A Comprehensive Survey on Sentiment Analysis Using Machine Learning Techniques

Naeem Ahmed, Tariq Shah, Wakeel Ahmed, S.M. Adnan Shah

Department Of Computer Science, University Of Engineering And Technology Taxila Pakistan
naeem.ahmed@students.uettaxila.edu.pk , tariq.shah@students.uettaxila.edu.pk
wakeel.ahmed@uettaxila.edu.pk, syed.adnan@uettaxila.edu.pk

Abstract: Analysis of sentiment is a method of retrieving the information or emotions of peoples about any product or service. We can use data from social media platforms, blogs, product reviews websites and E-Commerce websites for using this data for analysis purposes. So, by this process we can get thinking or feelings of others that helps us in decision making. Various machine learning techniques can be applied on this data for getting useful insights that helps in decision making. This paper discusses and compares various machine learning and non-machine learning sentiment analysis techniques, applications, and challenges.

Keywords: Sentiment Analysis; Natural Language Processing; Machine Learning; Text Mining.

I. INTRODUCTION

Sentiment means the emotions or feelings of people about any product, service, or event. Many people called it as emotion artificial intelligence [1]. Before the implementation of internet most of the information was transmitted by the means of leaflets, television, radio, or word of mouth [1]. After the implementation of internet, the transmission of information is now done by taking advantage of social media platforms like Facebook, twitter E-mails etc. Because it removes the restriction of space and distance all over the world and we can communicate with anyone in seconds around the globe [1]. Today social media is playing a vital role in exchanging of information and is a perfect platform for communication all over the world. Most of the data on internet produced by people is come from social media platforms [2]. This data can be very useful for decision making in any aspect. But the data on social media platform is unlimited and unstructured due to which social media platform cannot optimize this data at much greater accuracy. Sentiment analysis is generally concerned with the mining of textual data like surveys, comments, reviews on the websites and social media platforms.

Sentiment analysis uses some computational techniques, like natural language processing techniques, text processing, and machine learning approaches for extracting and classifying people's feelings from the given data. The input data for sentiment can be in the form of online reviews, opinions of people, comments on any matter etc. Today a huge number of people relay upon online sentiment reviews because these sentiments help them in decision making about any product or service. There are three different kinds or components of sentiment analysis currently being used are given below

- Sentiment analysis on document level
- Sentiment analysis on sentence level
- Feature or aspect level sentiment analysis.

Document level sentiment classification is the level of classification in which a whole document is classified into any one category i.e. positive, negative, unbiased document that is also called neutral document. Sentence level is a type of sentiment analysis classification in which we classify every sentence of document into a category as positive, negative, or neutral sentence [3]. Feature or Aspect level classification of sentiment is a level of classification which is used to classify sentences of a document into positive, negative or neutral sentences on the basis of some characteristics or properties of the sentences used in the document [3]. This level is also known as prospective level sentiment Classification. Sentiment analysis become a prominent and attractive area for election result prediction and decision making for the politicians and party leaders. A progressive approach has been used since last decade to extract the information from the data available on social networks [4] for the accurate prediction of any election results. Behavioral analysis of data can be applied for the prediction accuracy of the sentiment analysis process. Sentiment analysis can be done by using both supervised and unsupervised machine learning techniques. The primary objective of this article is to survey current techniques and tools used for the analysis of sentiments and to evaluate results on the bases of comparison of various research articles.

II. SENTIMENT ANALYSIS

Sentiment classification or analysis is a method of analyzing and classifying the opinion of group of peoples or individual into any category from negative, positive, or neutral. For instance, the opinion of a group of users of a brand or company or customers of a company with respect to any product or service of the company [5]. Sentiment analysis tools analyze the opinions and emotions and then on the bases of some rules it gives ratings to the documents or opinions. Analysis of sentiments is a means of evaluating written words or spoken languages to assess whether the impact is positive, negative, or unbiased and to what extent [6]. The existing sentiment analysis methods being used in industry are as smart as being able to deal efficiently with a large amount of data.

Opinion inspection [7] is a broad and rich research area aimed at improving fully automated perception of the feelings and emotions of human beings that are communicated in any content with a higher implementation resulting in a more powerful use of the inferred data. There are various web sites and some other social networking sites which can be used to collect data for sentiment analysis processing. Twitter [8] has instigated especially far-reaching customer appropriation and it is also used for the rapid development in terms of correspondence/volume.

Twitter is a tiny blogging network that provides customers with 'tweets' that are linked to their users or another client [8]. These tweets can also contain feelings and emotions of people so we can use tweets about concerned product or service for emotion detection and processing. A research [9] indicates that Social media platform twitter has more than 313 million dynamic customers within a particular month in 2016, including 100 million customers worldwide daily.

III. PROCESS OF SENTIMENT ANALYSIS

There are a few steps that are used to analyze any text for sentiment detection. By using these steps, we can easily perform sentiment analysis on a piece of textual data [10]. The steps that are required for sentiment analysis are data collection or data gathering, text preprocessing, sentiment detection or sentiment extraction, sentiment classification and presentation of output [10] at the end of analysis. Figure 1 shows the process of sentiment analysis. Let us discuss steps of sentiment analysis one by one.

A. Data Collection

Data collection is the process of collecting data from different locations. People usually express their feelings or emotion about any matter on public forums like product review pages, blogs, comments on videos, or on Social media sites such as Facebook, Twitter, or Instagram etc. So, we can grab data from these platforms, but the data gathered from these platforms is very huge and it is in unstructured

format [10]. So, for dealing with such data text preprocessing in needed.

B. Text Preparation

Text preparation is also called text preprocessing. It is a method of cleaning and filtering data before analysis. In this step non-textual and irrelevant contents are removed from the data. This step enhances the efficiency of the analysis [11] because it removes unnecessary data due to which analysis take less time.

C. Sentiment Detection

The detection of sentiment is a method of subjectivity analysis of the sentences used in the specified content. Those sentences that give subjective expressions are collected while the sentences with objective expressions are removed or discarded [11]. This work is done on different scales with the help of several computational techniques like Unigrams, lemmas, negations etc.

D. Sentiment Classification

Sentiment classification is a vital step in which sentiments are divided or classified into different groups or classes [12]. Emotions or human opinions can be divided into three categories [12] i.e. positive, negative, and unbiased or neutral. During this step, the sentiment of sentiment analysis each sentence is detected, observed, and classified into any of the groups. At the end we get classified form of sentences into three different groups.

E. Presentation of Output

When sentences are successfully classified into categories then it is necessary that the results are in such format that anyone can easily understand it [3, 12]. In fact, the primary purpose of the sentiment analysis is to transform or convert unstructured textual data into useful information which can be used by any organization or individual for decision making or policy making [13]. So, the results will be shown on graphs such as bar chart, pie chart, and line graphs after the analysis is completed. Fig. 1 illustrate the process of sentiment analysis.

Figure 1: Process of sentiment analysis
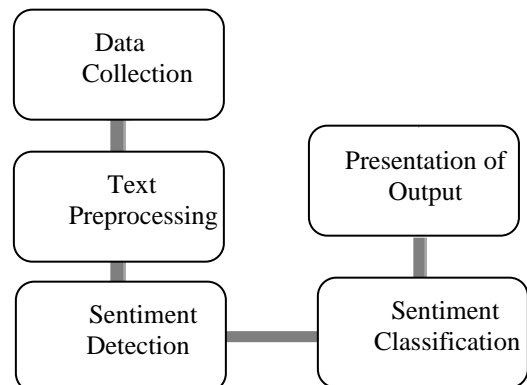
IV. LEVELS OF SENTIMENT ANALYSIS

Since the last decade, a handsome quantity of work has been done in the field of sentiment detection and analysis. The early efforts [14] in this field tended to apply the semantic orientation of adjectives as indicators of the overall orientation of textual content sentiment, and to classify objective and subjective sentences using a data set that was labeled with subjective adjectives [14]. There are different levels of analysis of sentiments, which are used according to the need and complexity of the problem.

- Document level sentiment [15] analysis is a process in which the goal is to determine whether a document contain positive emotions or statements or negative statements. In this level a whole document is declared as positive or negative document. This type of sentiment analysis is regularly known as document level sentiment classification.
- The Sentence level [15] of sentiment analysis, is the level in which the fundamental task is to determine that sentences of given document have positive, negative, or neutral sentiment. Neutral means no opinion about any sentence.
- Both the document level and the sentence level analysis [16] cannot find that what exactly people like and what they do not like. Aspect level sentiment analysis is used for this purpose as it is concerned with every fact in the sentence. It directly looks at the fact or opinion itself and extract results. Aspect level is depending on the possibility that opinion contains a sentiment is positive, negative, or neutral or an objective statement.

V. TECHNIQUES USED IN SENTIMENT ANALYSIS

Many techniques are used for analyzing sentiments. In this section we study various techniques used for analyzing sentiments. Machine learning and lexicon -based techniques [17] are commonly used for the sentiment analysis and classification of sentiments. We can use supervised machine learning as well as unsupervised learning techniques for sentiment analysis [18]. Different supervised and unsupervised learning algorithms like Naïve Bayes, Random forest, support vector machine [18] etc. are available that can be used for sentiment analysis on different kinds of data. Several methods are used for assessing sentiments in any document or given text. In this section we research various techniques used to evaluate sentiments. The figure below illustrates various machine learning and lexicon-based techniques that can be used for an interpretation of sentiments

Machine learning approach used various learning algorithms for the classification of sentiments from the desired data. These algorithms use some linguistic features [19] and different datasets to accurately classify any document, sentence, or feature into positive negative or

neutral sentiment result. Machine learning is dived into sub types i.e. supervised learning, unsupervised learning, reinforcement learning etc. These types have their own models and techniques that are used for various purposes. Fig. 2 shows various types and models of machine learning used for sentiment analysis. The Machine learning techniques (supervised and unsupervised) that are used for sentiment analysis and classification are described below:
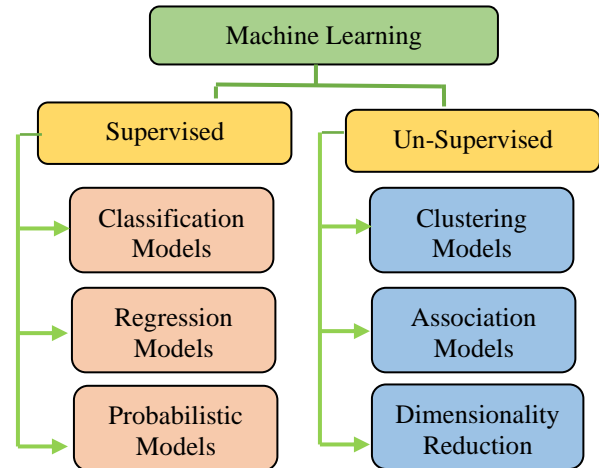

Figure 2: Techniques used for sentiment analysis

A. Supervised Learning

A labeled data set is used in supervised machine learning to train the machine learning model, then testing data set is used to test the model that determines that the model is specifically and reasonably trained or not [21]. After successful testing, the model can be used on unknown data for prediction. There are many supervised machine learning models some important and commonly used models are discussed below. Decision Tree Algorithm [22] is a simple, widely employed classification technique that is used to solve the classification and regression problems. Decision Tree Classification system poses a set of carefully crafted questions concerning the characteristics of test data. In every iteration a reply is received, a follow-up question is asked until a decision is made on regarding to the class of the data or record. Some decision tree algorithms [22] are:

- Random forest
- Classification and Regression Tree (CART)
- C4.5 and C5.0
- Chi-square
- Gradient boosting machines (GBM)

There are various linear classification and regression machine learning algorithms. Among all of them the SVM and artificial neural networks are models or algorithms that

are widely used now a day. Support vector machine (SVM) [23] is a supervised machine learning model that work on the idea of planes that specify some boundaries for decision making. It classifies different objects into different planes to show the dissimilarities between them. A simple example of support vector machine classification is shown in Fig. 3.
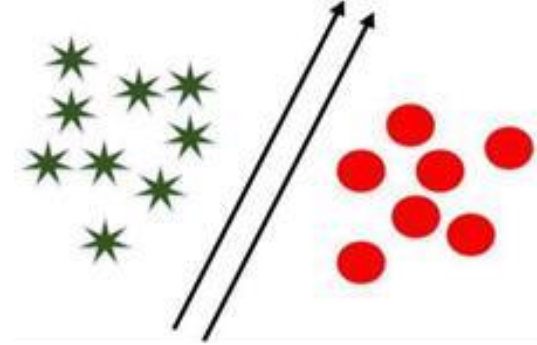

Figure 3: Support vector machine classification

The above Fig. 3 shows the classification concept of linear support vector machines. In the Figure there are circles that are called objects and these objects can belong to any of two classes i.e. the Green class or the Red class. An Isolated line determine the choice of objects between green/and red objects. On the correct hand side of the boundary, all objects are Green and to the side hand side of boundary, all objects are Red that shows that two colors are perfectly classified. If new object black circle is given to the model it will be classified as Green if it falls to the correct hand side of the boundary or classified as Red if it falls to the other side of the boundary. There are also probabilistic models [24] are statistical models that employ statistical techniques. These models are based on a strategy that defines the relationship between two variables. This relationship can be derived for sure as this involves using a random background process.

The first probabilistic model we will be addressing is classifier Naïve Bayes. This technique is based on the Bayes theorem [24] and assumes the predictors to be independent, which means knowing the value of one attribute put impact on the value of any other attribute. Naïve Bayes classifiers are easy to build, because they do not require any iterative process and, they perform very efficiently on large datasets with good level of accuracy. Despite its simplicity, Naïve Bayes is known to have often outperformed other classification methods. Naïve Bayes depends upon a theorem that is called Bayes theorem that is given below. In Bayes theorem it is all about discovering a likelihood while we know certain other probabilities. To this theorem, the formula is:

$$P(X|Y) = P(X)P(X|Y) \div P(Y) \qquad (1)$$

The (1) says that how often "X" occurs considering that "Y" occurs "P (X) P (Y|X)" if we know how many times "Y" occurs considering "X" occurs "P (Y|X) .And up to

what extend it is possible that X is to be its own P(X). And how "Y" to be its own P(Y)". For example, "Let us assume P (F) implies so often that fire happens, and P (S) implies how often smoke is seen, then by the theorem:

- P (F | S) means how often there is fire when we see smoke

- P (S| F) means how often we can see smoke when there is fire

Thus, by equation we can find that if hazardous fires are rare (1%), but smoke is relatively common (10%) because of barbecues, and 90% of hazardous fires create smoke then:

$$P (F| S) = P (F) P (S| F) \div P (S) \qquad (1)$$

Now by putting values in the above equation we get

$$= 1\% \times 90\% \div 10\% = 90\% \div 10\% = 9\%$$

Thus, the likelihood of hazardous fire when smoke will be around 9 percent.

B. Unsupervised Techniques

In unsupervised technique sentiment classification process is done by comparison. In this method of machine learning, the components of a given content are compared against words or lexicons whose sentiment values are chosen prior to their use. The main ides of this technique are this that we do not need to give any labels and definitions of training data. It is very useful technique when we do not have labeled data. This technique is useful because labeled data is not always present. Most of the times we deal with the unlabeled data. The Hierarchical clustering [25] and partial clustering are mostly utilized algorithms of unsupervised technique.

The first unsupervised clustering algorithm that we are about to explore is Hierarchical clustering. The hierarchical clustering involves the concept of clusters with a hierarchy, either by iteratively combining smaller clusters into a larger cluster or by splitting a larger cluster into smaller clusters [25]. This hierarchy of clusters that are produced by a clustering algorithm is called a dendrogram. A dendrogram is one of the ways in which the hierarchical clusters can be represented, due to this the user can easily analyze different clusters based on the level at which the dendrogram [25] is defined. It uses a similarity scale that represents the distance between the clusters that were grouped from the larger cluster. A simple structure of dendrogram is shown in Fig.4.
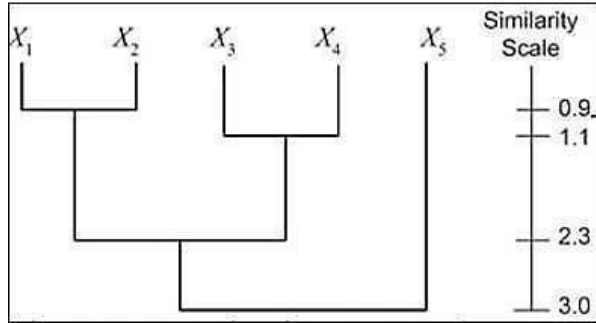
Figure 4: Classification of dendrogram [14]

Fig. 4 shows the physical illustration of dendrogram that classify the five objects based on similarities. On the right-hand side, a similarity scale is given which shows the similarity level of the objects. Partial clustering [25] is also an unsupervised machine learning technique that can be used for several classification problems. Objects are partitioned within partial clustering algorithms. Based on dissimilarity objects can change the clusters. K-mean Clustering algorithm [25] is mostly used partial clustering algorithm for sentiment analysis.

C. Lexicon Approches

Lexicon-based approach [26] relies on sentimental lexicons. Lexicon is an important indicator for any sentiment that is also called opinion word. This approach is divided into two categories that are:

- Corpus based approach
- Dictionary based approach

The first lexicon approach that is commonly used is dictionary-based approach [26]. In this approach sentiment words are gathered manually by using known instructions. The output set is called conclusion set that is created by the help of a corpora or dataset called WordNet [26] for their proportionate word and antonyms. In the next step more words are added into a special list that is called a Seed list and this process is continued and at the end the process will end when there is no new word for the iteration are found. This approach has a disadvantage that is the inability to discover feeling words with space and setting specific introductions. Corpus based approach [26,27] can also be used for sentiment analysis. This approach is used for dealing with the problem of

Table 1: Summary Of Some Existing Sentiment Analysis Techniques Used Of Machine Learning

Sr #	Author	Algorithm Used	Dataset	Accuracy %
1	Sefer Kurnaz et al. [1]	Deep Sparse Auto Encoder, SVM	Twitter Dataset	91.00%
2	S. Jagdale et al. [2]	Naive Bayes, SVM	Electronic Device on Amazon	98.17%
3	Chetana Heged et al. [3]	Naive Bayes, KNN Decision Tree, SVM	Custom dataset	94.66%
4	Hooi Mei et al. [4]	Naïve Bayes, NN, MLP, RNN	News headlines	96.695
5	Olivier Habimana et al. [5]	Deep Learning (DNN)	Amazon reviews	86.01%
6	Sunday Olusanya [7]	Naive Bayes, J48, BF tree	Standard database dataset	94.06%
7	Melvin Diale et al. [8]	SVM and ELM	Enron dataset	93.045
8	Songbo Tan et al. [9]	Hyper parameter Optimization, SVM	Spam Assassin	89.82%
9	Wadi Hijawi et al. [10]	KNN, SVM, Naïve Baye	Micro F1	89.70%
10	Hossam Faris et al. [11]	Naive Bayes, Random Forest Decision Tree	Publically available data	96.80%
11	Niketan et al. [12]	Random Weight Networks	Movies Reviews	95.10%

predicting feeling words with setting specific presentation.

Corpus based approach may also be used to analyze feelings. This technique is used to deal with the problem of seeking terms of meaning through setting specific appearance. Corpus based approach is a method in which the corpus serves as an empirical base from which

VI. APPLICATIONS OF SENTIMENT ANALYSIS

Sentiment analysis plays vital role in brand monitoring now a day we all know that reputation is very necessary for the success of any brand [28]. When most users check social media reviews of a brand or product and as review sites before purchasing it. Negative reviews put people off and how you handle can define your future as a business. Sentiment analysis is also used for customer support. For customer support social media is a platform where you can contact your customers. And whenever they are unhappy or unsatisfied with your products, they will call you on social media platforms. Whenever these mentions will appear in your dashboard with a flashing red color, and you better start engaging them as soon as they are there [28]. People nowadays expect brands to respond on social media almost immediately, and if you are not quick enough, you might as well see them moving on to your competitor's instead of waiting for your reply." Applications for sentiment analysis have spread to every imaginable space, items, administrations, workforce development, and budgetary authorities in order to reach gatherings and political decisions Sentimental analysis is also used in the forecasting of product sales, online video popularity checks etc.

VII. CHALLENGES OF SENTIMENT ANALYSIS

There are many challenges that influence the result of sentiment analysis process. Let us discuss some common challenges that are being faced by existing sentiment analysis techniques.

A. Language Issue

In sentiment extraction process English language is commonly used because its resources, API's, Datasets, dictionaries, and lexicons are easily available [29]. But for analysis of sentiment from other languages like Urdu, French, Hindi, and Arabic is very difficult because of in availability of resources for these languages.

B. Fake Reviews

Bogus or fake reviews also effect the result of sentiment analysis process. Fake opinions are used to misguide users by giving untruthful opinion about any matter. The identification of these fake reviews or opinion is a huge problem for sentiment analysis.

lexicographers extract their data and detect linguistic phenomena without prior assumptions and expectations [27]. This approach uses factual or semantic techniques to discover assumptions extremity.

C. Domain Dependency

Domain dependency also influence the result of sentiment analysis. This is a massive challenge for sentiment analysis in which the polarity of a same word changes from one domain to another domain [29]. For example, we have two statements "The story was unpredictable." And second statement is "The steering of car is unpredictable." Now for the same word "unpredictable" in the first statement the sentiment is positive while for the second statement the sentiment is negative."

D. Contradiction or Negation

Contradiction or negation identification [30] is also a challenging task for sentiment analysis. In Invalidations are communicated from different points of views without the reasonable utilization of any negative sentiment or feature in any sentence. The negation usually changes the polarity of the opinion in any sentence. For instance, it is not comfortable. The Table 2 shows [19] the accuracy, recall and precision of different machine learning classifiers when used for sentiment analysis on one thousand selected features.

Table 2 present the overall performance of various machine learning models of 100 selected features. Among all machine learning models supervised technique support vector machine outperforms in many research articles. All other techniques have their own uses. These models show different results on different datasets. In some conditions/ dataset support vector machine got low accuracy. So, we can say it depend on the type of data and how we are using the model.

Table 2: Performance Of Various Machine Learning Models On 100 Selected Features

Algorithm	Recall %	Precision %	Accuracy %
Naïve Bayes	98.46	99.66	99.46
SVM	95.00	93.12	96.90
KNN	97.14	87.00	96.20
NN	96.92	96.02	96.83
AIS	93.68	97.75	96.23
DT	94.36	91.35	93.55

VIII. CONCLUSION

Analysis of sentiments helps us to identify peoples' emotions or attitudes toward any matter. The feelings or emotions of peoples can be expressed in two ways. Positive or negative. This paper discusses different ways and techniques of machine learning i.e. supervised and unsupervised, and lexicon-based approaches for finding and dealing with the extraction and analysis of sentiments of peoples on different dataset. The survey paper also discusses the applications and challenges that are faced during the sentiment analysis process. Fake reviews language issues and domain dependency are the most important challenges faced by the existing sentiment detection models. Sentiment analysis techniques are being used in many fields like election result prediction, pricing of products, and feeling of peoples on any service etc. After the study and analysis of various machine learning techniques and models we came to conclusion that support vector machine (SVM) and Naïve Bayes are the most efficient algorithms that can be used for the sentiment analysis.

REFERENCES

[1] A, S. Sentiment Analysis Based on Financial Tweets and Market Information. in 2018 International Conference on Audio, Language, and Image Processing (ICALIP). 2018. IEEE Market Information. in 2018 International Conference on Audio, Language, and Image Processing (ICALIP). 2018. IEEE.

[2] Pai, P.-F. and C.-H. Liu, *Predicting vehicle sales by sentiment analysis of Twitter data and stock market values.* IEEE Access, 2018. **6**: p. 57655-57662. 3. Chen, Y. and Z. Zhang. *Research on text sentiment analysis based on CNNs and SVM. in 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA).* 2018. IEEEI. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] Jabbar, J., et al. *Real-time Sentiment Analysis on ECommerce Application. in 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC).* 2019. IEEE.

[4] Chang, W.-L., L.-M. Chen, and A. Verkholantsev, *Revisiting Online Video Popularity: A Sentimental Analysis.* Cybernetics and Systems, 2019. **50**(6): p. 563-577.

[5] Agarwal, A., et al. Sentiment analysis of twitter data. in Proceedings of the Workshop on Language in Social Media (LSM 2011). 2011. 7. Alsaeedi, A. and M.Z. Khan, A Study on Sentiment Analysis Techniques of Twitter Data.

[6] Jebaseeli, A.N. and E. Kirubakaran, A *survey on sentiment analysis of (product) reviews.* International Journal of Computer Applications, 2012. **47**(11).

[7] Yadav, S.K., *Sentiment analysis and classification: a survey.* International Journal of Advance Research in Computer Science and Management Studies, 2015. **3**(3): p. 113-121.

[8] Taboada, M., et al., *Lexicon-based methods for sentiment analysis.* Computational linguistics, 2011. **37**(2): p. 267-307.

[9] Ahmad, M., et al., *Machine learning techniques for sentiment analysis: A review.* Int. J. Multidiscip. Sci. Eng, 2017. **8**(3): p. 27-32.

[10] Gollapudi, S., *Practical machine learning.* 2016: Packt Publishing Ltd.

[11] Faris, H., et al. *Improving email spam detection using content-based feature engineering approach. in 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT).* 2017. IEEE.

[12] Mehta, R., et al., *Sentiment analysis and influence tracking using twitter.* International Journal of Advanced Research in Computer Science and Electronics Engineering, 2012. **1**(2): p. 72-79.

[13] Gitari, N.D., et al., *A lexicon-based approach for hate speech detection.* International Journal of Multimedia and Ubiquitous Engineering, 2015. **10**(4): p. 215-230.

[14] Pandey, M., et al., *Sentiment Analysis using Lexicon based Approach.* IITM Journal of Management and IT, 2019. **10**(1): p. 68-76.

[15] Alessia, D., et al., *Approaches, tools, and applications for sentiment analysis implementation.* International Journal of Computer Applications, 2015. **125**(3).

[16] Ravi, K. and V. Ravi, *A survey on opinion mining and sentiment analysis tasks, approaches and applications.* Knowledge-Based Systems, 2015.

[17] Hussein, D.M.E.-D.M., *A survey on sentiment analysis challenges.* Journal of King Saud University-Engineering Sciences, 2018. **30**(4): p. 330-338.

[18] Ebrahimi, M., A.H. Yazdavar, and A. Sheth, *Challenges of sentiment analysis for dynamic events.* IEEE Intelligent Systems, 2017. **32**(5): p. 7075.

[19] Kurnaz, A.P.D.S. and M.A. Mahmood, *Sentiment Analysis in Data of Twitter using Machine Learning Algorithms.* 2019.

[20] Jagdale, R.S., V.S. Shirsat, and S.N. Deshmukh, *Sentiment analysis on product reviews using machine learning techniques*, in *Cognitive Informatics and Soft Computing.* 2019, Springer. p. 639-647.

[21] Lim, S.L.O., et al., *Examining Machine Learning Techniques in Business News Headline Sentiment Analysis*, in *Computational Science and Technology.* 2020, Springer. p. 363-372.

[22] Singh, J., G. Singh, and R. Singh, *Optimization of sentiment analysis using machine learning classifiers.* Human-centric Computing and Information Sciences, 2017. **7**(1): p. 32.

[23] Tan, S. and J. Zhang, *An empirical study of sentiment analysis for Chinese documents.* Expert Systems with applications, 2008. **34**(4): p. 2622-2629.

[24] Tripathy, A., A. Agrawal, and S.K. Rath, *Classification of sentiment reviews using n-gram machine learning approach.* Expert Systems with Applications, 2016. **57**: p. 117-126.

[25] Amolik, A., et al., *Twitter sentiment analysis of movie reviews using machine learning techniques.* international Journal of Engineering and Technology, 2016. **7**(6): p. 1-7.

[26] Mehta, R., et al., *Sentiment analysis and influence tracking using twitter.* International Journal of Advanced Research in Computer Science and Electronics Engineering, 2012. **1**(2): p. 72-79.

[27] Diale, M., T. Celik, and C. Van Der Walt, *Unsupervised feature learning for spam email filtering.* Computers & Electrical Engineering, 2019. **74**: p. 89-104.

[28] Habimana, O., et al., Sentiment analysis using deep learning approaches: an overview. Science China Information Sciences, 2020. 63(1): p. 1-36.