

Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes

Rachmawan Adi Laksono

Department of Information Technology

Management

Institut Teknologi Sepuluh Nopember

Surabaya, Indonesia

rachmawanadi@gmail.com

Kelly Rossa Sungkono

Department of Informatics

Institut Teknologi Sepuluh Nopember

Surabaya, Indonesia

kelly@its.ac.id

Riyanarto Sarno

Department of Informatics

Institut Teknologi Sepuluh Nopember

Surabaya, Indonesia

riyanarto@if.its.ac.id

Cahyaningtyas Sekar Wahyuni

Department of Informatics

Institut Teknologi Sepuluh Nopember

Surabaya, Indonesia

cahyaningtyas.18051@mhs.its.ac.id

Abstract— Sentiment analysis is one method for classifying documents to identify positive or negative opinions. Customer satisfaction has an essential point for customer service. Customer behaviour is currently doing a lot of reviews in online media such as on trip advisor. A restaurant is a business that requires more attention in the service to consumers by improving service to customers continuously. This study tries to classify Surabaya restaurant customer satisfaction using Naïve Bayes. Data sampling is crawling by using WebHarvy Tools. The result from this research shows that these two methods get the customer response accurately and Naïve Bayes method is more accurate than TextBlob sentiment analysis with a different accuracy of 2.9%.

Keywords— *customer satisfaction, Naïve Bayes, sentiment analysis, textBlob*

I. INTRODUCTION

Customer satisfaction is an opinion or feeling between expectation and reality obtained by consumers [1]. Today, many customers write opinions in the form of reviews about their obtained satisfaction on online media, such as TripAdvisor. Customer reviews on the online media become important as it might increase the popularity of the product or service sold by the seller.

A restaurant is a business [2], [3] that prepares and serves food for customers and exchanges for a certain amount of money. Although improving quality through this method is considered useful, only some restaurants use customer satisfaction analysis to improve their services. Also, many algorithms might be used for doing the study [4], [5].

Restaurant customer satisfaction research through reviews on online media such as TripAdvisor is still rare. Generally, restaurant customer satisfaction analyses through product data, nutrition data and food preparation [6]. One of the previous studies is an analysis of hotel customer satisfaction in Ponorogo district. The study used hotel customer review data on TripAdvisor [7].

Restaurant reviews on TripAdvisor are still in the form of text, customer reviews are included in the text mining category, the results of these data will be classified into two values, positive or negative [8]. Retrieving data on TripAdvisor using WebHarvy software, for preprocessing review data such as remove stopword, remove punctuation is

done with the help of Python, while for classifying data using Waikato Environment for Knowledge Analysis (WEKA) software with the Naïve Bayes method and also using TextBlob which is a python-based sentiment analyzer to compare. Naïve Bayes is chosen because this method has been widely implemented in sentiment analysis [9]–[11].

This aim of this study is to analyze restaurant customer reviews from online TripAdvisor in the best ten Surabaya restaurants. Another aim is to find the best method for analyzing restaurant customer review data by comparing Naïve Bayes method and TextBlob [12], [13] sentiment analysis since the two methods have fundamental differences in terms of calculations.

II. LITERATURE REVIEW

Customer satisfaction is an essential concern in the field of marketing and research in terms of consumer behaviour. As in the habits of hotel consumers when they get excellent service, they will transmit to others mouth to mouth [14].

Text mining or retrieval of data from a collection of documents stores frequently with the help of analysis tools or manuals [13]. Through the analysis process of several text mining perspectives, information can be produced that can be used to increase profits and services.

Sentiment analysis is used to find opinions from the author about a specified entity [15]. Sentiment analysis of a review is an opinion investigation of a product [16]. The basis of sentiment analysis is using Natural Language Processing (NLP), text analysis and some computational portions to extract or omit unnecessary parts to see the pattern of the sentence negative or positive [17].

In the 18th century, Reverend Thomas Bayes developed a method known as Naïve Bayes that used probability and opportunity approaches. The workings of the Naïve Bayes algorithm can be seen in Equation (1). Naïve Bayes calculates future probability predictions from data or experiences that have been given, based on the opportunity point of view [18]. One characteristic of the Naïve Bayes Classification is the existence of independent input variables which assume the presence of an articlar feature from a class that is mutually independent of other features [19].

$$P(H_j | x) = \frac{P(x|H_j)P(H_j)}{P(x)} \quad (1)$$

where:

- $P(H_j | x)$ = states the probability arises H_j if known x .
- $P(x|H_j)$ = The likelihood function of H_j to x
- $P(H_j)$ = Prior probability
- $P(x)$ = evidence

Methods of precision, recall, and accuracy is used to check the accuracy of the results of the process. A confusion matrix is created to provide performance classification data. Elements of confusion matrix in Fig. 1 are True positive (TP) when both human and method predict are positive and True Negative (TN) when both human and method predict are negative. False negative (FN) is used when the human prediction is positive while method prediction is negative and False Positive (FP) is used when the human prediction is negative while method prediction is positive [20].

The level of accuracy between what the user wants and the results of the system process is called Precision, can be seen in Equation (2), whereas Recall is the average success of the system in the process of finding information, listed in Equation (3), Precision and Recall calculations are used to avoid measurement errors for deviation values as shown in Equation (4). Accuracy is the degree of truth between the predictive value, and the actual value is shown in Equation (5). Precision value is obtained by dividing TP with the positive results obtained by the method. Because this paper analyzes customer review, so the positive results are the reviews classifying as positive reviews. Recall value is obtained by dividing TP with positive results based on the real data. F Score value uses the Recall value and Precision value. The accuracy is the division of the amount of TP and TN to the amount of data.

$$\text{Precision} = \frac{TP}{(TP+FP)} \quad (2)$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (3)$$

$$F \text{ Score} = \frac{(2 \times \text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4)$$

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (5)$$

where: TP = the positive review of the real data is classified as the positive review obtained by the method
 TN = the negative review of the real data is classified as the negative result obtained by the method
 FP = the negative review of the real data is classified as the positive review obtained by the method
 FN = the positive review of the real data is classified as the negative review obtained by the method

		Real data	
		Positive Review	Negative Review
Methods	Positive Review	TP	FP
	Negative Review	FN	TN

Fig. 1. Confusion matrix

As a comparison, the results of this study are also compared to the TextBlob sentiment analysis, a sentiment analyzer that has a Natural Language Toolkit (NLTK) and Pattern processing basis [21]. TextBlob can also be used for text mining, text processing modules for python winners, and even text analysis. TextBlob also provides simple APIs for general Natural Language Processing (NLP) processing such as part-of-speech tagging, tokenizing sentences, noun phrase extraction, sentiment analysis, classification, translation [22].

WEKA, is one of the workbenches that has many choices of machine learning methods for research. Initiated since 1992, the use of WEKA allows users to try and compare several different machine learning methods on the new data set quickly, and there are several Algorithms available in several languages [23]. WEKA has been very accepted in various academics, the business environment, to be used as a data mining research tool. The analysis is working on environment using WEKA 3.8.2.

One of its abilities is the classification method either with supervision or no supervision. Some popular classifications are the Naïve Bayes method, Decision tree J48, Random Forest, k-Nearest Neighbors (k-NN), Sequential Minimal Optimization (SMO) or also called the Support Vector Machine (SVM) method..

III. RESEARCH METHOD

This paper takes restaurant review data on TripAdvisor, especially restaurants in Surabaya by web crawling method, and will be analyzed using the Naïve Bayes method on WEKA and a comparison with the TextBlob sentiment analysis. There are several steps of the research method.

A. Data collection

Retrieval of customer review data is done by crawling by using the WebHarvy tool as in the methods Fig. 2. Data is taken from web TripAdvisor, as in Fig. 3. TripAdvisor is the largest travel community review site on the web. This web was introduced in the year of 2000, and now covers 212.000 hotels, more than 30.000 destinations, and 74.000 attractions around the world. Fig. 3 shows an overview of reviews from TripAdvisor.

In Fig. 3, there are 2 data shown, in the red box is the title of the review, while the green box is a review of the customer. This paper will use customer review data as a data source.



Fig. 2. Data mining process

B. Data processing and analysis

Fig. 4 shows the work steps on processing and analysis data. On the processing data and analysis, there are several steps that the procedure must be followed. The Trim lowercase process is done to change all the letters in small bricks not mixed large and small so that the uniforms are like this: "Good", "Noisy" becomes "noisy" good.

← → C TripAdvisor LLC [US] | https://www.tripadvisor.com/Restaurant_Review-g297715-d11549002-Reviews-Ayam_Dakar_Frimarasa-Surabaya_Cat

Overview **Reviews** Location Q&A Details

iwillkiwil
Surabaya, Indonesia
143 12

Reviewed 1 week ago via mobile

Delicious

Its guramee bakar is highly recommended! Ayam bakar is another recommendation. Eat with sambal pencit and kangkung hotplate

Date of visit: December 2018

Thank iwillkiwil



Fig. 3. TripAdvisor Web

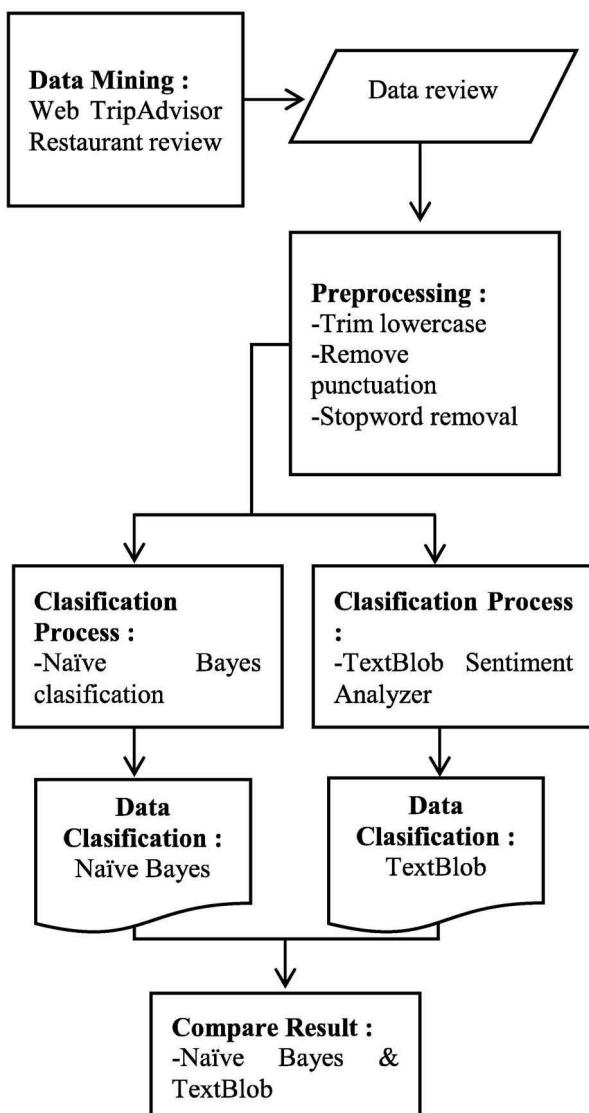


Fig. 4. Data analysis process

Stopword removal is a process of removing words that often appear but do not have meaning in languages, such as "the", "a", "an", "in". Like the example below: "let me preface this review," said gardeners " became " let preface review, saying, mad keen gardeners ".

The remove punctuation process is the process of removing punctuation that often appears and usually does not have much meaning like "-", "/", ";", "?". After the preprocessing text is complete, the next process will be text calcification using the Naïve Bayes method and the comparison using Blob Text.

IV. RESULT

Crawling techniques take the process of receiving customer review data on Tripadvisor, especially restaurants in Surabaya. The data taken is data on restaurant names and reviews. The results of the data collection obtain 337 data, 269 data are used for training data and 68 data for testing data. Some data results from the data collection process as shown in TABLE I. There are three kinds of information in TABLE I, i.e. several reviews (No.), the name of the restaurant, and the reviews. There are only ten reviews that are shown in TABLE I.

The next step is to prepare the review data for training and testing data. Trim lowercase or make the letters uniform into lowercase letters. TABLE II are examples of trim lowercase processing results. As seen in the first review of TABLE II, the sentence "About the steak, well, the taste was nice" has been processed into "about the steak, well, the taste was nice".

The next preprocessing is shown in Fig. 5. Fig. 5 explains the process of removing punctuation and words that often appear in the text but have no meaning. As in Fig. 5, an example of the results of eliminating punctuation and removing stop words.

TABLE I. PART OF THE RESULT OF DATA COLLECTION

No.	Restaurant Name	Reviews
1	Steak Hut Manyar Kertoarjo	About the steak, well, the taste was nice
2	Steak Hut Manyar Kertoarjo	The price is standard
3	Steak Hut Manyar Kertoarjo	I think, they need more innovation for their menu
4	Steak Hut Manyar Kertoarjo	About the place, felt comfortable
5	Steak Hut Manyar Kertoarjo	overall, worth to try this restaurant
6	Steak Hut Manyar Kertoarjo	best luck
7	Steak Hut Manyar Kertoarjo	Affordable price, good ambience, recommended for family event
8	Steak Hut Manyar Kertoarjo	So many promotion from any credit card
9	Steak Hut Manyar Kertoarjo	Great service & affordable price for steak
10	Steak Hut Manyar Kertoarjo	Usually I visited this restaurant once in a month

TABLE II. TRIM LOWERCAST RESULT

No.	Before Trim Lowercase	After Trim Lowercase
1	About the steak, well, the taste was nice	about the steak, well, the taste was nice
2	The price is standart	the price is standart
3	I think, they need more innovation for their menu	i think, they need more innovation for their menu
4	About the place, felt comfortable	about the place, felt comfortable
5	overall, worth to try this restaurant	overall, worth to try this restaurant
6	best luck	best luck
7	Affordable price, good ambience, recommended for family event	affordable price, good ambience, recommended for family event
8	So many promotion from any credit card	so many promotion from any credit card
9	Great service & affordable price for steak	great service & affordable price for steak
10	Usually I visited this restaurant once in a month	usually i visited this restaurant once in a month

“nice and clean restaurant, good service. good food, the best is garlic steam fish.. so yummy... price also ok”

“nice clean restaurant good service good food best garlic steam fish yummy price ok”

Fig. 5. Remove stopwords process

Time taken to Build model : 0.27 seconds
Correctly Classified 237 88.1041 %
Incorrectly Classified 32 11.8959 %

==== Confusion Matrix ====

a b <-- classified as
63 14 | a = -1
18 174 | b = 1

Fig. 6. Training data and result

TABLE III. THE CLASSIFICATION OF TESTING DATA BY USING NAÏVE BAYES AND TEXTBLOB

No.	Review	NB	TB	EX
1.	steak well taste nice	1	1	1
2.	place felt comfortable	1	1	1
3.	overall worth try restaurant	1	1	1
4.	best luck	-1*	1	1
5.	affordable price good ambience recommended family event	1	1	1
6.	many promotion credit card	-1*	1	1
7.	great service affordable price steak	1	1	1
8.	usually visited restaurant month	1	-1*	1
9.	recommended steak restaurant surabaya	1	1	1
10.	crowded	1	1	1
11.	cozy place decent food	1	1	1
12.	nice environment atmosphere	1	1	1
13.	nice meal larger portion perfect	1	1	1
14.	cozy place good service	1	1	1
15.	good food perfect salad	1	1	1
16.	chicken cordon blue delicious	1	1	1
17.	love	1	1	1
18.	thumbs up	-1*	1	1
19.	lunch family surabaya	1	1	1
20.	good taste good service steak hut manyar ketoarjo	-1*	1	1
21.	recommended	1	1	1
22.	great lunch family	1	1	1
23.	nice steak hut salad nz sirloin steak	-1*	1	1
24.	come next	-1*	1	1
25.	convenience environment great taste steak	1	1	1
26.	also love burger new mozzarella chicken schnitzel	1	1	1

27.	thanks good service delicious beef steak i will back here	-1*	1	1
28.	thanks lot	-1*	1	1
29.	steak hut restaurant specialties steak menu	-1*	1	1
30.	usually promotion price cooperation credit card certain bank	1	-1*	1
31.	good delicious food clean spacious place great large groups good friendly service	1	1	1
32.	good food good price comfort place	1	1	1
33.	must visit indonesian food surabaya	1	1	1
34.	grilled fish shrimp etc.	1	1	1
35.	think need innovation menu	-1*	1	1
36.	crowded visit weekend	1*	1*	-1
37.	area nonsmoking smoking small space	1*	-1	-1
38	come thought small restaurant totally wrong	-1	-1	-1
39.	long time tried dinner fish bit small tempe also small portion nasi goreng merah amazed us super tiny bit portion small bowl its	1*	-1	-1
40.	food price expensive quality average portion small	-1	-1	-1
41.	sorry coming back again	-1	-1	-1
42.	disappointed female staff	-1	-1	-1
43.	want ask spoon serve food without spoon take food and one coming	-1	1*	-1
44.	and response	-1	1*	-1
45.	look back and yell louder voice tunggu sebentar	-1	1*	-1
46.	end she is coming table all	-1	1*	-1
47.	and sometime ask staff take food without spoon food	-1	-1	-1
48.	know staff makes bad experience coming here	-1	-1	-1
49.	ac working well here	-1	1*	-1
50.	not cold	-1	1*	-1
51.	little bit expensive crabs prices beside parking area difficult imagine raining season	1*	1*	-1
52.	noisy lots loud people here	-1	1*	-1
53.	privacy dont come romantic dinner	-1	1*	-1
54.	good seafood want go prepared noisy evening	1*	1*	-1
55.	restaurant quite big crowded	1*	1*	-1

56.	you starving dont go there	-1	1*	-1
57.	quite take time especially restaurant crowded	-1	1*	-1
58.	lobster expensive	-1	-1	-1
59.	large restaurant crowded local foreign	1*	1*	-1
60.	prepare serve late menu	-1	-1	-1
61.	always crowded	-1	1*	-1
62.	thing parking fee flat rate longer expensive ticket is	-1	1*	-1
63.	place always crowd dinner time	1*	1*	-1
64.	variety foods many nothing special	1*	1*	-1
65.	enjoyed ice cream solace	-1	1*	-1
66.	The food terribly overrated	1*	-1	-1
67.	The chicken dry tough marinade taste like kecap manis	-1	-1	-1
68.	The price expensive quality poor rating	-1	-1	-1

where: NB : Naïve Bayes

TB : TextBlob

EX : Expert

TABLE IV. CONFUSION MATRIX OF NAÏVE BAYES AND TEXTBLOB

		Naïve Bayes		TextBlob	
TP	FP	25	10	34	19
FN	TN	9	24	2	13

After going through preprocessing the data, a model was made with training data using WEKA 3.8.2 with Naïve Bayes classifiers. The model data is stored and will be used for evaluation with other data or testing data, Fig. 6 is the result of the data model. As seen in Fig. 6, this research obtains 63 TN and 174 TP by using Naïve Bayes. The total of reviews that are correctly classified is 237 reviews.

After training data with the Naïve Bayes method, then re-evaluate the model with testing data. The testing data was also tested in the TextBlob Sentiment analyzer for comparison of accuracy with the Naïve Bayes method. The following is the result of testing data with the Naïve Bayes and TextBlob models that have been implemented written as in TABLE III. TABLE III shows 68 testing data. Each review will be classified by using Naïve Bayes as the method of this research, TextBlob as the comparison method, and by an expert. 1 is denoted the positive review and -1 is denoted as the negative review. The sign (*) means the result is different from the expert judgment.

All the results obtained by Naïve Bayes and Textblob will be analyzed by using confusion matrix. TABLE IV shows the confusion matrix of those two sentiment analysis method. Based on TABLE IV, Naïve Bayes obtains 25 TP and 24 TN. Then, TextBlob collects 34 TP and 13 TN. Naïve Bayes has a higher FN number than TextBlob.

Conversely, Naïve Bayes has a lower FP number than TextBlob. The accuracy of the confusion matrix listed in TABLE IV is processed by using Equation (5). The accuracy of Naïve Bayes is shown in Equation (6), and the accuracy of TextBlob is shown in Equation (7). The results of the calculation of the above formula for the Naïve Bayes method have data accuracy of 72.06%, whereas by using the TextBlob analyzer obtained an accuracy of 69.12%. From these results, the Naïve Bayes method using WEKA tools has a higher percentage of accuracy than the TextBlob sentiment analysis.

$$\text{Accuracy Naïve Bayes} = 72.06\% \quad (6)$$

$$\text{Accuracy TextBlob} = 69.12\% \quad (7)$$

V. CONCLUSION

Based on the result of this paper that done on restaurant customer reviews in Surabaya, customer satisfaction analysis might be used by the Naïve Bayes classification method and TextBlob sentiment analysis that might be able to learn sentiment from customers, since customer satisfaction is essential in terms of the restaurant business. The results also indicate that the Naïve Bayes method has a value of 72.06% accuracy and is slightly better (2.94%) than TextBlob sentiment analysis. Further research can be done by increasing the number and variety of review data, or by other methods, to increase the value of accuracy.

ACKNOWLEDGMENT

The authors would like to thank to Institut Teknologi Sepuluh Nopember, The Ministry of Research, Technology and Higher Education of Indonesia, *Direktorat Riset dan Pengabdian Masyarakat*, and *Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia* for supporting the research

REFERENCES

- [1] K. Ilieska, "Customer satisfaction index--as a base for strategic marketing management," *TEM Journal*, vol. 2, no. 4, p. 327, 2013.
- [2] Y. Caesarita and K. R. Sungkono, "Identifying Bottlenecks and Fraud of Business Process using Alpha ++ and Heuristic Miner Algorithms (Case study : CV . Wicaksana Artha)," pp. 143–148, 2017.
- [3] R. Sarno and K. R. Sungkono, "Coupled Hidden Markov Model for Process Discovery of Non-Free Choice and Invisible Prime Tasks," *Procedia Computer Science*, vol. 124, pp. 134–141, 2018. <http://doi.org/10.1016/j.procs.2017.12.139>.
- [4] M. Fikri and R. Sarno, "A Comparative Study of Sentiment Analysis using SVM and SentiWordNet," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 13, no. 3, 2019.
- [5] B. Rintyarna, R. Sarno, and C. Fatichah, "Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context," *International Journal of Information and Decision Sciences*, vol. 11, 2018.
- [6] J. E. Mills and L. Thomas, "Assessing customer expectations of information provided on restaurant menus: A confirmatory factor analysis approach," *Journal of Hospitality & Tourism Research*, vol. 32, no. 1, pp. 62–88, 2008.
- [7] D. A. K. Khotimah and R. Sarno, "Sentiment Detection of Comment Titles in Booking. com Using Probabilistic Latent Semantic Analysis," in *2018 6th International Conference on Information and Communication Technology (ICoICT)*, 2018, pp. 514–519.
- [8] D. Gräßner, M. Zanker, G. Fiedl, M. Fuchs, and others, *Classification of customer reviews based on sentiment analysis*. Citeseer, 2012.
- [9] B. Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," in *Data and Software Engineering (ICoDSE), 2015 International Conference on*, 2015, pp. 170–174.
- [10] F. S. Fitri, M. N. S. Si, and C. Setianingsih, "Sentiment Analysis on the Level of Customer Satisfaction to Data Cellular Services Using the Naive Bayes Classifier Algorithm," in *2018 IEEE International Conference on Internet of Things and Intelligence System (IOT AIS)*, 2018, pp. 201–206.
- [11] U. W. Wijayanto and R. Sarno, "An Experimental Study of Supervised Sentiment Analysis Using Gaussian Naive Bayes," in *2018 International Seminar on Application for Technology of Information and Communication*, 2018, pp. 476–481.
- [12] H. Gauba, P. Kumar, P. P. Roy, P. Singh, D. P. Dogra, and B. Raman, "Prediction of advertisement preference by fusing EEG response and sentiment analysis," *Neural Networks*, vol. 92, pp. 77–88, 2017.
- [13] A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine learning-based sentiment analysis for twitter accounts," *Mathematical and Computational Applications*, vol. 23, no. 1, p. 11, 2018.
- [14] K. Berezina, A. Bilgihan, C. Cobanoglu, and F. Okumus, "Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews," *Journal of Hospitality Marketing & Management*, vol. 25, no. 1, pp. 1–24, 2016.
- [15] R. Feldman, "Techniques and applications for sentiment analysis," *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2013.
- [16] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," *Procedia Computer Science*, vol. 17, pp. 26–32, 2013.
- [17] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *Journal of machine learning research*, vol. 9, no. Aug, pp. 1871–1874, 2008.
- [18] M. Jupri and R. Sarno, "Taxpayer compliance classification using C4. 5, SVM, KNN, Naïve Bayes and MLP," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, 2018, pp. 297–303.
- [19] T. Pranckevičius and V. Marcinkevičius, "Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification," *Baltic Journal of Modern Computing*, vol. 5, no. 2, pp. 221–232, 2017.
- [20] D. L. Gupta, A. K. Malviya, and S. Singh, "Performance analysis of classification tree learning algorithms," *International Journal of Computer Applications*, vol. 55, no. 6, 2012.
- [21] B. Agarwal, N. Mittal, P. Bansal, and S. Garg, "Sentiment analysis using common-sense and context information," *Computational intelligence and neuroscience*, vol. 2015, p. 30, 2015.
- [22] S. Vijayarani, R. Janani, and others, "Text mining: open source tokenization tools-an analysis," *Advanced Computational Intelligence: An International Journal (ACII)*, vol. 3, no. 1, pp. 37–47, 2016.
- [23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.