Mehran University Research Journal of Engineering and Technology

https://doi.org/10.22581/muet1982.2301.19

2023, 42(1) 207-215

Comparative study on sentimental analysis using machine learning techniques

Murali Krishna Enduri [a], Abdur Rashid Sangi [b,*], Satish Anamalamudi [a], R. Chandu Badrinath Manikanta [a], K. Yogeshvar Reddy [a], P. Lovely Yeswanth [a], S. Kiran Sai Reddy [a], Asish Karthikeya [a]

[a] *Department of Computer Science and Engineering, SRM University-AP,Amaravati, Guntur India*

[b] *Department of Computer Science, College of Science and Technology, Wenzhou-Kean University, Ouhai, Wenzhou Zhejiang China*

[*] Corresponding author: Abdur Rashid Sangi, Email: sangi_bahrian@yahoo.com

Received: 26 October 2022, Accepted: 15 December 2022, Published: 01 January 2023

K E Y W O R D S

Sentimental Analysis

Machine Learning

Textual Opinions

A B S T R A C T

With the advancement of the Internet and the world wide web (WWW), it is observed that there is an exponential growth of data and information across the internet. In addition, there is a huge growth in digital or textual data generation. This is because users post the reply comments in social media websites based on the experiences about an event or product. Furthermore, people are interested to know whether the majority of potential buyers will have a positive or negative experience on the event or the product. This kind of classification in general can be attained through Sentiment Analysis which inputs unstructured text comments about the product reviews, events, etc., from all the reviews or comments posted by users. This further classifies the data into different categories namely positive, negative or neutral opinions. Sentiment analysis can be performed by different machine learning models like CNN, Naive Bayes, Decision Tree, XgBoost, Logistic Regression etc. The proposed work is compared with the existing solutions in terms of different performance metrics and XgBoost outperforms out of all other methods.

1. Introduction

Access Sentiment analysis is study of the distinct existence of people such as opinion, emotions, thoughts and reviews of the people. Mainly it determines the quality or quantity of the article, news, service, product etc by using the negative, neutral, and positive comments of the user. In other words, sentiment analysis is analyze the user's comments and information and gives better knowledge and idea about the product or service to new user. It means, rating, comments and reviews of product by user's inputs a high impact while choosing any product online by other user. Different user have various opinions. So in one case, a person treated as negative and it will be taken as positive in another case. A lot of the mixed opinions can be misleading so that why sentiment analysis comes into picture. This helps them to determine their business model by better understanding the general reaction into their products and their business stand in the market. Furthermore, it allows to understand the consumer's needs and wants in the better possible way [6,15,25,19]. Reviews of movie plays an important role in gauge the performance and publicity of a movie. While rating the movie, the writer tells about the status, failure or success of a movie quantitatively to the end users. A textual movie review helps to the end users about the the merits and demerits

of the movie. Furthermore, the deeper analysis of a movie review can help the end users to conclude whether the general expectations of the audience can be attained by the reviewer. In this paper, authors also to propose an sentiment analysis on movie reviews which are given by reviewers and gives information of the overall output of the movie. In other words, the conclusion would be whether the movie is good or bad based on the story or other technical aspects. Further, authors goal to extract the connection or correlation of the movie keywords in the review process to identify or predict the overall status of the movie review[3,22,5]. In general, there are two different approaches that can be used in the sentimental analysis. First One is with the consideration of the symbolic methods whereas the second method is through machine learning approach. In the symbolic learning approach, it is grouped based on some of the learning strategies namely learning from discovery, analogy, and through learning from root. In the machine learning based approach, supervised learning, weakly supervised learning and unsupervised learning is mainly used. Further, with linguistic method, lexicon based and machine learning can be approached as one of the mainly used approach in sentiment categorized[6,20]. Sentiment analysis is defined as the process of determining the information from Social media reviews, i.e., the articles can be rated as Neutral, Negative or Positive. From the text mining technique, one can use Machine learning to analyze the text for the sentiment to detect whether it is Neutral, Negative or Positive opinions in customer feedback[10,16].

It is broadly divided into three types

Fine-Grained:

• It helps us to derive the polarity precision.

• It categories Neutral, Negative or Positive.

Aspect-Based:

• It helps us to determine the specific aspects where the users are talking about Emotion Detection.

• It helps us to detect the emotions.

• It can include panic, fear, frustration, happiness, sadness, and anger, etc.

Intent Analysis:

• It helps you to identify the intent of the customer.

• Ex: if a customer is willing to purchase, you target them with advertisements. If not, you can save your time and resources by not advertising them.

Applications:

1. Product Design and Improvement Phase: In this, companies try to clear about how the customers can understand their products and how customers feel\ about their products. It also captures the customers' complaints about a product and improve their products based on the customer's needs.
2. Sentiment Analysis in terms of Call center perspective: It helps to monitor and manage the call center agent's live phone interactions with the end users. It further helps to analyze and understand how customer satisfaction can be varied by product and call center services.
3. Monitoring the Brand reputation: It helps the companies to monitor the brand's reputations across different Social media broadcast. With this, companies usually get the valuable insight for their products, services, and brands through application of the sentiment analysis to the social media posts.
4. Sentiment analysis on Movie reviews: It helps the user to decide whether to watch the movie or not. It helps to make decision by not wasting their time for reading all the reviews. It gives overall rating for a movie how positive, negative, or neutral review on movie.

2. Related and Recent Work

The In the current scenario, the dissemination of the pretence is mainly altered into an extraordinarily very simple promise because of the media which is based on web. In order to minimize or stop this issue, the authors in this paper would like to find a news which is false or authentic. To attain this, this paper uses a machine learning algorithm to construct a process which it is identify that the news given is authentic or not with using the optimized ML and NLP methods.

2.1 Bag Of Words(BoW)

In general, Bag Of Words which usually called as BoW is most commonly utilized in the techniques for classifying the reports[24,29]. BoW is defined as a Natural Language Processing methodology and an Information retrieval technique which is mainly proposed to be utilised on the numbers rather than text information which will be utilized in our models. Hence, the BoW model is mainly used to utilise for the

pre-processed text information by converting it into a group or set of words. With this process, number of occurrences of each term/word is used as a component in the sequence. Bag-of-Words is a less request record method which is mostly important in the frequency of words

2.2 N-Grams Model

In N-grams, a word categorisation model is being proposed to use for the most portion of utilization in NLP and text mining [17,3,4]. N-grams is mainly helpful in the utilization of the group of co-occurrences words in given text data and it also finds useful in processing a single word ahead.

2.3 TF-IDF Model

Term Frequency–Inverse Document Frequency, which refers as TF-IDF is most commonly a numerical statistic model which is used to indicate and identifies the mostly utilized word in a document from dataset [28,2]. It also proposed to mainly use as a weighting factor in computing the data extraction and word/text abstraction.

2.4 Term Frequency (TF) Model

In Term Frequency, the considering of the texts occurred in the data or a finding of the inconsistency between the data is being defined [17,11]. Each data is being depicted in a simple vector which it has a frequency of the word. This word/text in general can be find by number of occurrences word seen in a record compared by sum of number of terms in the given record or data[3].

2.5 Word Embedding model

Text inserting can be defined as a group of language reveal and highlighting retrieval techniques in Natural Language Processing (NLP). In text embedding model, jargon are usually converted into a simple vectors of real numbers. Text including is a kind of word depiction which allows the text with a similar importance to have a similar depiction. Sentiment analysis can be used to compute the significance of the express the polarity. The authors [21] applied the Naive Bayes, Logistic Regression with Stochastic Gradient Descent and Vector Machine for Roman Urdu data set. Tripathy et al. [27] researched on data set movie reviews in sentiment analysis of English language using Machine Learning methods. By using ML methods, in the paper [5,9], they are predicted the false news using sentiment analysis. By using this sentiment analysis, the researchers can also predicted pattern of citations with these ML methods by study the text from the research articles [13,8,12]. Akmal proposed a clustering system with a reduced feature set to sentiment analysis with dimensionality reduction[4]. By using deep learning techniques, they have sentiment analysis on COVID-19 reviews [23]. Their algorithm is used on an LSTM-RNN network and enhanced featured by attention layers. By using improved Neural Network approach, they study the customer preferences are identified from online reviews of hotels[7]. By using Long short term memory (LSTM) model, authors study the sentiment analysis in social data for e-commerce products reviews in Hindi languages [31]. Authors done a sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence [1] More details and survey papers see in [30,26,19,18].

3. Methodology

In this paper, the authors used some models like Logistic Regression, Decision Tree,XGBoost, KNeighbors Classifier, Multinomial Naïve Bayes, Bernoulli Naïve Bayes and Random Forest Classifier.

3.1 Logistic Regression model

Logistic Regression is defined as a supervised learning methodology which is mainly proposed to use for calculating or predicting the output of a categorical dependent variable. In addition, it is also proposed to use for solving the classification problems. Furthermore, it generates the values which ranges in between 0 and 1.

It can be usually imported through SKlearn library. In the real world scenarios, we can mainly use it for the Logistic Regression with the application of various areas and fields like Movie Review. Logistic regression can also be applied to predict whether a movie is expected to be watched or not based on the review rating. This method can be applied to predict whether a selective audience will active or not for the given issue in marketing perspective.

$$P = \frac{e^{a+bx}}{1+e^{a+bx}} \tag{1}$$

3.2 K-Nearest Neighbour:

K-Nearest Neighbour is defined as a supervised learning methodology which is mainly used for the

Regression and Classification problem solving. It is mainly proposed for a non-parametric algorithm and also can be called as a lazy learner algorithm. This is because it does not learn from the given training set. In other words, it usually stores the given dataset and performs the action on it during the time of classification. It is more effective towards data set which contain more reviews and robust to the given noisy training data.

$$d(a,b) = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2} \quad (2)$$

3.3 Bernoulli Naïve Bayes Model:

It is a part of the Naive Bayes family which is mainly used for the discrete data. It is mainly based on the Bernoulli Distribution that accepts only binary values, i.e., 0 or 1, Positive, Negative and so on. It is extremely fast compared to other classification models and able to make real time predictions.

$$p(d) = p[D = d] = \{^{y=1-x;d=0}_{x;d=1} \quad (3)$$

Where, $D = \{^{1:Bernoullitrail=S}_{0:Bernoullitrail=F}$

3.4 Multinominal naïve bayes:

It is a supervised learning methodology which is mainly proposed to use for the analysis of the categorical text data. It is purely based on the bayes theorem, which is a powerful algorithm for the text data analysis. It further can be used on both continuous and discrete data processing. It is also used for predicting the real time applications like movie reviews, fake news detection and so on.

$$P\left(\frac{d_2}{d_1}\right) = \left(\frac{P(d_1 \bigcap d_2)}{P(d_1)}\right) \quad (4)$$

3.5 XGBoost

XGBoost can be defined as "eXtreme Gradient Boosting".

It can be defined as a decision-tress-based algorithm which can be mainly used for a gradient boosting frame work that are created in sequential form. It is also proposed to use for solve regression, classification, ranking, and user-defined prediction problems.

$$f(p) \approx f(a) + f^{'(a)}(p-a) + 0.5(f^{"}(a)(p-a)^2) \quad (5)$$

$$L^t \cong \sum_{i=1}^{n} [l(q_i, y^{\wedge(t-1)} + g_i f_t(X_i) + 0.5(h_i f_t^2(X_i)] + \Omega(f_t) \quad (6)$$

3.5 Decision Tree Model

Decision Tree is proposed to be a Supervised Learning Technique which is mainly used in both classification and regression problem solving scenarios. It is a tree structured classifier where the Internal Nodes are defined as a Features of a Data set, Branches are defined as Decision rules and Leaf nodes are the outcome. Decision Tree is usually a mimic human thinking ability in making a decision of specific issue and also it shows a tree like structure. Hence, it is easy to understand and analyse the given problem.

3.6 Random Forest Classifier Model

It is also proposed to be a supervised learning model which is mainly used for both regression and classification problems. The main functionalities of the Random Forest algorithm is to manipulate the data set which contains a continuous elements in scenario of Regression and categorical elements. Random Forest Classifier also generates a decision tree on the given data samples which helps to get the prediction and finally selects the best possible solution.

3.6.1 Dataset

The given data is in text format in csv files. While extraction, a comma-separated values (CSV) file saved in a tabular format of the data (numbers and text) in a plain text. Every line in the given file is a data information. Furthermore, each line record in the file gives of one or more fields which is separated by a comma. Formally, in a given training sample of reviews and sentiment, the sentiment ‘1’ represents the review as negative and the sentiment ‘0’ represents the review as positive. Our aim is to predict the categorise the sentiments on the given text dataset. In this paper, We trained on two different data sets which contains 50,000 reviews each. The test dataset mainly contains 2 attributes i.e., reviews and sentiment. The reviews collected from IMDB are either having negative or positive or sentiments integrated with it whereas a review with sentiment '0' is of positive thought and a review with sentiment '1' is of negative thought.

Overview of all data sets we have shown in Table.1. In this paper, before comparing the seven models, we need to import some packages and perform data pre-processing to train the accuracy of each model.

Table 1

Data sets

Data sets	Total reviews	Number of 0s	Number of 1s
Movie Reviews	34000	16993	17007
Coursera Reviews	107018	102298	4720
Google Play Store Reviews	12485	7635	4850
Flipkart Reviews	9976	8975	1001

3.6.2 Operation on datasets

Import all the necessary packages of Python like Pandas, Numpy, NLTK, SKlearn, seaborn, etc. which are used for analyzing the text reviews. Read the dataset using Pandas and perform the data cleaning operation like removing special characters, numbers, and punctuation's, from the text using NLTK, re-libraries. Later on, for cleaned reviews, we tokenize from our dataset. Tokens are defined as a individual words or terms. Tokenization is nothing but the method of dividing a text/sentence into tokens. In the perform stemming stage, stemming is applied which is a rule-based method of cleaning the suffixes i.e., it usually separates the "s", "es", "ly", and "ing", etc from a given term of the sentence. For better understanding, in the given example – "write", "writer", "writes" and "writing" are the different alteration of the word – "write". Extracting the Features from the cleaned reviews: We use Bag-of-words-features. In this process, collection of the words are used to extract the features from word documents. These features in general will be utilised to train the machine learning techniques. It also creates a vocabulary of all the distinct terms occurring in all the records in the training dataset. TF-IDF Feature: Tf-idf is defined as a term frequency-inverse document frequency which is mainly used for a text weight analysis in the information extraction and word/text mining. This weight can be a proportional measure which is signifies the how significant a word can be in the record/document. The importance could be further increased certain portion to the number of occurrences of a term seen in the record/document. It is also frequently used as a offset by the occurrences of the term in the corpus.

TF: Term Frequency is a measure about how frequently a word occurs in a given document. Since, every given record is distinct in terms of the length and it is finely feasible that a word would appear many more times in the given lengthy records or documents than the shorter record or document. hence, the keyword count is further divided based on the record length which is also known as the the total number of words in the given document for the normalization process. Thus, TF(t) is defined as a number of occurrences the character 't' appears in a given record to the total number of characters in the given record. IDF refers to Inverse Document Frequency which mainly measures about how significant a word is in the given record. During the process of computing the TF, all terms in the given document are considered as equally significant. But it is also known that in different scenarios, the term such as "that","of",and "is", appear more number of occurrences but projected to have very less significance. Hence, we need to focus more on the make it down with the frequent words during the scale up of the irrelevant ones. Function of the IDF is defined as the exponential logarithm of the total number of records with respect to the number of records with the word 't' in it. With this, we can apply all the proposed seven machine learning algorithms on the given data set to get the accuracy of each and every model.

4. Results

In this section, we start analyzing with the data-processing stage to the feature extraction, classification and prediction analysis through different algorithms**.**

4.1. Data Pre-processing

In data pre-processing stage, data contains individual information which needs to be carefully cross verified and completely checked to apply the pre-processing operation. In order to start this operation, we had gone through the training phase, later on testing phase to approve the given information records. Subsequently, we applied some pre-processing like tokenizing, stemming and so on to the given datasets. At this point, the information derived is completely tested and checked by assuming that data has a missing quality.

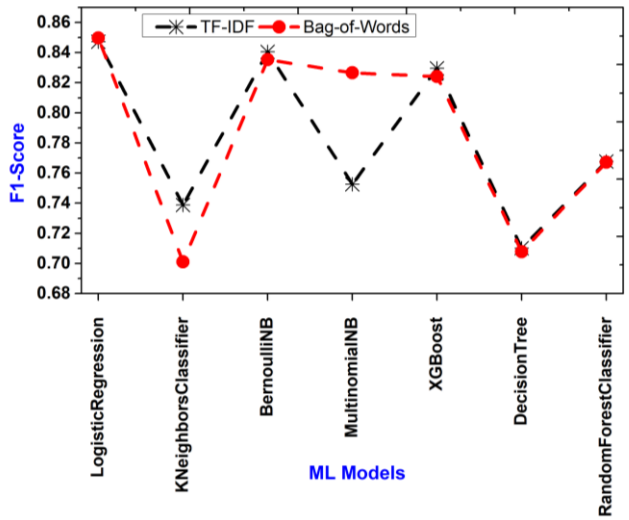

Fig. 1. F1-score for a dataset (movie reviews) using seven ML algorithms with TF-IDF and bag of words

4.2. Feature Extraction

In this feature retrieved phase, we mainly deal with the element retrieval and options of different techniques from the scikit and python. To apply the highlight options, we further utilized a technique named TF-IDF (briefly explained in Section-II). In this, we propose to use a word to separate the vector for measuring the specific highlights. In addition, the pipelining process has been utilized efficiently to utilize the facilitation in the given code.

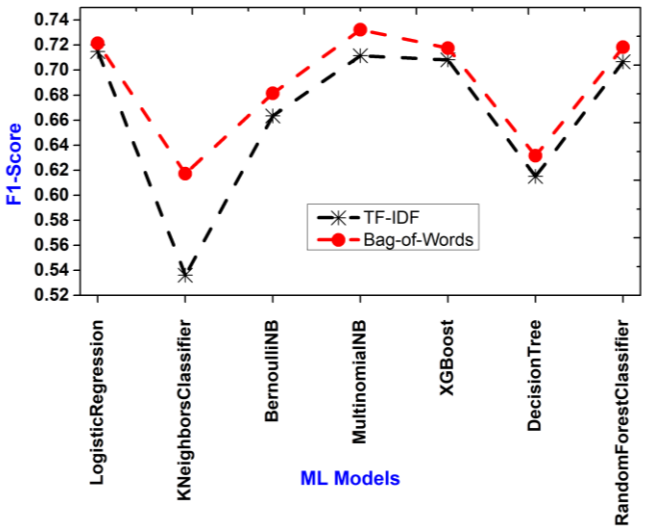

Fig. 2. F1-score for a dataset (google play store reviews) using seven ML algorithms with TF-IDF and bag of words

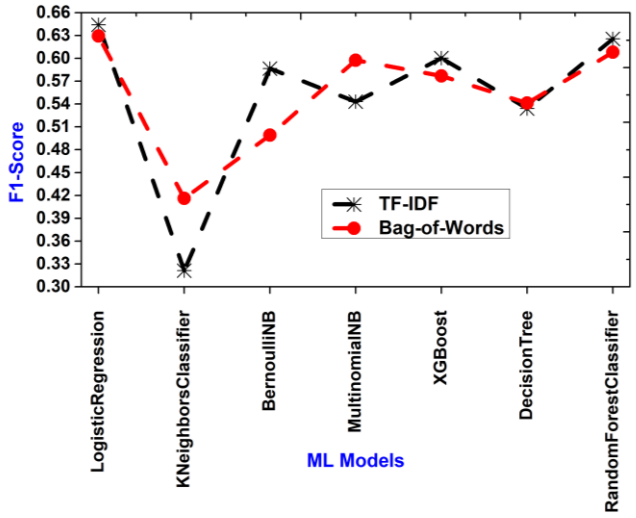

Fig. 3. F1-score for a dataset (Flipkart Reviews) using seven ML algorithms with TF-IDF and bag of words

4.3. Classification

In this phase, we tested the classification of the information into different parts. This is further verified with the test data and train data. Furthermore, the train dataset is mainly categorized into group of matching words with a similar substance. Later on, the test data is coordinated and gathered in such a way that it is classified for further the Naïve Bayes classification. This can be further processed, and the likelihood of every unique term is determined independent with each other.

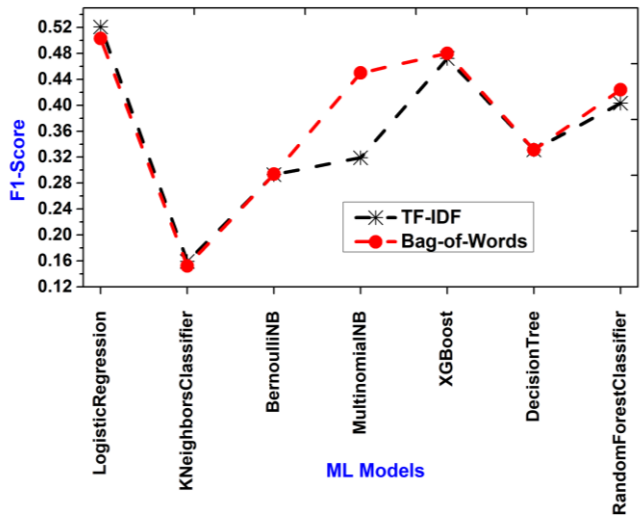

Fig. 4. F1-score for a dataset (Coursera Reviews) using seven ML algorithms with TF-IDF and bag of words

4.4. Prediction

In the last phase, the proposed model was used. The proposed technique is well mapped in the client's computers, and it is also implemented to be used by the app.py file document to find the false news with a good accuracy. Further, it takes a random news record which acts as a connection from the client. At this point, the machine learning technique can be used for a correct accuracy yield which can be established to the client alongside with the similarity of the truth. After training with 7 machine learning models with extracting features i.e. Bag-of-words-features, TF-IDF Feature on 4 different type of data sets, we got accuracy of each model as : Movie Review: This Data Set contain 34000 reviews which mainly consist of 2 attributes, i.e., review and sentiment. Review: It is about the opinion of the user on the movie. Sentiment: A review with sentiment '0' is of positive review while a review with sentiment '1' is of negative review. Coursera Review: This Data Set contain 107018 reviews which mainly consist of 2 attributes, i.e., review and label. Review: It is about the opinion of the user on the course. Label: A review with label '0' is of positive review while a review with label '1' is of negative review. From Table 3, in Movie Reviews data set, Logistic regression and Bernoulli Na¨ıve Bayes performing better accuracy with Bag-of-words and TF-IDF feature respectively see Fig 1. For Coursera data set, Logistic regression and XGBoost algorithms giving good F1-Score for TF-IDF feature extraction from Fig 2. For Google Play Store review data set, Logistic regression and Random Forest classifier algorithms shows more performance compared with other methods for Bag-of-words-feature see Fig 3. For Flipkart reviews, Logistic regression and Random Forest classifier algorithms gives out standing f1-score for Bag-of-words and TF-IDF feature respectively see Fig 4. Clearly, we can observe from these four data sets, Logistic regression is outperforms compared with other machine learning algorithms.

Table 3

Accuracy for various data sets by using machine learning techniques

Model	Features	Movie Accuracy	Coursera Accuracy	Google Play Store Accuracy	Flipkart Accuracy
Logistic Regression	Bag-of-words-feature	84.96	50.29	72.14	62.93
	TF-IDF Feature	84.72	52.05	71.48	64.42
KNeighbors Classifier	Bag-of-words-feature	70.10	15.22	61.73	41.60
	TF-IDF Feature	73.88	15.91	53.62	32.11
Bernoulli Na¨ıve Bayes	Bag-of-words-feature	83.53	29.39	68.15	49.93
	TF-IDF Feature	84.05	29.30	66.35	58.67
Multinomial naive bayes	Bag-of-words-feature	82.66	44.98	73.23	59.76
	TF-IDF Feature	75.25	31.92	71.15	54.32
XGBoost	Bag-of-words-feature	82.41	47.99	71.75	57.70
	TF-IDF Feature	82.99	47.20	70.82	60.03
Decision Tree	Bag-of-words-feature	70.78	33.12	63.16	54.13
	TF-IDF Feature	71.01	33.12	61.51	53.43
Random Forest Classifier	Bag-of-words-feature	76.72	42.40	71.83	60.82
	TF-IDF Feature	76.77	40.31	70.66	62.56

4. Conclusion and Future Work

This paper makes use of the various existing algorithms to compare and contrast the sentimental classification and analysis. Based on the theoretical and experimental analysis, we estimated that the supervised learning methods namely Naïve Bayesian, Logistic regression and XGBoost are concluded as standard learning techniques for sentimental classification and analysis. Within the supervised learning, based on the experimental analysis, we conclude that XGBoost is providing an excellent accuracy (82.99%) when compared with other classification algorithms. In words of category, it is observed that for the the small feature set, Naive Bayes performs very well whereas for the large feature set, XGBoost will be the perfect option. In addition, Lexical based methods are identified as an ideally more significant because of its requires usual task on the given record. Sentiment

analysis can be performed by different machine learning models like CNN, Naive Bayes, Decision Tree, XgBoost, Logistic Regression etc. The proposed work is compared with the existing solutions in terms of different performance metrics. Clearly, we can observe from these four data sets, Logistic regression is out performs compared with other machine learning algorithms.

5. References

[1] A. Adak, B. Pradhan, and N. Shukla, "Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review", Foods, MDPI, 11(10):1500, 2022.

[2] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using n-gram analysis and machine learning techniques", International Conference on Intelligent, Secure, And Dependable Systems in Distributed And Cloud Environments, Vol? pp. 127–138. Springer, 2017.

[3] H. Ahmed, I. Traore, and S. Saad, "Detecting opinion spams and fake news using text classification", Security and Privacy, Vol? pp? 1(1):e9, 2018.

[4] Saba Akmal and Hafiz Muhammad Shahzad Asif. Sentiment analysis based on soft clustering through dimensionality reduction technique. Mehran University Research Journal Of Engineering and Technology, 40(3):630–644, 2021.

[5] K. R. Asish, A. Gupta, A. Kumar, A. Mason, M. K. Enduri, and S. Anamalamudi, "A tool for fake news detection using machine learning techniques", 2nd International Conference on Intelligent Technologies, India IEEE, pp. 1–6, 2022.

[6] B. K. Bhavitha, A. P Rodrigues, and N. N Chiplunkar, "Comparative study of machine learning techniques in sentimental analysis", International conference on inventive communication and computational technologies, India IEEE, pp. 216–221, 2017.

[7] Y. Bian, R. Ye, J. Zhang, and X. Yan, "Customer preference identification from hotel online reviews: A neural network based finegrained sentiment analysis", Computers and Industrial Engineering, 172:108648, 2022.

[8] M. K. Enduri, V U. Sankar, and K. Hajarathaiah, "Empirical study on citation count prediction of research articles", Journal of Scientometric Research, 11(2):155–163, 2022.

[9] R. Gupta, S. Sameer, H. Muppavarapu, M. K. Enduri, and S. Anamalamudi. "Sentiment analysis on zomato reviews", 13th International Conference on Computational Intelligence and Communication Networks, pages 34–38. IEEE, 2021.

[10] S. Gupta, R. Thirukovalluru, M. Sinha, and S. Mannarswamy, "Cimtdetect: a community infused matrix-tensor coupled factorization based method for fake news detection", International Conference on Advances in Social Networks Analysis and Mining, pages 278–281. IEEE, 2018.

[11] Q. He, "Knowledge discovery through co-word analysis", 1999.

[12] S. Himani, M. H. Kumar, M. K. Enduri, S. S. Begum, G. Rageswari, and S. Anamalamudi, "An empirical study on impact of news articles", 13th International Conference on Computational Intelligence and Communication Networks, pages 39–43. IEEE, 2021.

[13] H. Shaik, M. H. Kumar, M. K. Enduri, B.S. Shaik, G. Rageswari, and S. Anamalamudi, "A comparative study on machine learning based prediction of citations of articles", 6th International Conference on Trends in Electronics and Informatics, pages 1819–1824. IEEE, 2022.

[14] M. D. Ibrishimova and K. F. Li,"A machine learning approach to fake news detection using knowledge verification and natural language processing", International Conference on Intelligent Networking and Collaborative Systems, pages 223–234. Springer, 2019.

[15] K. Jain and S. Kaushal, "A comparative study of machine learning and deep learning techniques for sentiment analysis", 7th International Conference on Reliability, Infocom Technologies and Optimization,

pages 483–487. IEEE, 2018.

[16] J. Zhiwei, J. Cao, Y.G. Jiang, and Y. D. Zhang, “News credibility evaluation on microblog with a hierarchical propagation model”, IEEE international conference on data mining, pages 230– 239. IEEE, 2014.

[17] R. K. Kaliyar, “Fake news detection using a deep neural network”, 4th International Conference on Computing Communication and Automation, pages 1–7. IEEE, 2018.

[18] A. K. Bhadra, S. G Shaila, and M. K. Banga, “Review on sentiment analysis and polarity classification of sarcastic sentences using deep learning in social media”, Data Engineering and Intelligent Computing, pages 225–237, 2022.

[19] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey”, Ain Shams engineering journal, 5(4):1093– 1113, 2014.

[20] A. J. Priyavrat, “Sentiment analysis: A comparative study of supervised machine learning algorithms using rapid miner”, International Journal for Research in Applied Science and Engineering Technology, pages 2321–9653, 2017.

[21] A. Rafique, M. K. Malik, Z. Nawaz, F. Bukhari, and A. H. Jalbani, “Sentiment analysis for roman urdu”, Mehran University Research Journal of Engineering and Technology, vol. 38, issue 2, pp. 463–470, 2019.

[22] S. Shabani and M. Sokhn, “Hybrid machine crowd approach for fake news detection”, International Conference on Collaboration and Internet Computing, IEEE, pp. 299–306. 2018.

[23] C. Singh, T. Imam, S. Wibowo, and S. Grandhi, “A deep learning approach for sentiment analysis of covid-19 reviews”, Applied Sciences, 12(8):3709, 2022.

[24] B. S. Desikan, “Natural language processing and computational linguistics: a practical guide to text analysis with Python, Gensim, spaCy, and Keras”, Packt Publishing Ltd, 2018.

[25] P. Sudhir and V. D. Suresh, “Comparative study of various approaches, applications and classifiers for sentiment analysis”, Global Transitions Proceedings, 2(2):205–211, 2021.

[26] H. Tanaya, S. Sagnika, and L. Sahoo, “Survey on sentiment analysis: a comparative study”, International Journal of Computer Applications, 159(6):4–7, 2017.

[27] A. Tripathy, A. Agrawal, and S. Kumar Rath, “Classification of sentimental reviews using machine learning techniques”, Procedia Computer Science, 57:821–829, 2015.

[28] S. Vijayarani, M. J Ilamathi, M. Nithya, et al., “Preprocessing techniques for text mining-an overview”, International Journal of Computer Science and Communication Networks, 5(1):7–16, 2015.

[29] J. Wang, P. Liu, M. F. H. She, S. Nahavandi, and A. Kouzani, “Bag-of-words representation for biomedical time series classification”, Biomedical Signal Processing and Control, 8(6):634–644, 2013.

[30] M. Wankhade, A. C. S. Rao, and C. Kulkarni, “A survey on sentiment analysis methods, applications, and challenges”, Artificial Intelligence Review, pages 1–50, 2022.

[31] V. Yadav, P. Verma, and V. Katiyar, “Long short-term memory (LSTM) model for sentiment analysis in social data for e-commerce products reviews in hindi languages”, International Journal of Information Technology, vol.? pp. 1–14, 2022.

