Proceedings of the 6th International Conference on Advances in Artificial Intelligence (ICAAI 2022), 21-23 October, Birmingham, UK

Measuring Airport Service Quality Using Machine Learning Algorithms

MOHAMMED SALIH HOMAID

Department of Computational Engineering Sciences, Cranfield University, United Kingdom

IRENE MOULITSAS

Department of Computational Engineering Sciences, Cranfield University, United Kingdom

The airport industry is a highly competitive market that has expanded quickly during the last two decades. Airport management usually measures the level of passenger satisfaction by applying the traditional methods, such as user surveys and expert opinions, which require time and effort to analyse. Recently, there has been considerable attention on employing machine learning techniques and sentiment analysis for measuring the level of passenger satisfaction. Sentiment analysis can be implemented using a range of different methods. However, it is still uncertain which techniques are better suited for recognising the sentiment for a particular subject domain or dataset. In this paper, we analyse the sentiment of air travellers using five different algorithms, namely Logistic Regression, XGBoost, Support Vector Machine, Random Forest and Naïve Bayes. We obtain our data set through the SKYTRAX website which is a collection of reviews of around 600 airports. We apply some pre-processing steps, such as converting the textual reviews into numerical form, by using the term frequency-inverse document frequency. We also remove stopwords from the text using the NLTK list of stopwords. We evaluate our results using the accuracy, precision, recall and F1_score performance metrics. Our analysis shows that XGBoost provides the most accurate results when compared with other algorithms.

CCS CONCEPTS • Computing methodologies • Machine Learning • Machine Learning Algorithms • Ensemble Methods • Boosting •

Additional Keywords and Phrases: Airport service quality, Sentiment analysis, Machine Learning, Natural Language Processing.

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

Air travel has become one of the most common and popular ways of transportation worldwide. According to the International Civil Aviation Organization, the number of air travellers in 2019 was 4.5 billion [1]. Travellers have at times faced challenges and problems, but digital solutions have played a significant part in resolving these issues. According to [2], airport investment increased by 40% in 2020 in an attempt to improve airport capacity and operations, as well as to provide a better passenger experience. An airport is a complex place where travellers can access variety of services. Therefore, developing an effective tool to assess the quality of services is a critical issue for all involved stakeholders. Furthermore, airports are competitive environments and airport management should focus on enhancing operations and businesses [2]. Providing high quality services is an indicator of passenger satisfaction. Assessing airport services requires continuous observation and monitoring, to guarantee that high-quality services are provided. Nowadays, the number of web-based opinion platforms has increased. Platforms such as SKYTRAX [3], Google reviews [4] and Tripadvisor [5] allow travellers to rate their experience and express their opinions on a specific airport. These platforms receive a large number of reviews. Thus, an automated technique based on machine learning is an excellent candidate to efficiently and effectively evaluate Airport Service Quality.

In machine learning, historical data is used to make predictions. Thus, understanding and pre-processing the data is a critical part in any machine learning model. In this paper, we removed stopwords from the textual data. Then, the textual data was converted into a numerical format using the term frequency – inverse document frequency (TF-IDF) technique [6]. Then, we employed five machine learning algorithms on the processed dataset. Many sentiment analysis approaches can be used, but the key to success is determining which technique to employ on specific data. We compared between the most common algorithms used for determining sentiment analysis in such dataset and for developing an effective model for decision-making and business planning in organisations. We used the evaluation metrics of accuracy, recall, precision and F1-Score to evaluate the performance of the selected algorithms.

The rest of this paper is structured as follows. Section 2 presents the related work and data context is introduced in section 3. In section 4, the classification methods are explained, and the performance metrics are described in section 5. The experiments and evaluation are demonstrated in section 6 and finally the conclusion is summarised in section 7.

2 RELATED WORK

Over the last two decades, there has been a considerable attention on evaluating Airport Service Quality (ASQ) in the literature. Most of the studies generally applied user surveys and expert opinions to analyse and evaluate ASQ, for example [6, 7]. In [7] the authors used a fuzzy multi-attribute decision-making approach to assess the quality of passenger service at 14 Asia-Pacific international airports. The airports were ranked based on six different attributes. This ranking helps airport management in comprehending the services given to travellers. In [8] the authors used a questionnaire inspired by [9] to assess the quality of service at Melbourne Airport in Australia. They used the service quality attributes provided by the Airports Council International (ACI), the global representative of the airports around the world. The findings revealed that passengers' perceptions and expectations can differ significantly.

Over the last few years, machine learning techniques and natural language processing have been applied in the field of ASQ. In this area, user-generated content has become a common source for evaluating user satisfaction. Some studies have employed lexicon-based sentiment analysis where lexicon dictionaries have been utilised to measure the travellers' sentiment [10, 11], while other studies employed the more conventional machine learning algorithms [12, 13]. In [14], the authors proposed an analysis to compare between these two popular techniques. They evaluated the efficiency of VADER sentiment lexicon and logistic regression, to identify which method is appropriate depending on the context. A study conducted by [15] evaluated ASQ by employing sentiment analysis on Google maps' reviews using AFINN sentiment lexicon [16]. They also examined how ACI service attributes match the service attributes in Google reviews by using the Latent Dirichlet Allocation (LDA) [17]. In study [18], the authors measured the travellers' sentiment of six major US Airlines. Seven different classifiers were implemented in order to compare between the results. They found that Random Forest is the winning classifier for this particular application domain and given dataset. Table 1 illustrates the used data and methods in the domain of ASQ.

Table 1: Summary of the past and recent studies using online reviews in the domain of ASQ

Study Number	Method used	Data context	
		Source	Review sample size
[10]	Sentimentr sentiment lexicon, topic modelling	SKYTRAX	1224
[14]	Vader sentiment lexicon, logistic regression	SKYTRAX	38,105
[15]	AFINN sentiment lexicon, Latent Dirichlet Allocation	Google reviews	42,137
[19]	Sentiment analysis tool 'Theysay'	Twitter	4392
[20]	Sentiment analysis tool 'Knime' and 'Semantria'	SKYTRAX	895
[21]	Latent Dirichlet Allocation	SKYTRAX	7437

Regarding the data of SKYTRAX, there are few studies have utilised this data and applied different machine learning methods to analyse the sentiments of travellers and the provided services in airports. Study [20] looked into the level of customer satisfaction with airport services by applying sentiment analysis. The reviews of passengers on the SKYTRAX website were gathered and analysed. This study only looked at five international airports, and the data was gathered from the website between September 2013 and February 2014. Aviation and non-aviation services have been separated into two categories. Following that, the sentences are processed via Semantria, which is an automated sentiment analysis tool. After that, the sentences are graded as positives, negatives, or neutral. In study [21], the authors analysed 1,095 traveller reviews from SKYTRAX website to identify what are the key drivers for passenger satisfaction and dissatisfaction. The findings indicated that key satisfiers in the airport context such as pleasant environment and cleanliness. Whereas security-check, poor dining and confusing signage are recognised as key dissatisfiers. Study [22] proposed a Latent Dirichlet Allocation-based sentiment analysis approach. They only focused on the best 50 ranked airports. The result indicated that the model outperforms pervious model in forecasting the polarity of airport reviews.

3 DATA COLLECTION AND CLEANING

SKYTRAX is an international organisation which focuses on evaluating services and providing certifications to airlines and airports where passengers can write about their experience and rate the provided services for a specific airport. To extract the sentiment analysis of passengers based on the reviews, we scraped the reviews for all airports from SKYTRAX (https://www.airlinequality.com/). This dataset consists of textual description/reviews along with multiple features where reviewers can rate the provided services by selecting a numerical form from 1 to 5, where 1 accounts for poor and 5 accounts for excellent. The dataset comprised a total of 38,105 reviews together with 8 rated services by passengers. These reviews are labelled as recommended (yes) and not recommended (no). It contained 74.81% negative reviews and 25.19% positive, covering the time span from July 2004 to November 2020. Table 2 shows the actual attributes of the dataset used by the selected machine learning classifiers in this work.

Table 2: Dataset description

Attribute Name	Data Type	Description
Airport Name	String	Name of the airport to which the review belongs
Review Text	String	The textual review provided by the user
Queuing Times	Integer	Score between 1 and 5 regarding the time spent on queues
Terminal Cleanliness	Integer	Score between 1 and 5 regarding the cleanliness of the terminal
Terminal Seating	Integer	Score between 1 and 5 regarding the availability and quality of seats in the terminal
Terminal Signs	Integer	Score between 1 and 5 regarding the efficiency of the signs used in the terminal
Food Beverages	Integer	Score between 1 and 5 regarding the food beverages served in the airport
Airport Shopping	Integer	Score between 1 and 5 regarding the shops quality and availability in the airport
Wifi Connectivity	Integer	Score between 1 and 5 regarding the availability and quality of the provided Wi-Fi in the airport
Airport Staff	Integer	Score between 1 and 5 regarding the airport staff and their efficiency at work
Recommend	String	Contains “yes” or “no” values, whether or not the reviewer recommends the airport.

3.1 Data Pre-processing

Understanding the data is the most critical part of any machine learning model and is crucial in ensuring a successful data pre-processing step. Pre-processing steps were implemented which involve filling the missing values for the numerical attributes in the dataset. We applied the statistical mean to fill these missing values. The next step was the removal of stopwords by utilising the Natural Language ToolKit (NLTK) stopwords for English language [23]. In addition to previous steps, a further step was required for the use of the selected machine learning classifiers which expect the input variables to be in numerical form. In order to convert the review text into a numerical format, we applied the TF-IDF technique [6]. This is a statistical measure that assesses a word's relevance to a document in a collection of documents. It is accomplished by multiplying two metrics: the number of times a word appears in a document and the word's inverse document frequency over a set of documents. Each word is represented as a separate entry with a significant numerical value in the resulting data. Our dataset has 39,981 attributes, which will be subjected to the machine learning classifiers for training and testing purposes.

4 CLASSIFICATION METHODS

In this section, we describe the five different machine learning classifiers that we employed in our study. These classifiers are trained and employed on the processed dataset to detect the sentiment of air-travellers. In the field of sentiment analysis, there are two types of sentiment classifications, the lexicon-based approach and the learning-based approach. Lexicon-based approach requires effort and time to build an effective domain-specific lexicon dictionary whereas the learning-based algorithms can be trained to detect the associated class of a review [14]. Moreover, for this particular dataset, conventional machine learning classifiers can take advantage of all the numerical columns whereas lexicon-based approaches only expect textual inputs. Thus, conventional algorithms have the potential to produce more accurate results. In our dataset, we employed both numerical and textual columns as inputs to the machine learning models to compare between the results of the selected classifiers. Figurer 1 depicts our methodology steps to classify passengers' sentiments.

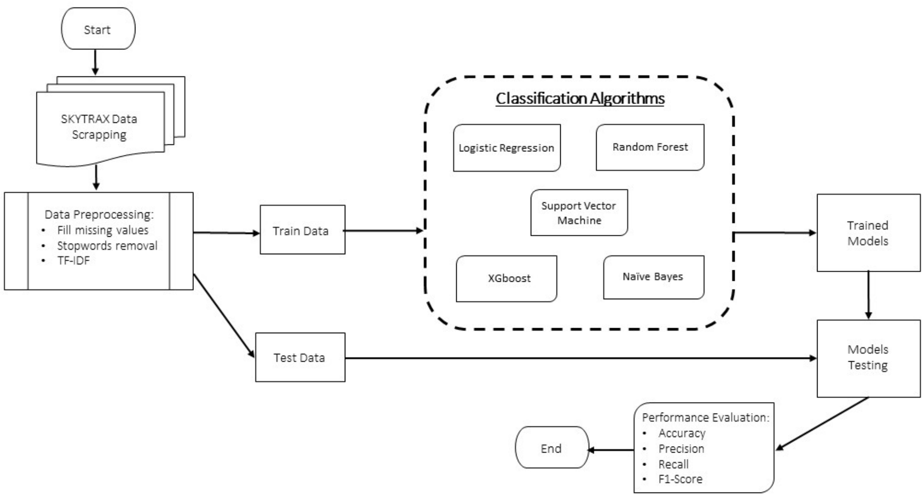

Figure 1: Research method steps

4.1 Logistic Regression

Logistic Regression is a machine learning classification approach which is used to predict the likelihood of a categorical dependent variable. The dependant variable in logistic regression is a binary variable that comprises data coded as 0 (NO), or 1 (YES). Logistic regression predicts the sentiment classification from the input of term frequency integer vector [15, 16].

4.2 XGBoost

The Extreme Gradient Boosting or XGBoost algorithm was introduced in [26]. XGBoost is a gradient boosted decision tree implementation designed to improve speed and performance. It is shown to also work well in imbalanced dataset. It is essentially a decision tree ensemble algorithm, which uses a gradient boosting framework [27].

4.3 Support Vector Machine

The Support Vector Machine (SVM) algorithm, introduced in [28], has been proven to work well for text classification due to its capability of handling large features. It is a classifier that aims to determine the hyperplane which is used to separate and categorise the data into different classes. In previous sentiment analysis studies [26, 27] SVMs have demonstrated promising results.

4.4 Random Forest

The Random Forest algorithm is an ensemble learning method for classification that is operated by constructing decision trees. Each tree has random samples from the dataset. The output of the ensemble datasets is aggregated to predict the final classification. Random forest has been used in a range of applications and domains including the medical domain [31].

4.5 Naïve Bayes

The Naïve Bayes classifier is a popular supervised classification algorithm. It is a probabilistic classifier based on the Bayes' theorem that takes strong independence assumptions into account [32]. Despite its strong assumptions and simplicity, the Naïve Bayes has been proven to work well in many domains. It is a popular method for text categorisation, where the problem is that of evaluating documents as belonging to one of two categories based on word frequencies. The core idea behind this algorithm is to use the joint probabilities of words and categories to find the probabilities of categories in a given text document [33]. Given a document x_i, the probability of each class y is computed as:

$$P(y|x_i) = \frac{P(y) \cdot P(x_i|y)}{P(x_i)} \quad (1)$$

where $P(y)$ is the probability of the classification, x is the given review and i represents the number of words [34], [35]. Algorithm 1 summarised the proposed approach for extracting passengers' sentiments.

ALGORITHM 1: Pseudocode of proposed experiments	
	Input: Data
	Output: Classification results
1	Begin
2	Dataset scrapping
3	Read dataset
4	Data preparation & cleaning
	Non-values filling
	Stopwords removal
	TF-IDF
5	Data splitting
6	Model building
7	Classification results
8	Evaluation of results
9	End

5 PERFORMANCE METRICS

To evaluate the performance of our models, we formulated this into a classification problem and employed the measures of accuracy, recall, precision, and F1-score to assess the performance of the chosen classifiers. These metrics are defined as follows:

The definition of accuracy is the closeness of the measurements to a specific value [36]

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives} \quad (2)$$

While Accuracy gives an indication of a classifier's overall performance, on its own is not sufficient and needs to be complemented with the notions of Recall and Precision. The concept of Recall is to determine the elements of actual positive that were correctly identified [36]

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

Recall also provides an assessment of the accuracy of our model's performance for positive classes. Precision is used to define the percentage of positive predictions were correct [37]

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (4)$$

The F1-score is used in addition to precision and recall, and it effectively indicates the balance between precision and recall. The F1-score is used to assess the accuracy of the test and seek a balance between precision and recall [36]

$$F1\ score = 2\ \frac{Precision * Recall}{Precision + Recall} \qquad (5)$$

6 EXPERIMENTS AND ANALYSIS

We perform the experiments on a personal computer with Intel(R) Core (TM) i7-9700 CPU, a processor speed of 3.00 GHz and 32 GHz RAM. Python Programming Language implemented in a Jupyter Notebook. Also, we employed libraries such as NLTK, ScikitLearn for the analysis, and for scraping the reviews, Beautiful Soup library was used.

Unlike lexicon-based tools, conventional machine learning algorithms require training. We divided our dataset into training and testing subsets. The training and testing ratio used for these experiments was 67:33. This resulted in a total of 25,530 entries for training and 12,575 for testing purposes. The test set will be used to test the performance of the trained model, while the training set will be used to train the chosen classifiers. We predicted sentiments from the testing data after training the classifier models. These predictions were then used to evaluate the model's performance using the performance metrics described earlier in Section 5.

In previous work, few studies used SKYTRAX reviews to assess ASQ. They mostly focused on analysing the data from business perspective such as identifying the key drivers of passenger satisfaction employing specific portion of the data. Whereas in this study, we focused on what is the best approach to classify passengers' sentiment. Thus, all available SKYTRAX reviews were used to assure the accuracy and validity of our models.

In this study, the binary classification of Recommended and Not Recommended is used to classify passengers' sentiments by applying different machine learning approaches. The Recommended is used when the reviews are positive while Not Recommended when they are negative. The dataset we have used contains some features such as Queuing Times, Terminal Cleanliness, Terminal Seating, Terminal Signs, Food Beverages, Airport Shopping, Wi-Fi Connectivity and Airport Staff. These features have ratings from 1 to 5, with 1 being the worst and 5 being the best. Therefore, these features behave as categorical in nature. For the textual reviews, we used TF-IDF features. Looking into the nature of data, we can conclude that most of the features are categorical in nature. Hence, all tree-based classifiers such as Random Forest and XGBoost are expected to outperform the non-tree-based classifiers, such as Support Vector Machines and Naïve Bayes.

We presented and evaluated five different machine learning algorithms to extract sentiment from the reviews data from the SKYTRAX website. To compare the performance of these algorithms, we used the recommended column in the dataset as a target column. In our study, the results in term of accuracy, precision, recall and F1-Measures were close to each other. We observe that the data is not balanced in terms of the number of positive classes being significantly lower than that of negative class. Due to this fact, algorithms such as Naïve Bayes might have weak performance. In such cases, tree-based ensemble methods, such as Random Forest and XGBoost classifiers, are expected to produce more accurate results.

Table 3 shows the prediction results for the chosen classifiers as measured by the performance metrics. The results of the classifiers in terms of accuracy, precision, recall and F1-Score are comparable. However, XGBoost clearly produces the best results across all the performance measures.

Table 3: Performance Results

Classifier	Accuracy	Precision	Recall	F1-Score
Logistic Regression	87%	84%	82%	83%
XGBoost	**88%**	**85%**	**83%**	**84%**
SVM	86%	83%	81%	82%
Random Forest	87%	85%	79%	82%
Naïve Bayes	84%	79%	78%	79%

Moreover, Logistic Regression is a classification algorithm so it is best applied to categorical data, hence it can also produce accurate results in certain datasets. Figure 2 shows the confusion matrices of each classifier. XGboost proved to be the winner among the used classifiers as it has the least number of false positive and false negative resulted in 1528 reviews.

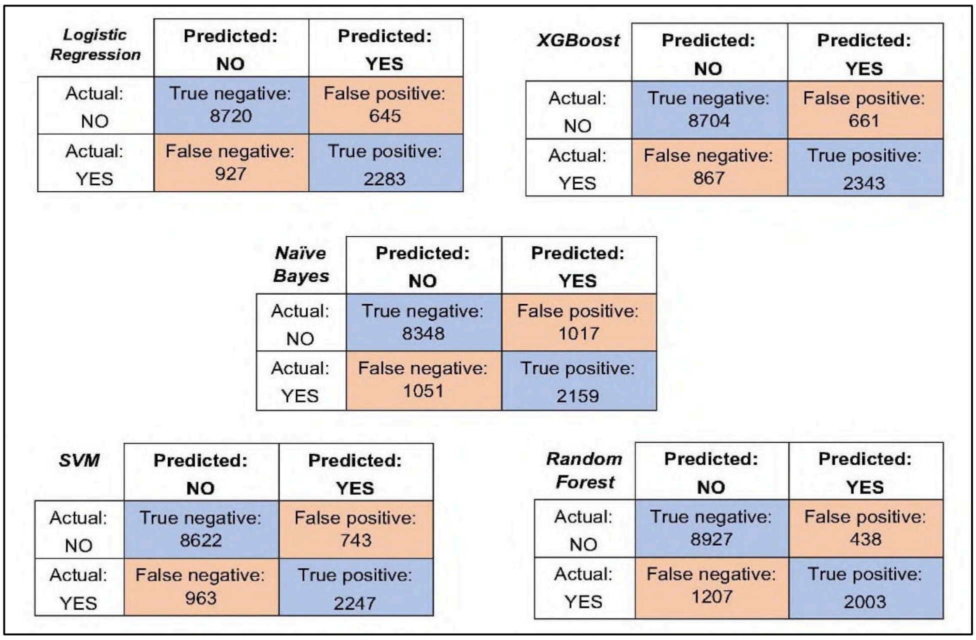

Figure 2: Confusion matrices for each classifier

7 CONCLUSION

In 2020, nearly half of the world's population, 3.6 billion people used social media. This provides an important source of data that can be studied and analysed and be used as a useful source of information. The number and frequency of people flying has risen considerably as a result of globalisation. Hence, analysing passenger feedback has become critical for airport management in order to improve the services provided within an airport. Developing an automated and effective model to accurately assess and analyse passenger input will allow stakeholders to not only improve the quality of services provided, but also to save time and effort compared to traditional techniques such as user survey and expert consultations. In this paper, we extracted data from the SKYTRAX website covering a time span of 16 years, pre-processed the dataset by removing stopwords and applying the TF-IDF to convert the review text into numerical form and analysed the resulting data. Our results have shown that given a sufficiently large dataset and good pre-processing, conventional machine learning algorithms can produce accurate and reliable results in the aviation sector. The assessment of the results was carried out using the metrics of accuracy, recall, precision and F1-score. The XGBoost algorithm outperformed the other selected approaches for extracting sentiments in this study. Our model could be further developed to automatically identify and rank in terms of importance the challenges that the passengers are facing, hence, provide further assistance to airport operators.

DATA ACCESSIBILITY

The code used for the simulations is freely available under the MIT license and can be downloaded from the Cranfield University repository: https://doi.org/10.17862/cranfield.rd.20169254.v1

REFERENCES

[1] "Air transport, passengers carried | Data." [Online]. Available: https://data.worldbank.org/indicator/IS.AIR.PSGR?end=2018&start=1970. [Accessed: 30-Jan-2022].

[2] S. E. Zaharia and C. V. Pietreanu, "Challenges in airport digital transformation," in Transportation Research Procedia, 2018, vol. 35, pp. 90–99, doi: 10.1016/j.trpro.2018.12.016.

[3] "A-Z Airport Reviews - SKYTRAX." [Online]. Available: https://www.airlinequality.com/review-pages/a-z-airport-reviews/. [Accessed: 28-Jan-2022].

[4] "Google Maps." [Online]. Available: https://www.google.com/maps/@21.5848387,38.9160206,10z. [Accessed: 28-Jan-2022].

[5] "Trip Advisor." [Online]. Available: https://www.tripadvisor.co.uk/.

[6] J. Ramos, "Using TF-IDF to Determine Word Relevance in Document Queries Juan," Urol. Clin. North Am., vol. 2, no. 1, pp. 29–48, 2003.

[7] C. H. Yeh and Y. L. Kuo, "Evaluating passenger services of Asia-Pacific international airports," Transp. Res. Part E Logist. Transp. Rev., vol. 39, no. 1, pp. 35–48, 2003, doi: 10.1016/S1366-5545(02)00017-0.

[8] H. Jiang and Y. Zhang, "An assessment of passenger experience at Melbourne Airport," J. Air Transp. Manag., vol. 54, pp. 88–92, Jul. 2016, doi: 10.1016/j.jairtraman.2016.04.002.

[9] D. Fodness and B. Murray, "Passengers' expectations of airport service quality," J. Serv. Mark., 2007, doi: 10.1108/08876040710824852.

[10] S. Kiliç and T. O. Çadirci, "An evaluation of airport service experience: An identification of service improvement opportunities based on topic modeling and sentiment analysis," Res. Transp. Bus. Manag., no. xxxx, p. 100744, 2021, doi: 10.1016/j.rtbm.2021.100744.

[11] C. Song, J. Guo, and J. Zhuang, "Analyzing passengers' emotions following flight delays- a 2011–2019 case study on SKYTRAX comments," J. Air Transp. Manag., vol. 89, no. August, p. 101903, 2020, doi: 10.1016/j.jairtraman.2020.101903.

[12] E. Prabhakar, M. Santhosh, A. H. Krishnan, T. Kumar, and R. Sudhakar B B Student, "Sentiment Analysis of US Airline Twitter Data Using New Adaboost Approach," Int. J. Eng. Res. Technol., vol. 7, no. 01, pp. 1–3, 2019.

[13] D. Sulu, H. Arasli, and M. B. Saydam, "Air-Travelers' Perceptions of Service Quality during the COVID-19 Pandemic: Evidence from TripAdvisor.com," Sustain., vol. 14, no. 1, Jan. 2022, doi: 10.3390/SU14010435.

[14] M. H. Salih, B. D. Bala, M. Irene, and Jenkins Karl, “Analysis the Sentiment of Air-Traveller: A Comparative Analysis,” no. March, 2021, doi: 10.7763/IJCTE.2022.V14.1309.

[15] K. Lee and C. Yu, “Assessment of airport service quality: A complementary approach to measure perceived service quality based on Google reviews,” J. Air Transp. Manag., vol. 71, pp. 28–44, Aug. 2018, doi: 10.1016/j.jairtraman.2018.05.004.

[16] F. Å. Nielsen, “A new ANEW: Evaluation of a word list for sentiment analysis in microblogs,” CEUR Workshop Proc., vol. 718, pp. 93–98, 2011.

[17] J. C. Campbell, A. Hindle, and E. Stroulia, “Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data,” Art Sci. Anal. Softw. Data, vol. 3, pp. 139–159, 2015, doi: 10.1016/B978-0-12-411519-4.00006-9.

[18] A. Rane, “Sentiment Classification System of Twitter Data for US Airline Service Analysis,” 2018 IEEE 42nd Annu. Comput. Softw. Appl. Conf., vol. 01, pp. 769–773, 2018, doi: 10.1109/COMPSAC.2018.00114.

[19] L. Martin-Domingo, J. C. Martín, and G. Mandsberg, “Social media as a resource for sentiment analysis of Airport Service Quality (ASQ),” J. Air Transp. Manag., vol. 78, pp. 106–115, Jul. 2019, doi: 10.1016/j.jairtraman.2019.01.004.

[20] S. Gitto and P. Mancuso, “Improving airport services using sentiment analysis of the websites,” Tour. Manag. Perspect., vol. 22, pp. 132–136, Apr. 2017, doi: 10.1016/j.tmp.2017.03.008.

[21] D. Hutchinson, V. Bogicevic, W. Yang, A. Bilgihan, and M. Bujisic, “Airport service quality drivers of passenger satisfaction,” Tour. Rev., vol. 68, no. 4, pp. 3–18, 2013, doi: 10.1108/TR-09-2013-0047.

[22] K. Mizufune and S. Katsumata, “Joint classification model of topic and polarity: Finding satisfaction and dissatisfaction factors from airport service review,” IEEE Int. Conf. Data Min. Work. ICDMW, vol. 2018-Novem, pp. 856–863, 2019, doi: 10.1109/ICDMW.2018.00126.

[23] S. Bird, “NLTK: The natural language toolkit,” COLING/ACL 2006 - 21st Int. Conf. Comput. Linguist. 44th Annu. Meet. Assoc. Comput. Linguist. Proc. Interact. Present. Sess., pp. 69–72, 2006.

[24] A. Poornima and K. S. Priya, “A Comparative Sentiment Analysis of Sentence Embedding Using Machine Learning Techniques,” 2020 6th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2020, pp. 493–496, 2020, doi: 10.1109/ICACCS48705.2020.9074312.

[25] H. Hasanli and S. Rustamov, “Sentiment Analysis of Azerbaijani twits Using Logistic Regression, Naive Bayes and SVM,” 13th IEEE Int. Conf. Appl. Inf. Commun. Technol. AICT 2019 - Proc., 2019, doi: 10.1109/AICT47866.2019.8981793.

[26] D. A. Al-Qudah, A. M. Al-Zoubi, P. A. Castillo-Valdivieso, and H. Faris, “Sentiment analysis for e-payment service providers using evolutionary extreme gradient boosting,” IEEE Access, vol. 8, pp. 189930–189944, 2020, doi: 10.1109/ACCESS.2020.3032216.

[27] K. Afifah, I. N. Yulita, and I. Sarathan, “Sentiment Analysis on Telemedicine App Reviews using XGBoost Classifier,” pp. 22–27, 2022, doi: 10.1109/icaibda53487.2021.9689762.

[28] N. Zainuddin and A. Selamat, “Sentiment analysis using Support Vector Machine,” I4CT 2014 - 1st Int. Conf. Comput. Commun. Control Technol. Proc., no. I4ct, pp. 333–337, 2014, doi: 10.1109/I4CT.2014.6914200.

[29] R. Passonneau, “Sentiment Analysis of Twitter Data,” no. June, pp. 30–38, 2011.

[30] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? Sentiment Classification using Machine Learning Techniques,” 2002.

[31] A. Gupte, S. Joshi, P. Gadgul, and A. Kadam, “Comparative Study of Classification Algorithms used in Sentiment Analysis,” Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 5, pp. 6261–6264, 2014.

[32] A. Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining,” Proc. 7th Int. Conf. Lang. Resour. Eval. Lr. 2010, pp. 1320–1326, 2010, doi: 10.17148/ijarcce.2016.51274.

[33] A. Prabhat and V. Khullar, “Sentiment classification on big data using Naïve bayes and logistic regression,” 2017 Int. Conf. Comput. Commun. Informatics, ICCCI 2017, 2017, doi: 10.1109/ICCCI.2017.8117734.

[34] C. S. Division, M. Park, P. Langley, and P. Smyth, “Bayesian Network Classifiers *,” vol. 163, pp. 131–163, 1997.

[35] J. Chen, H. Huang, S. Tian, and Y. Qu, “Expert Systems with Applications Feature selection for text classification with Naïve Bayes,” Expert Syst. Appl., vol. 36, no. 3, pp. 5432–5435, 2009, doi: 10.1016/j.eswa.2008.06.054.

[36] J. Brownlee, “Machine Learning Mastery With Python Understand Your Data, Create Accurate Models and Work Projects End-To-End,” vol. 91, pp. 399–404, 2017.

[37] M. Buckland and F. Gey, “The Relationship between Recall and Precision,” vol. 45, no. 1, pp. 12–19, 1994.