"Improved Tweet Sentiment Classification Using semantic Base Features with an ensemble learning Approach"

Swati Bhola[1], Dr Rahul Malhotra[2], Mr.Gatuam Arora[3]

Computer Science Engineering Swami Devi Dyal Institute of technology Affilated to
Kurukshetra University,Kurukshetra

Abstract— **The available data on social media has contributed to vast research using sentiment analysis. The twitter-based social media represents a gold-mine approach for analyzing the performance of the brand. Large opinions of the people are found over Twitter that are honest, informative, and casual as compared to the formal type of data-survey analysis using magazines or reports. Millions of people share and express their sentiments over the media discussing about the brands whom they interact with. Sentiment analysis has turned out one of the most significant tools in natural language processing because it opens up numerous possibilities to understand people's sentiments on different topics. The purpose of an aspect-based sentiment analysis is to understand this further and find out what someone is talking about, and whether he likes it or does not like it. There have been various ways to deal with handle this issue, utilizing machine learning. In this thesis, labelled data is used based on polarity and Tweets pre-process and extract unigram features after pre-processing of the tweets. In pre-processing, the noisy data is removed by using tokenization; stop word removal and stemming these processes remove the duplicate data like to repeat words, hash tags and emoji's. These features and label learn by SVM, ANN and a hybrid of KNN and ANN (hybrid). In proposed experiment Ensemble approach shows improvement in precision and accuracy than other.**

Keywords-component; Sentiment Analysis,Semantic Base feature, social media, natural language processing,twitter

I. INTRODUCTION

The huge influence of social media around the world contributed to sentiment analysis being discovered. Huge data generation has been involved in recent developments in smart technology using mobile communication. The social media helps you to exchange ideas, thoughts and emotions. The word sentiment analysis (SA) is generally known as opinion mining, which is typically transmitted in a text that is positive , negative or neutral.

Approximately 500 million daily tweets are produced, representing a huge / extensive data collection for analysing the brand performance of manual members or company teams[1, 4]. The variety of tweets will possibly not be captured with constant or set laws. Due to its nuanced actions in contrast to a well-formatted documentary it is very hard to quantify tweet feeling analysis. The tweets are not based on any formal language style or on any formal terms. In theory, symbols and points are used to communicate thoughts, including emoticons, smileys and so on. The thesis discussed therefore the controlled learning approaches and methods of natural tweets, including feeling-based querys, for the understanding of the concept of tweets.

A. Problem Formulation

Using a large set of datasets has shown an improved accuracy and better outcomes. Naïve Bayes performed satisfactorily but It did not exceed the expectations. Support VectorMachines provided better accuracy, and Logistic Regression as well as the Support Vector Machines took very less time as compared to Naive Bayes.

The results obtained presents that K-NN classifier outperforms well for analyzing sentiments.The classifier along with bag of words-based feature selection outperformed well dependent over polarity-based sentiment classification.Username vs. Influential and Line graphs User Tweet Flow were used to provide recommendations, hence minimum and maximum tweet out-reach was calculated. It also analyzed how beneficial a user would be in tweet spread.

A sentiment prediction model was developed using a Swedish

model, elastic net method, for computing the polarity of Swedish tweets sentiment The experiment analysis has shown that the proposed model out performed well compared to existing approaches.Based on distinct strategies, one can have the consumer interest, customer opinions related to specific product. The major challenge was to pin-point the polarity- negative, neutral, and positive reviews of customer according to products.

The results have shown that Deep Learning model can be used efficiently and effectively for financial sentiment analysis. The model of convolutional neural is considered to be the best model for prediction of author's sentiment.

The research novelty represents the combination of temporal and spatial aspect of twitter-based data while implementing the emotion and sentiment for the park visits in the overall city.

B. Objective

1. To study different approaches and methods of tweet sentiment analysis.
2. To proposed semantic base feature selection with ensemble learning base classification.
3. To compare and analysis of proposed approach with existing approach.

C. Need Of Study

A social network and sentiment analysis framework has been developed that can be run on Twitter data to study work. Twitter is a popular social networking and microblogging site that counts hundreds of millions of active users and regular posts. As a social networking site, Twitter is designed as a guided graph that allows individual users to monitor a number of other users (followers) and other users (followers) to do the same. Therefore the "follow" relation is asymmetrical, needs no compulsory acknowledgment and is used primarily to accept all public messages from any user. Twitter is used as a microblog service for the posting of short messages of up to 140 characters (tweets) which may contain views, feelings, facts, photos and other media references. In addition, the @ symbol allows to implement mentions, i.e. references to other users, and hashtags, i.e. references to topic themes can be added by means of the # symbol. As a result, we gathered three types of data through our research. The User type is user profiles; we get user id from Twitter, name , location, number followers, number tweets. The tweet form refers to posted messages; we obtain tweet id, user id, message and date from Twitter. The following are given. Finally, the sort of friend is the "follow" of users. We have added a field to both tweets and users, in addition to the data obtained directly from Twitter in order to connect a sentiment with them according to our research.

II METHODOLOGY

This chapter describes the proposed methodology of the work with the algorithm used in it. The data is collected for experiment from Tweeter and stored in database for preprocessing. In the below given section we describe the flowchart and the techniques used in this work in detail.

METHODOLOGY STEPS

Data Collection: data given in input is collected from the Tweeter in the form of tweets which are related to the social welfare scheme started by govt. the tweets given in input are showing the opinion of peoples related to the schemes.

Storing and Fetching the data: In this we store the data in the database to make the graph which shows the progress in the work.

Data Preprocessing: – It is a data cleaning process which noisy data like repeated words, hash tags and stop words are removed and data in uppercase is converted into lower case

Applying various Learning techniques:The Machine learningis the mechanism that makes mining models. To form a model, associate degree formula 1st analyzes a group of information for specific patterns and trends.

Result optimization: the goal of optimization is to determine the optimal input values to obtain a desired output.

III LITERATURE REVIEW

Twitter

Twitter for data examination, where Twitter is a web arranging organization that enables customers to send and peruse short 140-character messages called —tweets. Not with standing its introduction, Twitter is open for unregistered customers to examine and screen most tweets, not in any manner like Facebook where customers can control the assurance of their profiles. Twitter is similarly a generous in length run relational correspondence microblogging site page. The enormous information gave by twitter, for instance, tweet messages, customer profile information and the quantity of fans/followings in the framework expect a gigantic part in data examination, which therefore makes most examinations investigate and take a gander at various examination methods to understand the progressing used developments.

Retweet in twitter is the declaration movement to a specific tweet, as occasionally the customer passes information to his/her social events of individuals to express their inclination on a particular tweet. The instrument of retweetability accept an undeniable part in information scattering. The retweet rate of the primary tweets and the quantity of notification related to those tweets to look into whether the amount of retweet and number of notification are related to a comparable framework.

TWEETS SEMANTIC FEATURES

- Bag-of-words: How to break up long text into individual words.
- Filtering: Different approaches to remove uninformative words.
- Bag of *n*-grams: Retain some context by breaking long text into sequences of words.
- Log likelihood ratio test: Identify unique combination of words that are more likely to be used together than not.
- Parts of speech: Tag words with their *parts of speech* (i.e. noun, verb, adjective).

Social Media

In 1990s, the clients of the system were capable to make the content on their own in order to show the process by broadband internet representing a very popular/common technology. In early 1997, first website based on social network was appeared famously known as SixDegrees.com. From the year 2002 onwards, most number of informal type of organization-based locales are derived. For instance, most of them are similar to Friendster whereas the others are similar to the concept of MySpace. In 2000, there was a broad increase in the process of online-based networking and it also increased the client's quantity/number. A large number of elements were involved into the area of online-based networking process. These mainly embodies broadband connection accessibility, new devices of programming, and cell phones and personal computers development; most of the developed form of elements are

financial and social and quick adoption by youth generation, and the process of expansion of programming, personal computers, social media sites commercialization.

In 21st century, with enhancement of Internet (Web 2.0), platform of social media for instance, the websites helps to allow their user clients to work and connect together in practical groups. Openness, rich experience of client, adaptability, insight aggregation, etc. are added form of newly built attributes in the process of online-based networking. The user of the system (clients) usually get supplementary facilities of make post pictures, like, audio upload, video-based content on the environment of online-based networking. The posted data can be easily shared with most of the chosen users or clients or it navigate freely over the technology of web. The concept of social media has quickly expanded its roots all over the world due to merger of social links and innovations for worth co-making.

.

D. Twitter Sentiment Analysis

In the field of sentiment analysis, the techniques of SA represent a major challenging research topic. The main aim of such techniques is to mainly classify the positive or negative opinions expressed by the document. The classification of SA is divided mainly into two distinct methods/approaches. One is the machine learning approach and the other is the lexicon- based approach. The machine learning methodology uses different algorithms of machine-learning whereas the lexicon- based approach is divided into dictionary based-approach and corpus-based approach.

1. Lexicon-based Approach

This method represents another type of unsupervised approach, but in such cases, a dictionary with synonyms and antonyms of opinionated phrases and words could be used along with their respective orientation of sentiments. Here, two methods named dictionary-based and the corpus-based methods are used commonly in its automated form. In case of dictionary-Based approach, the main strategy involves the manual collection of small opinion words sets and then the set is grown further by finding large text collections like Word Net. The new searched words are added to opinion set and the cycle gets repeated till no more words are found further. The major disadvantage of such a method is that it completely relies over corpora and there is no collection of opinion words largely each time with its available domain. Note, here all the lexicon-based words do not express the negative or positive opinion regarding an entity.

The Corpus-Based approach is used in two of the following conditions. The first is the discovery of new sentiment words from a corpus domain by using the list given of known opinionated words and secondly, the creation of sentiment lexicon from another one. Such an approach is not as effective as compared to dictionary-based approach as it needs a corpus with all words in English. The corpus-based method is divided in semantic and static approach depending over the used techniques.
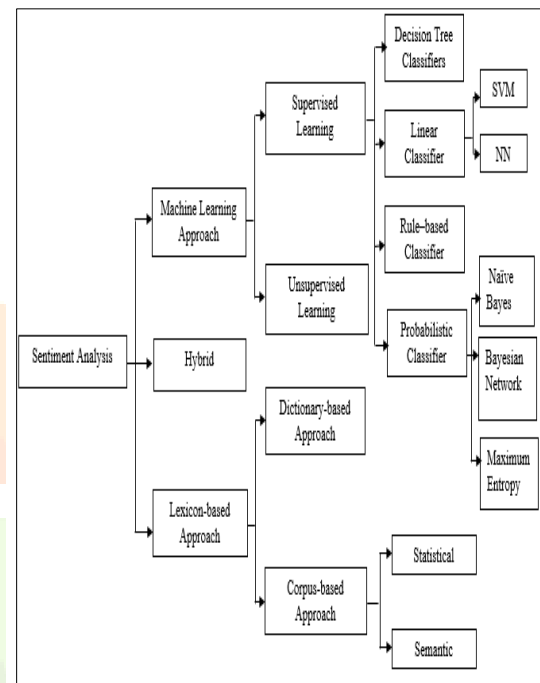
2. Machine-learning-based Approach

It relies over sentiment analysis treatment as a problem of text classification. Here, classification of text is basically used in business automated decision that need proper text processing. It uses records (trained set) in order to train a design or model which is used in prediction of fresh records without any kind of label. Each and every record gets assigned to a specified class [19]. If a new form of unlabelled record is known (given), then the model helps in predicting its class label. Such classes may be positive, neutral, and negative. In this kind of approach, two types of its sub-approaches are distinguished based on the methods of learning. One is the supervised learning and the other is un-supervised learning.

Fig.3.3

Sentiment Analysis Techniques

Techniques of Sentiment Analysis

So many techniques of machine learning are adopted to categorize several reviews.

1. Decision Tree: The approach of Decision Tree useful is mainly use for classifying issues. In case of option-based testing, the experts use split percentage as the favored method.

2. Naive Bayes: The NB classifier presents basic probabilistic form of classifier on the basis of applying the Bayes based theorem. Naïve Bayes evaluates probabilities set or group by consolidation-based values in a known dataset. Also, this classifier has a quick process of decision-based making.

3. K-nearest neighbours: Also known as K-NN representing an algorithm based on lazy learning and it also presents an approach of non-parametric nature for classifying objects on the basis of closest form of training. The algorithm of K-NN indicates a very easy type of algorithm

for the process of machine learning . The performance of algorithm based on K-NN classifier mainly depends over various distinct key factors, such as proper measure of distance, K parameter, and a measure of similarity .

*4. Support Vector Machine:*in machine learning, SVM presents a learning supervised model that helps in examining data and identifies several patterns. This technology is mainly used for analysis of classification and regression. Recently, a large number of classification-based algorithms have been planned and scheduled, yet SVM is considered as the most popular and widely used classifiers.

5. Artificial neural networks: Connectionist systems or ANNs represents the system of computing as encouraged by constituting the brain of animals . Here, neural network does not represent an algorithm on itself basis, instead it presents a design for various distinct algorithms of in order to perform combined analysis and process the complex form of data-based inputs. Such type of systems mainly learns to execute various tasks by using examples that are not programmed in general with any kind of task-particularized rules. For example, in case of identifying, that on physical or manual basis has been as "cat" or "no cat" and in addition by using numerous results, the identification of cats in other type of images can be done. They perform the analysis without any kind of prior information about cats, for instance, they have cat-like faces, tails, fur and whiskers. Instead of it, they generate automatically by identification of attributes from the material of learning the process they like.ANN is usually based on an assortment of associated nodes or units known as that loosely designs in biologically built brain. Each of its connection, such as the property of in it, helps in transmitting a signal from a single artificially built neuron to the other. Such type of neuron which receives a signal can perform the operating process and further the signal artificial neurons in addition are associated to it. Commonly, in case of implementing ANNs, signal at a links among several artificial neurons presents a, and each artificial neuron output is mainly evaluated by various non-linearized functions of the input sums. ANN approach main aim was to solve issues. However, the analysis based on attention, over time moved to perform particular tasks, resulting in discrepancy. ANN has been used excellently over large number of tasks involving computer-based vision, recognition of speech, machine-based translation, social networking filtering process, medical and video games.

6. Maximum Entropy: The maximum entropy classifiers, does not contain assumptions regarding the relation among features. These kind of classifiers always help in trying to maximize the system-based entropy by estimation of conditional class label distribution

The sub classes of supervised learning include Linear, Decision Tree, Probabilistic and Rule based classifier.

- *Linear classifiers:* They are simple in nature. The basic idea is to count the sentence-based negative and positive words and to compare the number of negative and positive words for the determination of sentence-based polarity [21]. These classifiers add weight to the words. In such a case, the "most positive" words represent highest weight and the "most negative" words represents the lowest weight. The most popularly used linear classifier is SVM Classifiers which aims to pin-point the linear separator with class-based best separation method.

Decision Tree: These methods are used for the purpose of prediction. They can be used easily for classification. Suppose, a record is known (given) with its unknown class label, such a record gets tested against decision tree, and the route gets traced from root to the node which when determines the prediction of class for such type of record [18]. Such methods are famous as their modelling does not need any expert domain or settings. The most widely used decision tree packages for the purpose of text classification implementation are C5 and ID3.

Probabilistic classifiers: Also known as generative classifiers as they generate or form a model for each of the class. They basically use mixture of models assuming that each of the class represents the component of a model. The most famous type of Probabilistic classifier is the Naïve Bayes Classifier. This classifier is of very simple form and it easily gets coded in each and every language of programming as these involve simple mathematical analysis. Such a model works with bag-of-words (unordered-set). The frequency of each word is kept generally instead of its location. This classifier uses Bayes Theorem in order to determine the label from where the known type of feature-set belongs.

*Rule Based Classifier: This classifier is similar to*decision tree classifier as both of the classifiers on its feature space encode the rules. The major difference is that in case of decision tree classifier, it uses hierarchical approach whereas the rule-based classifiers allows the mechanism of decision space overlapping. Various multiple research work suggests distinct ways to transform decision tree to a rule-based classifier. In case of rule-based classifier, the phase of training generates different criteria rule base. The most used popular forms are the confidence and the support.

F. Application Programming Interface(API)

API is third parties libraries which we use in our program to make our application work efficently.In this project I am using twitter API to collect the data from twitter. From the tweets analysis , the data provided as an input to the proposed model were collected. In order to extract useful information from twitter data set, several pre-processing steps are needed.

The tweets collected are saved to format .csv files, which are then fetched to python PyCharm. There are about 3000 tweets available to train the datasets and test them. For the training and testing of tweets fetched, the Data Mining Algorithm (ANN, KNN and Hybrid) is used.

IV. RESULT ANALYSIS
This section deals with the result and analysis of the results obtained.

Fig1. Accuracy on different classifiers

Fig.1 shows the accuracy of the ANN, KNN and Hybrid classifiers, The X-axis on the graph shows the validation fold and the Y-axis shows the accuracy values. In a five- and ten-fold validation process, the hybrid algorithm demonstrates maximum accuracy. In a 5-fold validation procedure, KNN demonstrates minimum precision 49.23.
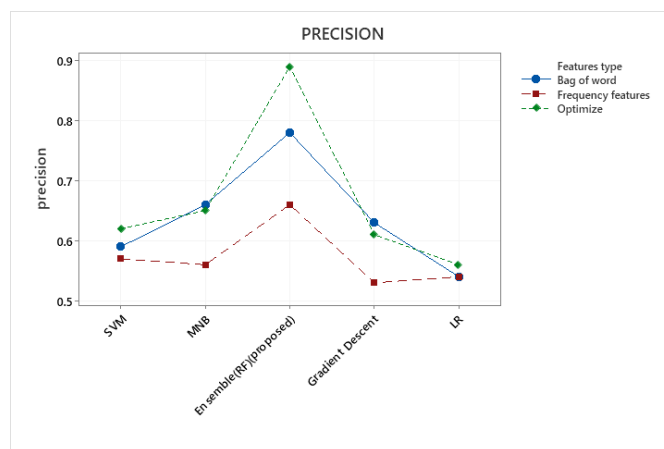

Fig.2 Precision on different classifiers

In Figure 2, it demonstrates the accuracy of the various ANN, KNN and Hybrid classifiers. The X-axis shows the validation fold and the Y-axis shows the precision values. In 5- and 10-fold testing, the hybrid algorithm shows the highest accuracy. In five-fold validation, KNN shows the minimum Precision 50.23.
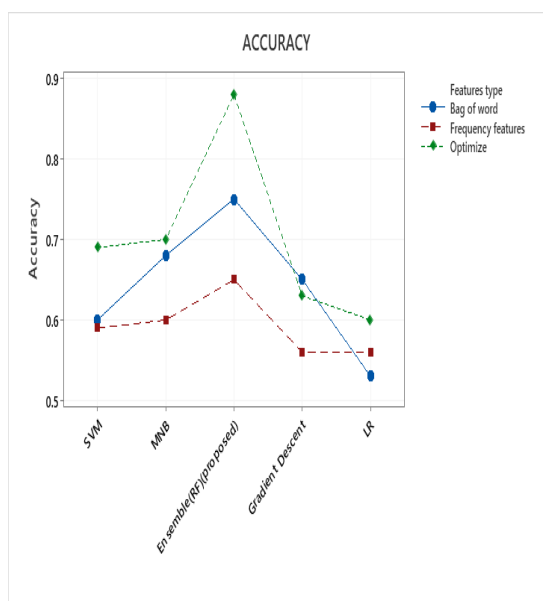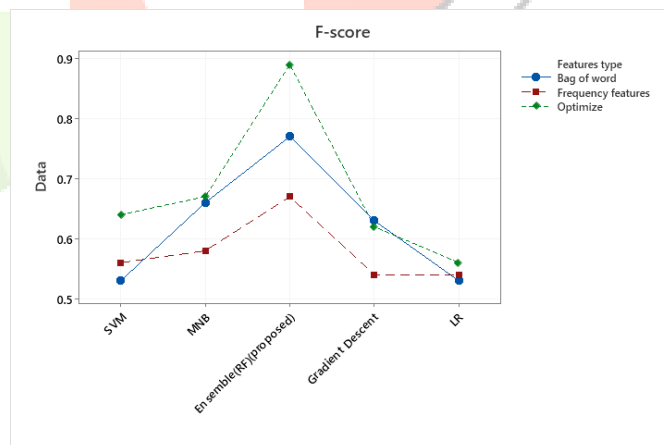

Fig.3 Comparison of different classifiers

It shows a contrast between ANN, KNN and Hybrid classifiers in fig.3. The X-axis displays the graphs and the Y-axis represents the accuracy, precision , recall and F-measurement values. The highest results on all parameters are seen in the hybrid algorithm, and KNN has the least result of all.

V CONCLUSION

For classification of algorithms with binary classification method, the Support Vector Machine principle (SVM) is used in this analysis. Such a method helps analyse various characteristic vectors of a given class in order to define the relationship dependence between a sensation and each characteristic. Each vector is regarded here as a data point in the dimension of the vector space which corresponds to the size of the feature set. The ANN helps to classify the hyperplanes based on the vector dimension that separate the class into two forms. One is considered to be "Highest," i.e. a good separation between hyperplane and training data type known as a functional margin. The hyperplane has the great distance from the training data type. Generally, the classificator error will be reduced if the margin is high. When a new type of tweet, i.e. unlisted, is added to the framework, we have found that our proposal to improve the output of hybrid ANN and KNN for the analysis of feeling based on aspect using unigrammatic features will help with the extraction of the same feature vector as the feature labels. The hybrid solution was also substantially better, in direct contrast. It can inspire additional data to use the results. In the thesis the study and many aspects of the sentiment analysis of the Indian welfare scheme is available. The extent of this study will contribute to a more general understanding of the analysis of sentiments. In the KNN experiment, precision and accuracy are effective than other methods.
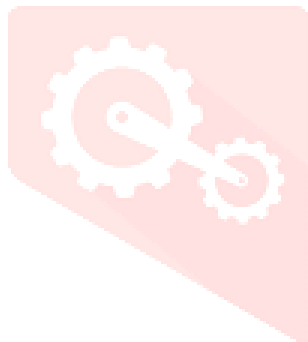
REFERENCES

1. [1] Gimpel, Kevin, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. "Part-of-speech tagging for twitter: Annotation, features, and experiments." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pp. 42-47. Association for Computational Linguistics, 2011.

2. [2] Agarwal, Apoorv, BoyiXie, Ilia Vovsha, Owen Rambow, and Rebecca Passonneau. "Sentiment analysis of twitter data." In *Proceedings of the workshop on languages in social media*, pp. 30-38. Association for Computational Linguistics, 2011.

3. [3] Kouloumpis, Efthymios, Theresa Wilson, and Johanna D. Moore. "Twitter sentiment analysis: The good the bad and the omg!" *Icwsm* 11, no. 538-541 (2011): 164.

4. [4] Gokulakrishnan, Balakrishnan, PavalanathanPriyanthan, ThiruchittampalamRagavan, Nadarajah Prasath, and AShehanPerera. "Opinion mining and sentiment analysis on a twitter data stream." In *Advances in ICT for emerging regions (ICTer), 2012 International Conference on*, pp. 182-188. IEEE, 2012.

5. [5] Mohammad, Saif M., Svetlana Kiritchenko, and Xiaodan Zhu. "NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets." *arXiv preprint arXiv: 1308.6242* (2013).

6. [6] Kontopoulos, Efstratios, Christos Berberidis, TheologosDergiades, and Nick Bassiliades. "Ontology-based sentiment analysis of twitter posts." *Expert systems with applications* 40, no. 10 (2013): 4065-4074.

7. [7] Da Silva, Nadia FF, Eduardo R. Hruschka, and Estevam R. Hruschka Jr. "Tweet sentiment analysis with classifier ensembles." *Decision Support Systems* 66 (2014): 170-179.

8. [8] Sahayak, Varsha, Vijaya Shete, and Apashabi Pathan. "Sentiment analysis on twitter data." International Journal of Innovative Research in Advanced Engineering (IJIRAE) 2, no. 1 (2015): 178-183.

9. [9] Ficamos, Pierre, and Yan Liu. "A topic based approach for sentiment analysis on Twitter data." *International Journal of Advanced Computer Science and Applications* 7, no. 12 (2016): 201-205.

10. [10] Onam Bharti, Monika Malhotra. "Sentiment Analysis on Twitter Data"International Journal of Computer Science and Mobile Computing, Vol.5, no.6, 2016, pp: 601-609

11. [11] Nakov, Preslav, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and VeselinStoyanov. "SemEval-2016 task 4: Sentiment analysis in Twitter." In *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)*, pp. 1-18. 2016.

12. [12] Bagheri, Hamid, and Md Johirul Islam. "Sentiment analysis of twitter data." *arXiv preprint arXiv: 1711.10377* (2017).

13. [13] Wehrmann, Joonatas, Willian Becker, Henry EL Cagnini, and Rodrigo C. Barros. "A character-based convolutional neural network for language-agnostic Twitter sentiment analysis." In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pp. 2384-2391. IEEE, 2017.

14. [14] Pabreja, Kavita. "GST sentiment analysis using twitter data." *IJAR* 3, no. 7 (2017): 660-662.

15. [15] Roy, Kaustav, Disha Kohli, RakeshkumarKathirvel Senthil Kumar, Rupaksh Sahgal, and Wen-Bin Yu. "Sentiment Analysis of Twitter Data for Demonetization in India–A Text Mining Approach." *Issues in Information Systems* 18, no. 4 (2017).

16. [16] Krebs, Florian, Bruno Lubascher, Tobias Moers, Pieter Schaap, and GerasimosSpanakis. "Social Emotion Mining Techniques for Facebook Posts Reaction Prediction." *arXiv preprint arXiv: 1712.03249* (2017).

17. [17] Farias, DI Hernández, and Paolo Rosso. "Irony, sarcasm, and sentiment analysis." In *Sentiment Analysis in Social Networks*, pp. 113-128. 2017.

18. [18] Nakov, Preslav. "Semantic Sentiment Analysis of Twitter Data." *arXiv preprint arXiv: 1710.01492* (2017).

19. [19] Shoeb, Md, and Jawed Ahmed. "Sentiment Analysis and Classification of Tweets Using Data Mining." *work* 4, no. 12 (2017).

20. [20] Krebs, Florian, Bruno Lubascher, Tobias Moers, Pieter Schaap, and GerasimosSpanakis. "Social Emotion Mining Techniques for Facebook Posts Reaction Prediction." *arXiv preprint arXiv: 1712.03249* (2017).

21. [21] Rosenthal, Sara, NouraFarra, and PreslavNakov. "SemEval-2017 task 4: Sentiment analysis in Twitter." In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pp. 502-518. 2017.

22. [22] Kovacs-Györi, Anna, Alina Ristea, Ronald Kolcsar, Bernd Resch, Alessandro Crivellari, and Thomas Blaschke. "Beyond Spatial Proximity—Classifying Parks and Their Visitors in London Based on Spatiotemporal and Sentiment Analysis of Twitter Data." *ISPRS International Journal of Geo-Information*7, no. 9 (2018): 378.

23. [23] Tyagi, Abhilasha, and Naresh Sharma. "Sentiments Analysis of Twitter Data using K-Nearest Neighbour Classifier." *International Journal of Engineering Science* 17258 (2018).

24. [24] Desai, Radhi. "Sentiment Analysis of Twitter Data: A Survey." International Journal of Scientific Research in Computer Science, Engineering and Information Technology IJSRCSEIT Vol. 3, no. 1, pp: 2456-3307 (2018).

25. [25] Dawar, Avnish, Archana Purwar, Nikhil Anand, and Chirag Singla. "TweetRush: A Tool for Analysis of Twitter Data." *I.J. Education and Management Engineering,* 2, (2018), no: 31-40

26. [26] Mensikova, Anastasija, and Chris A. Mattmann. "Ensemble Sentiment Analysis to Identify Human Trafficking in Web Data." GTA3 2018, February 2018, Marina Del Rey, CA USA.

27. [27] Brinda, H. S. Nagashree, and Madhura Prakash. "Sentiment analysis of Twitter data: A machine learning approach to analyse demonetization tweets." (2018).

28. [28] Kapase, Harshal, Kalyani Galande, Tanmay Sonna, Deepali Pawar, and DipmalaSalunke. "A review on: Sentiment polarity analysis on Twitter data from different Events." (2018).

29. [29] Barnes, Jeremy, Roman Klinger, and Sabine Schulte imWalde. "Projecting Embeddings for Domain Adaption: Joint Modeling of Sentiment Analysis in Diverse Domains." *arXiv preprint arXiv: 1806.04381* (2018).

30. [30] Cummins, Nicholas, Shahin Amiriparian, Sandra Ottl, Maurice Gerczuk, Maximilian Schmitt, and Björn Schuller. "Multimodal Bag-of-

Words for Cross Domains Sentiment Analysis." In *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Calgary, Canada*, pp. 1-5. 2018.

31. [31] Li, Yujiao, and Hasan Fleyeh. "Twitter Sentiment Analysis of New IKEA Stores Using Machine Learning." In *2018 International Conference on Computer and Applications (ICCA)*, pp. 4-11. IEEE, 2018.

32. [32] Sohangir, Sahar, Dingding Wang, Anna Pomeranets, and Taghi M. Khoshgoftaar. "Big Data: deep learning for financial sentiment analysis." *Journal of Big Data* 5, no. 1 (2018): 3.

33. [33] Jadon, Priyanshu, and Miss Rupali Dave. "A Big Data Approach for Sentiment Analysis of Twitter Data using Naive Bayes." (2018).