Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine

Anang Anggono Lutfi[1), Adhistya Erna Permanasari[2)*], Silmi Fauziati[3)]

[1)2)3)] *Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada*
Jl. Grafika 2, Yogyakarta - 55281, Indonesia

[1)]anang.anggono.lutfi@mail.ugm.ac.id

[2)]adhistya@ugm.ac.id

[3)]silmi@ugm.ac.id

Abstract

The online store is changing people's shopping behavior. Despite the fact, the potential customer's distrust in the quality of products and service is one of the online store's weaknesses. A review is provided by the online stores to overcome this weakness. Customers often write a review using languages that are not well structured. Sentiment analysis is used to extract the polarity of the unstructured texts. This research attempted to do a sentiment analysis in the sales review. Sentiment analysis in sales reviews can be used as a tool to evaluate the sales. This research intends to conduct a sentiment analysis in the sales review of Indonesian marketplace by utilizing Support Vector Machine and Naive Bayes. The reviews of the data are gathered from one of Indonesian marketplace, Bukalapak. The data are classified into positive or negative class. TF-IDF is used to feature extraction. The experiment shows that Support Vector Machine with linear kernel provides higher accuracy than Naive Bayes. Support Vector Machine shows the highest accuracy average. The generated accuracy is 93.65%. This approach of sentiment analysis in sales review can be used as the base of intelligent sales evaluation for online stores in the future.

I. INTRODUCTION

Information technology is utilized in many sectors of communitie's life. The utilization of information technology helps people to solve various problems. The online store is an example of information technology utilization in the economic sector. The existence of online stores has indirectly changed people's shopping behavior in purchasing goods and service. The change of behavior happens around the world, including in Indonesia.

In 2016, the internet user in Indonesia reached 132.7 million people [1]. Over half of the users (84.2 million) have ever conducted an online transaction. There are around 46.1 million users doing online transaction more than once a month, 18.8 million users doing it less than once in a month, 6.2 million users doing it once a week, and over 5 million doing it more than once a week. The data show that the active internet users in Indonesia are active in doing online transaction.

Online stores usually provide a facility for providing a review that is accessible to the visitors. The review contains a short description of testimony from prior users about the service or goods provided by the sellers. For sellers, a review can be a tool to evaluate their sales. For customers, the review can be a benchmark to assess the quality of the goods and service that are sold by the sellers. The review is helpful for the potential buyers to get familiar with the quality of the goods and services. That thing will influence the potential buyers to decide whether they will buy the products. The review may contain positive or negative testimony. The more positive review an online store has, the higher the trust of the customers will be. It works vice versa. The customers often write the

---

review using a short sentence and informal language. This will be confusing for the readers who are not familiar yet with online stores. Sentiment analysis is used to extract the polarity of the unstructured texts.

Sentiment analysis is a branch of text mining that intends to clarify a review into a certain class. The review can be classified into positive or negative class. This will help to figure out the polarity of the review. In a company scale, sentiment analysis of sales review can be used by the company's management board as a base of various decision-making processes in the company. Sentiment analysis can be a quite popular research domain. There are several machine learning approaches popularly used in sentiment analysis. Support Vector Machines (SVM), Naive Bayes (NB), and Maximum Entropy (ME) showed good results in text classification [2]. The approach that is employed in this research is the use of machine learning by comparing SVM and NB. The approach is employed to figure out the method that provides higher accuracy when the dataset of sales review is used.

Most of the researchers in sentiment analysis use social media as the data resources. Fiarni et.al conducted sentiment analysis of online stores in Indonesia, with social media as the main data sources [3]. The same is for research [4][5], social media is used as the data. Research of Muthia [6] focused on the review data from a website of the restaurant, while research [7] focused on tourism sites. Sghaier and Zrigui conducted sentiment analysis in one of the commercial sites, in Arabic [8]. Saragih and Girsang conducted a sentiment analysis in customer engagement on online transportation in Indonesia [5].

Based on the description above, sentiment analysis of Indonesian in online stores using a dataset from commercial websites is still rare to be found. This research focuses on the data on sales review in Indonesian taken from one of the biggest Indonesian marketplaces, Bukalapak. The data taken from the marketplace site will better illustrate the sales that occur rather than data taken from social media. This research proposed an analysis of the polarity of sales reviews on online stores. The sentiment analysis in sales reviews can be used as an evaluation tool for the store's sales.

II. RELATED WORKS

Some researchers in sentiment analysis had been conducted in various aspects. In research [3], Fiarni et.al conducted research in sentiment analysis of online store's reviews using Hierarchy Naive Bayes technique. The sentiment class was divided into positive, negative, and neutral. The data were taken from the review of the online stores given by the customers on Facebook. The gathered data were 1442 reviews, divided into training data and testing data. After various analyses, 217 training data were used. The object of the review was divided into 8 categories: material, product, price, quality, design, service, room exhibition, and general. The steps of the research were: preprocessing, feature extraction, and text classification using Naive Bayes Classifier (NBC). Classification process using NBC gave the accuracy value of 89.21%, precision of 97.25%, and recall of 89.83%.

Muthia conducted a research in sentiment analysis of restaurant reviews using Naive Bayes [6]. The data consist of 100 positive reviews and 100 negative reviews in Indonesian. The data are gathered from a restaurant website, Zomato. The preprocessing consists of two steps: tokenization and N-grams generate. Naive Bayes was used to classifying data into two classes: positive or negative. A comparison of the accuracy result of classification process with feature selection was done using a genetic algorithm or not. The result of the treatment showed that the classification process using Naive Bayes had the accuracy of 86.50%. While the classification process using Naive Bayes with the combination of feature selection using genetics algorithm had the accuracy of 90.50%.

A sentiment analysis of e-commerce in Arab was conducted by Sghaier dan Zrigui [8]. The data were in Arabic. They were gathered from commercial magazines and websites. The data consisted of 125 positive and 125 negative data. The preprocessing included normalization, segmentation, deletion of stop words, non-Arabic words removal, deletion and conversion of an emoticon, elongated words correction, special character removal. After preprocessing, the data were classified using SVM, K-Nearest Neighbors (KNN), and NB. The result shows that SVM and NB algorithms provided higher performance than KNN. The research showed that the precision reached 93.9% when using SVM.

A research of sentiment analysis in travel destination in Indonesia was conducted by Windasari and Eridani [7]. The data text was in Indonesian and gathered from a tourism website, Trip Advisor. The data gathered were given certain level manually: negative or positive. The preprocessing was conducted by conversion of an emoticon, cleansing, and case folding. Then, a feature extraction was done using stemming, stop words removal, negation conversion, and tokenization. The weighting of features was conducted using TF-IDF. Then, classification using SVM was done. The research showed that the accuracy reached 85%.

The other research of sentiment analysis in the online transportation industry in Indonesia was conducted by Windasari et.al [4]. The data of online transportation were extracted from Twitter. The texts were classified into positive and negative classes. The research procedure included data gathering, manual class labeling, preprocessing,

feature extraction, and classification. The classification used machine learning approach with SVM algorithm. The research resulted in the accuracy of 86%.

Saragih and Girsang conducted a sentiment analysis in customer engagement on online transportation in Indonesia [5]. The research tried to figure out, to what extent the customer's engage in the online transportation industry. The customer engagement rate could be figured out by doing analysis in the comments of Facebook and tweets on Twitter, on the accounts that belonged to the online transportation providers. The data were categorized into three classes; positive, negative, and neutral, using TF-IDF. The research shows that categories of "Feedback system by driver" and "Feedback system by user" had the most feedbacks. The negative comments (complaints) from the drivers got the highest rank.

III. METHODS

In the research, the sales reviews were classified into positive and negative classes. Some steps were required to get a good classification result. The steps were data gathering, preprocessing, feature extraction, classification, and evaluation. The research method diagram is available in Figure 1.
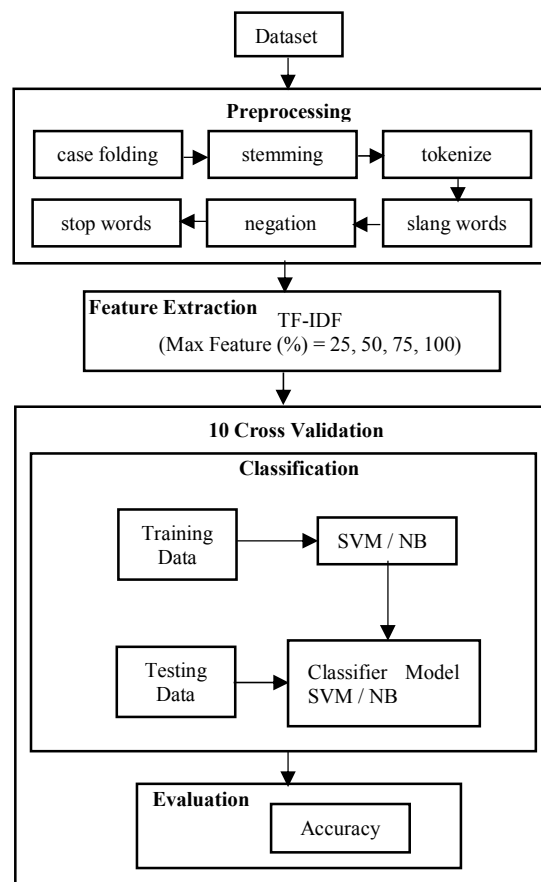

Figure 1. Research method

A. Dataset

The data of the review were taken from online stores in Bukalapak. After the data were gathered, they were then validated. Irrelevant data were removed. The data were labeled with the positive or negative class manually. There are 3177 reviews gathered for the research. They consist of 1521 negative reviews and 1656 positive reviews. Next step was preprocessing. After the preprocessing, the number of data was reduced from 3177 to 3077. From those data, 2770 data were used as training data and 307 data were used as testing data. The research used supervised learning approach so that the existing review data were divided into training data and testing data.

Other than that, the dictionary data of Indonesian and English stop words, slang words dictionary, and negation dictionary were required. The dictionary data were used in the preprocessing step. The dictionary data were acquired from the research of Prasetyo [9].

B. Preprocessing

The gathered review data were not well structured yet, therefore, some preprocessing steps were required to make them well structured. Structured data will simplify the classification process. Preprocessing was conducted through some steps, including case folding, stemming, tokenization, slang words conversion, negation words conversion, and stop words removal.

Case folding was the step of font conversion, changing all the letters into lower-cased letters. This step is the fundamental step that is used the most in the natural language processing. Stemming is a process to change a word into its stem from. The implementation of stemming uses Nazief Andriani algorithm with the help of libraries in Phyton, Sastrawi 1.0.1. Tokenization is a process to segregate word per word in a document.

The internet users often use informal words when input their messages. The informal words may be in the form of slang words or abbreviations that are often used in daily life like cpat (from "cepat" or fast), blum (from "belum" or not yet), and gak (from "tidak" or no). The conversion of slang words was conducted to solve this case. The slang words dictionary and their conversion utilize the dictionary based on Prasetyo research with some modification words adjusted to the needs of the research [9].

Negation conversion is a process of changing the word that follows a negation to become their antonym. It is conducted because the negation words may affect the polarity of the word. For example, tidak cepat ("not quick") is changed into lambat ("slow"). The negation dictionary is acquired from the research by Prasetyo with some modification words adjusted to the needs of the research [9].

Stop words are words that have no influence on the polarity of any review, but they still appear often. The stop words dictionary uses Indonesian and English stop words based on Prasetyo research with some modification words adjusted to the needs of the research [9].

C. Features Extraction

The feature extraction aims to extract fewer features only significant features shall be generated. The researchers extracted the features by doing the weighting of words. The words in a document were changed into their weight. The weighting of words was conducted using Term Frequency-Inverse Document Frequency (TF-IDF) method. TF-IDF is frequently used because it was relatively simple but it can generate high accuracy and recall [10]. The value of TF-IDF was calculated using (1).

$$\text{tf-idf}_{t,d} = \text{tf}_{t,d} \text{ x idf}_t \qquad (1)$$

Where tf-idf$_{t,d}$ is the value of TF-IDF in term t within document d. The value of tf$_{t,d}$ is generated from the value of term frequency in the term t from document d. While the value of idf$_t$ is the inverse document frequency from the term t. The value of idf$_t$ is generated from (2).

$$\text{idf}_t = \log (N/\text{df}_t) \qquad (2)$$

N is the number of all documents. Meanwhile, dft is the number of documents containing term t.

The weighting of words uses Phyton library, Scikit-learn [11]. After each word has its weighting, some experiments were conducted using 25%, 50%, 75%, and 100% of features with highest TF-IDF.

D. Text Classification

After the feature extraction, text classification was then conducted to define the sentiment of a document. The classification used machine learning approach with the algorithms of SVM and NB. The implementation of classification process used Scikit-learn library in Phyton[11]. In the research, SVM used linear kernel. SVM is a popular technique that is used for classification. This technique attempts to find the most optimum separation function (hyperplane) to separate data from different classes [12]. The illustration of a hyperplane in SVM can be seen in Figure 2.
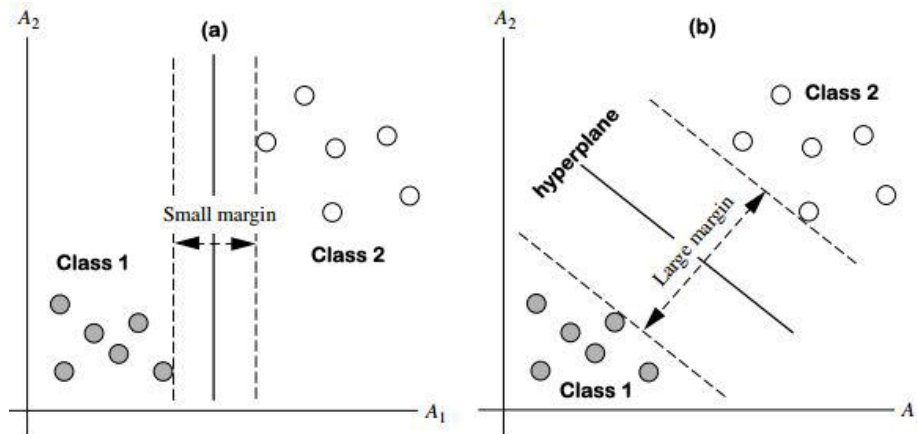
Figure 2. The Illustration of Hyper plane in SVM [13]

Figure 2(a) has the separation function that separates Class 1 and Class 2, however not optimum. While Figure 2(b) has the separation function that separates Class 1 and Class 2 effectively. In the SVM, the most optimum separation function is searched as illustrated in Figure 2(b).

The Bayes theorem is based on the statistics of probability and cost generated from the decision of the classification [12]. NB is one of the simple implementations of Bayes theorem. The formula of NB is written in (3).

$$P(B|A) = \frac{P(A|B)\ P(B)}{P(A)} \qquad (3)$$

Where $P(B|A)$ is the probability of the appearance of B when A is known. The value of $P(A|B)$ is the probability of the appearance of A if B is known. $P(A)$ is the probability of the appearance of A, while $P(B)$ is the probability of the appearance of B.

E. Evaluation

The evaluation is conducted by testing the system that has been created. The system is tested for its accuracy. The accuracy is measured to figure out the number of the documents to be classified correctly by the system. The accuracy is calculated using (4). Testing is conducted using 10 cross-validation techniques. This technique lets 10 times iteration of data division. In every iteration, the data is divided into 10 equal parts, 9 of which are training data and 1 of which is testing data.

$$A = \frac{(tp+tn)}{(tp+fp+tn+fn)} \qquad (4)$$

Where tp is the number of documents in positive class that is correctly classified by the system. Where tn is the number of documents in negative class that is correctly classified by the system. The number of positive class document that is incorrectly classified by the system is saved in fn variable. The number of negative class document that is incorrectly classified by the system is saved in fp variable.

IV. RESULTS

The number of data gathered is 3177. The data consists of 1521 negative reviews and 1656 positive reviews. The initial process begins with preprocessing data. The number of data was reduced to 3077 after preprocessing. The sample of preprocessing is available in Table I.

After going through the preprocessing step, the feature extraction is conducted using TF-IDF. The process generated 1452 words features. The experiment was conducted using SVM and NB accuracy results when it is using 25%, 50%, 75%, and 100% of the features with highest TF-IDF. The testing was conducted 10 cross-validation technique so that each experiment got 10 times iteration. The result is available in Table II.

TABLE I
PREPROCESSING SAMPLE RESULT

| Step	Result
initial review	tanggapan pesan Cpt bungkus Baik Barang Sudah Diterima Cepat Sampai Barang Sesuai Pesanan...tidak mengecewakan
case folding	tanggapan pesan cpt bungkus baik barang sudah diterima cepat sampai barang sesuai pesanan...tidak mengecewakan
stemming	tanggap pesan cpt bungkus baik barang sudah terima cepat sampai barang sesuai pesan tidak kecewa
tokenize	'tanggap' 'pesan' 'cpt' 'bungkus' 'baik' 'barang' 'sudah' 'terima' 'cepat' 'sampai' 'barang' 'sesuai' 'pesan' 'tidak' 'kecewa'
slang words	tanggap pesan cepat bungkus baik barang sudah terima cepat sampai barang sesuai pesan tidak kecewa
negation	tanggap pesan cepat bungkus baik barang sudah terima cepat sampai barang sesuai pesan puas
stop words	tanggap pesan cepat bungkus baik barang terima cepat barang sesuai pesan puas
final review	tanggap pesan cepat bungkus baik barang terima cepat barang sesuai pesan puas
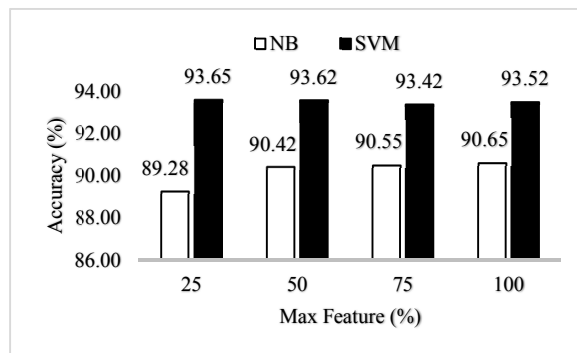
Figure 3. Accuracy average

TABLE II
ACCURACY RESULT

Iteration No.	Max Feature 25 %		Max Feature 50 %		Max Feature 75 %		Max Feature 100 %	
	NB	SVM	NB	SVM	NB	SVM	NB	SVM
1	82.08	95.77	82.41	96.42	82.41	96.09	82.41	96.42
2	99.67	95.44	99.67	95.11	99.67	94.46	99.67	95.11
3	92.51	93.16	91.21	92.83	89.58	92.18	89.90	92.18
4	94.46	93.81	93.49	93.49	92.83	93.49	92.83	93.49
5	99.02	89.90	98.70	89.90	98.37	89.90	98.05	90.55
6	92.83	99.35	96.42	99.35	96.42	99.35	96.74	99.02
7	92.18	98.70	97.72	98.70	97.72	98.37	97.72	98.05
8	96.09	98.70	96.42	98.70	96.74	98.37	96.74	98.37
9	91.86	97.07	95.44	97.07	95.44	97.07	95.77	97.39
10	52.12	74.59	52.77	74.59	56.35	74.92	56.68	74.59
Average	89.28	93.65	90.42	93.62	90.55	93.42	90.65	93.52

In each experiment, the highest accuracy is always produced by NB in the 2nd iteration. The highest accuracy is 99.67%. The lowest accuracy is always produced by NB in the last iteration. The lowest accuracy is 52.12%.

The average of accuracy on each experiment is available in Figure 3. The average of highest accuracy, 93.65%, was generated by SVM in the first experiment. Meanwhile, the average of lowest accuracy, 89.28%, was generated by NB in the first experiment.

V. DISCUSSIONS

Sales reviews are often written in unstructured texts. The sentiment analysis can be used to know the polarity of the unstructured texts. Sales reviews can be taken from various sources such as social media. In research with similar domains however with different dataset and method conducted by Fiarni et.al [3]. A dataset of online store reviews gathered from social media. The method used Naive Bayes Classifier. The accuracy of the research is 89.21%. But in this research, the data are taken from a marketplace site. It will illustrate the concurring sales better. A better sales illustration will give a more evident picture of sales.

The result is in correspondence with the statements of researchers [2] and [14] which indicated that SVM and NB provide a good result in text classification. The comparison between SVM and NB shows that the accuracy average of SVM is higher than NB. The average reaches 93.65% using 25% features with highest TF-IDF. This suggests that important feature extraction may affect accuracy results. This result affirms the result of research [15] which shows that SVM with linear kernel provides higher accuracy compares to NB. NB shows that the increase in accuracy is in concordance with the increase in the number of features that are used in the analysis. Meanwhile, in SVM, the increase in the number of features is not necessarily showing the increase in accuracy.

VI. CONCLUSIONS

This paper intends to classify the texts using Support Vector Machine and Naive Bayes to figure out the sentiment of the sales review. A sentiment analysis in the marketplace sales reviews will provide a more evident picture of real sales. The analysis will be able to improve the precision of decision-making process on sales strategy in the future.

The result of the experiment shows that SVM generates a good result. The extraction of features is conducted using TF-IDF. The experiment is conducted using 25%, 50%, 75%, and 100% of the features with highest TF-IDF. The experiment shows that Support Vector Machine with linear kernel provides higher accuracy than Naive Bayes. The accuracy reaches 93.65%; using 25% features with highest TF-IDF. The approach of the research that has been conducted can be used as the base of intelligent sales evaluation in the online stores. In the next experiment, the review can be categorized to particular aspects such as the qualities of goods, services, and shipment. This will provide a more detail polarity result in every aspect.

REFERENCES

[1] Indonesia Internet Service Provider Association (APJII), "Penetrasi dan Perilaku Pengguna Internet Indonesia," 2016. [Online]. Available: https://apjii.or.id/content/read/39/264/Survei-Internet-APJII-2016. [Accessed: 07-Aug-2017].

[2] D. V. N. Devi, C. K. Kumar, and S. Prasad, "A Feature Based Approach for Sentiment Analysis by Using Support Vector Machine," in *6th International Conference on Advanced Computing*, 2016, pp. 3–8.

[3] C. Fiarni, H. Maharani, and R. Pratama, "Sentiment Analysis System for Indonesia Online Retail Shop Review Using Hierarchy Naive Bayes Technique," in *International Conference on Information and Communication Technologies (ICoICT)*, 2016, vol. 4.

[4] I. P. Windasari, F. N. Uzzi, and K. I. Satoto, "Sentiment Analysis on Twitter Posts : An analysis of Positive or Negative Opinion on GoJek," in *Proc. of 2017 4th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE)*, 2017, pp. 266–269.

[5] M. H. Saragih and A. S. Girsang, "Sentiment Analysis of Customer Engagement on Social Media in Transport Online," in *International Conference on Sustainable Information Engineering and Technology (SIET)*, 2017, pp. 24–29.

[6] D. A. Muthia, "Analisis Sentimen Pada Review Restoran dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," *J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 2, No.2, pp. 39–45, 2017.

[7] I. P. Windasari and D. Eridani, "Sentiment Analysis on Travel Destination in Indonesia," in *Proc. of 2017 4th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE)*, 2017, pp. 276–279.

[8] M. A. Sghaier and M. Zrigui, "Sentiment Analysis for Arabic e-commerce websites," in *2016 International Conference on Engineering & MIS (ICEMIS)*, 2016, pp. 1–7.

[9] V. R. Prasetyo, "Penggolongan Program Sinetron Berdasarkan Opini Masyarakat di Twitter Dengan Cosine Similarity," Thesis, Department of Computer Science and Electronics Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, 2016.

[10] A. Guo and T. Yang, "Research and improvement of feature words weight based on TFIDF algorithm," in *2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference*, 2016, pp. 415–419.

[11] F. Pedregosa *et.al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[12] B. Santosa, *Data Mining: Teknik Pemanfaatan Basis Data untuk Keperluan Bisnis*, 1st ed. Yogyakarta: Graha Ilmu, 2007.

[13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. San Francisco: Morgan Kaufmann Publishers Inc., 2011.

[14] H. Zhang, W. Gan, and B. Jiang, "Machine Learning and Lexicon based Methods for Sentiment Classification : A Survey," in *Web Information System and Application Conference Machine*, 2014, pp. 262–265.

[15] S. Rana and A. Singh, "Comparative Analysis of Sentiment Orientation Using SVM and Naïve Bayes Techniques," in *International Conference on Next Generation Computing Technologies*, 2016, pp. 106–111.