowledgments and info

Support Info: A part of this study was presented as an oral presentation at the 20th National Tourism Congress held in Eskişehir in 2019 with the title of "Otel Yorumlarının Sınıflandırılmasında Makine Öğrenmesi Algoritmalarının Karşılaştırılması ve Eskişehir Otellerine Yönelik Yorumların Duygu Analizi" and published in the proceedings book.

Ethical Approval: In the article, the authors declare that they comply with national and international research and publication ethics. In case of detection of a contrary situation, **GSI Journals Serie A: Advancements in Tourism Recreation and Sports Sciences Journal** has no responsibility, and all responsibility belongs to the article's authors.

Ethics Committee Approval: Ethics Committee Approval was not required for the study.

Conflict of Interest: There is no conflict of interest or gain in the article.

Contribution Rate of Researchers: The study was prepared with the contribution of a single author. Contribution rates: 1. Author = 100%.

References

Acar, A. & Uğur, İ. (2021). Uluslararası Zincir Otellere Yönelik Tripadvisor Yorumlarının Duygu Analizi Yöntemi ile Değerlendirilmesi: Ankara Örneği, Türk Turizm Araştırmaları Dergisi, 5(3): 1803-1814.

Akgöbek, Ö., ve Çakır, F. (2009). Veri Madenciliğinde Bir Uzman Sistem Tasarımı. Akademik Bilişim, 9 , 801-806.

Akın, B., Gürsoy Şimşek, U. T., (2018) Sosyal Medya Analitiği ile Değer Yaratma: Duygu Analizi İle Geleceğe Yönelim. Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 5(3), 797-811.

Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., & Gupta, B. (2018). Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. Journal of computational science, 27, 386-393.

Altunışık, R. (2015). Büyük Veri: Fırsatlar Kaynağı mı Yoksa Yeni Sorunlar Yumağı mı?. Yıldız Social Science Review, 1(1), 45-76.

Altunkaynak, B. (2017). Veri Madenciliği Yöntemleri ve R Uygulamaları. Seçkin Yayıncılık, Ankara.

Balaban, M. E., & Kartal, E. (2015). Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları. Çağlayan Kitabevi, İstanbul.

Bagherzadeh, S., Shokouhyar, S., Jahani, H., & Sigala, M. (2021). A generalizable sentiment analysis method for creating a hotel dictionary: using big data on TripAdvisor hotel reviews. Journal of Hospitality and Tourism Technology, 12(2), 210-238.

Bastem, K., & Şeker, Ş. E. (2017). Veri Madenciliği Yöntemleri ile Twitter Üzerinden Mbti Kişilik Tipi analizi. YBS Ansiklopedisi, 4(2), 3-10.

Bayer, R. U. & Aksöz E. O. (2015). Otel Tercihinde Çevrimiçi Otel Değerlendirme Yorumlarına Ait İçeriklerin Önemi. Seyahat ve Otel İşletmeciliği Dergisi, 12(3).

Çalış, K., Gazdağı, O., & Yıldız, O. (2013). Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti. Bilişim Teknolojileri Dergisi, 6(1), 1-7.

Çoban, Ö., Özyer, B., & Özyer, G. T. (2015). Türkçe Twitter mesajlarının duygu analizi. In Signal Processing and Communications Applications Conference (SIU).

Demir, H., Erdoğmuş, P., & Kekeçoğlu, M. (2018). Destek Vektör Makineleri, YSA, K-Means ve KNN Kullanarak Arı Türlerinin Sınıflandırılması. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 6(1), 47-67.

Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). Sentiment Analysis of Review Datasets Using Naive Bayes and K-NN Classifier. International Journal of Information Engineering and Electronic Business 8(4):54-62.

Gelb, BD. (2002), Sundaram, S. Adapting to "Word Of Mouse". Business Horizons; 45(4): 21-25.

Göker, H., & Tekedere, H. (2017). FATİH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri ile Otomatik Değerlendirilmesi. Bilişim Teknolojileri Dergisi, 10(3), 291-299.

Haltaş, A., Alkan, A., & Karabulut, M. (2015). Metin Sınıflandırmada Sezgisel Arama Algoritmalarının Performans Analizi. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 30(3).

Han, J., & Kamber, M. (2006). Data Mining: Concepts and Techniques (Second ed.): The Morgan Kaufmann Series in Data Management Systems.

İlhan, S., Duru, N., Karagöz, Ş., & Sağır, M. (2008). Metin Madenciliği ile Soru Cevaplama Sistemi. Elektronik ve Bilgisayar Mühendisliği Sempozyumu (ELECO), Bursa, 26-30.

Kalpaklıoğlu, N. Ü. Bir Pazarlama İletişimi Unsuru Olarak E-Wom'un Turizm Ürünleri Tercihine Etkisi. Maltepe Üniversitesi İletişim Fakültesi Dergisi,2015; 2(1): 66-90.

Kaynar, O., Görmez, Y., Yıldız, M., & Albayrak, A. (2016). Makine öğrenmesi yöntemleri ile Duygu Analizi. In International Artificial Intelligence and Data Processing Symposium (IDAP'16), September (pp. 17-18).

Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., & Özçift, A. (2016). KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi. Marmara Fen Bilimleri Dergisi, 28(3), 89-94.

Li, H. X., Xu, L. D., Wang, J. Y., & Mo, Z. W. (2003). Feature space theory in data mining: transformations between extensions and intensions in knowledge representation. Expert Systems, 20(2), 60-71.

Oğul, B. B., & Ercan, G. (2016). Türkçe Otel Yorumlarından Duygu Analizi. In 24th Signal Processing and Communication Application Conference (SIU), 16-19 May 2016.

Parlar, T. (2022). Nitelik Seçimi Kullanarak Twitter Kullanıcısının Cinsiyet Sınıflandırması üzerine Bir Çalışma. International Journal of Engineering Research and Development, 14(3), 204-210.

Pektaş, A. O. (2013). SPSS ile Veri Madenciliği. Dikeyeksen Yayın Dağıtım, İstanbul.

Polat, H. & Ağca, Y. (2022). Tripadvisor Kullanıcılarının Türkçe ve İngilizce Yorumları Kapsamında Duygu Analizi Yöntemlerinin Karşılaştırmalı Analizi. Abant Sosyal Bilimler Dergisi,22(2): 901-916.

Sarıışık, M., & Özbay, G. (2012). Elektronik Ağızdan Ağıza İletişim ve Turizm Endüstrisindeki Uygulamalara İlişkin Bir Yazın İncelemesi. Uluslararası Yönetim İktisat ve İşletme Dergisi, 2012; 8(16): 1-22.

Sayed, A. A., Elgeldawi, E., Zaki, A. M., & Galal, A. R. (2020). Sentiment analysis for arabic reviews using machine learning classification algorithms. In 2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE) (pp. 56-63). IEEE.

Schmunk, S., Höpken, W., Fuchs, M., & Lexhagen, M. (2014). Sentiment Analysis: Extracting Decision-Relevant Knowledge From UGC. In Information and Communication Technologies İn Tourism (pp. 253- 265). Springer, Cham.

Shi, H. X., & Li, X. J. (2011). A sentiment analysis model for hotel reviews based on supervised learning. In 2011 International Conference on Machine Learning and Cybernetics (Vol. 3, pp. 950-954). IEEE.

Şeker, S. E. (2015). Metin Madenciliği (Text Mining). YBS Ansiklopedi, 2(3), 30-32.

Şeker, S. E. (2016). Duygu analizi (Sentimental analysis). YBS Ansiklopedi, 3, 21-36.

Taşcı, E., & Onan, A. (2016). K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. Akademik Bilişim, 1(1), 4-18.

Tuna, M. F., Kaynar, O., & Akdoğan, M. Ş. (2021). Otellere ilişkin çevrimiçi geribildirimlerin makine öğrenmesi yöntemleriyle duygu analizi. İşletme Araştırmaları Dergisi, 13(3), 2232-2241.

Tunç, A. & Ülger, İ. (2016). Veri Madenciliği Uygulamalarında Özellik Seçimi İçin Finansal Değerlere Binning ve Five Number Summary Metotları ile Normalizasyon İşleminin Uygulanması, XVIII. Akademik Bilişim Konferansı, Adnan Menderes Üniversitesi

Vinodhini, G. et al. (2012). Sentiment Analysis and Opinion Mining: A Survey, International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Vol 2, Issue 6, June.

Yang, L., Xu, L., & Shi, Z. (2012). An enhanced dynamic hash TRIE algorithm for lexicon search. Enterprise Information Systems, 6(4), 419-432.

Zheng, L., Wang, H., & Gao, S. (2018). Sentimental feature selection for sentiment analysis of Chinese online reviews. International journal of machine learning and cybernetics, 9, 75-84.