

Sentiment Analysis: Quantitative Evaluation of Machine Learning Algorithms

Harjasdeep Singh
 CSE Department
 SJTU

Jhunjhunu, Rajasthan, India
 deep.harjas@hotmail.com

Dr. Durgesh Srivastava
 CSE Department
 SJTU

Jhunjhunu, Rajasthan, India
 drdkumar.ptu@gmail.com

Abstract— Sentiment analysis is a computer method for analyzing the opinions, feelings, and thoughts expressed in online comments and reviews. Sentiment analysis is a topic that has recently gained attention in the research areas of NLP and text analysis. Understanding public sentiment is greatly aided by doing sentiment analysis on this user-generated data. Two methods are used to extract emotions from text: the knowledge base approach and the machine learning methods. This research work has developed a machine learning model to find the best technique for performing sentiment analysis. For this purpose, various machine learning classifiers were compared by including reviews of movies. Accuracy and f1 metrics were also used for the evaluation of algorithms.

Keywords— *Sentiment analysis, Machine learning, NLP, Opinion mining.*

I. INTRODUCTION

Social media websites have become popular in recent years. On these websites, users are so engaged that they discuss and share their opinions on newsworthy topics, their goals, and goods and services that they have used recently or in the past. Businesses are trying to understand the importance of social media postings and tweets from individuals throughout the globe. This is done with the goal of amassing as much information as possible that could provide additional insight about the relevant topics in which individuals are engaged. This information can be used to recommend products and know the interests of the people. User opinions obtained from these microblogging websites is very useful.

Sentiment analysis is the area which deals with people's opinions, behavioural responses, attitudes and sentiments. With the development of deep learning and machine learning, sentiment analysis has become one of the hot research areas.[1] Sentiment analysis offers a broad picture of the general public opinion on subjects that appear in a range of articles, from those on politics to user reviews. This boom in the area of opinion mining comes with the interest and participation of the people in social media website and by writing reviews of products on websites.[2]

The processes of evaluating, analysing, summarizing, and extracting conclusions from subjective writings are referred as sentiment analysis. It uses natural language processing (NLP) to determine whether or not a set of data is positive, negative, or neutral, sentiment analysis is also commonly referred to as opinion mining. It is common practice for companies to do sentiment analysis on text data in order to monitor how customers discuss their brands and goods in online reviews and to gain a deeper comprehension of the demographics of their ideal customers.[3] The two primary classification approaches used in today's text sentiment

analysis research are rule-based classification method and machine learning classification.

Types of Sentiment Analysis: Forms of Sentiment Analysis that are most frequently used are:

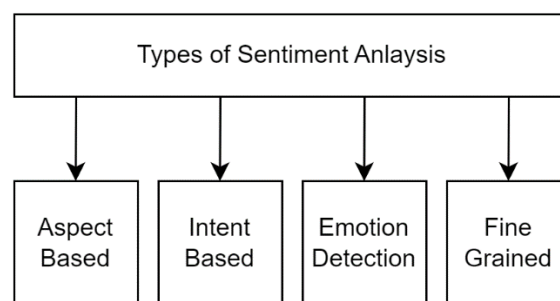


Fig. 1. Types of sentiment analysis

Fine grained: If polarity accuracy is crucial to a company, then company or organization might consider fine grained sentiment analysis.[8] In fine grain opinion mining different levels of polarity are: positive, very positive, neutral, negative and very negative.

Aspect Based: In this type text classification is based on aspect of the text and then to determine its sentiment. Aspect based opinion mining is used to determine user feedback through the association of sentiments with different features of a product.

Emotion detection: The goal of emotion detection is to identify several sorts of sentiments expressed in texts, including disgust, fear, pleasure, sorrow, and surprise. This has the benefit of helping businesses understand how a consumer feels a certain way, but it may also be extremely difficult because some people describe their emotions by using long, complex language.

Intent-based: This type user input is inspected and attitude is determined as neutral, positive and negative. In intent-based sentiment analysis, intent of customer is detected and also customer can be tracked and targeted for advertisement

II. PREVIOUS WORK

The problem of opinion mining using textual evaluations has attracted the attention of several researchers. The overall sentiment either positive or negative, reflected in the textual reviews is the subject of our research. The overall sentiment (positive or negative) reflected in the review is the subject of our research.

The relevant research that had done in the subject of opinion mining is discussed in the section that follows along with various methods and techniques used in sentiment analysis for textual review data. There are microblogging data and text reviews available now a days which helped researchers to perform opinion mining on that data.

In the paper by [1], tweets data obtained by using various API was used to pass from several machine learning algorithms for classification. After that final prediction was made using a voting classifier to predict the class of a tweet. In this work, DL techniques were used to know the class of the text reviews obtained from several social media platforms. For text classification, deep learning methods outperform machine learning methods.

In [2] research paper authors worked on the sentiment analysis over a movie review dataset. Dataset was divided in test and train parts. Researchers used sentence level opinion mining because it reduced computational power required to evaluate text review. SVM, KNN, Decision Tree and Naïve Bayes algorithms were compared using accuracy measure technique. In their experimental study, SVM classifier performed best among other classifiers. In their study support vector machines achieved accuracy of 72.7%. Authors evaluated the performance of machine learning classifiers based on metrics like accuracy, Sensitivity.

In the paper by [3], text review data extracted from many major cities was used. For the polarity, Textblob library was used. Using various scores obtained from Textblob library, polarity of the tweets was labeled as positive or negative. In the classification task, authors used four types of naïve bayes algorithm. According to their experiment Naïve Bayes multinomial classification algorithm performed better than others.

In [4], authors suggested a lexicon method for opinion mining using word2vec embedding technique. This study proposes utilising sentiwordnet to increase sentiment words. Authors have evaluated the effectiveness on movie review dataset. According to them, Senti2vec technique developed by them is more effective than other methods of opinion mining.

Research paper [5] had implemented sentiment analysis by using naïve bayes machine learning method. The authors had used the Indonesia election dataset obtained from twitter social media by using crawlers. The researchers compared the performance of naïve bayes algorithm with other methods of classification such as KNN and SVM. Naïve bayes method performed better than KNN and SVM algorithm. Further sentiment analysis can also be performed on date from Facebook and Instagram social media websites.

In paper [9] authors focused on implicit aspects. Authors have discussed applications and terminology related to implicit aspect-based sentiment analysis. Hybrid, supervised and unsupervised were the categorization of opinion mining. According on authors' survey, unsupervised classification techniques were prevalent because these techniques do not require training data. In case of performance, Supervised classification methods were efficient. In hybrid techniques, performance was good as number of techniques were combined.

In [10] paper, authors examined Recurrent Neural Networks on three datasets for sentiment analysis. When compared with other recurrent neural networks, the performance of the gated variant was superior. According to authors, deep Recurrent Neural Network should only be used when dataset for classification is big. Convolutional neural networks along with recurrent neural networks could be combined with RNN to make a hybrid model.

[11] have used four different machine learning methods to compare the accuracy of the models created. Textblob library was used to identify the polarity of given text reviews or tweets. Tweets can be categorized as positive, neutral and negative according to sentiment words present in those tweets. In this research work both supervised and unsupervised techniques were used. The tweets data was evaluated using a number of different machine learning techniques, including the Naive Bayes, the Random Forest, the Logistic Regression and the Support Vector Machine learning methods. The results of these models were put to the test using a variety of testing criteria, including precision, recall, and f1-score.

In [13] paper a new sentence vector technique was used which is using information obtained from opinion dictionaries and word vectors. In this purposed technique weights of sentiment and neutral words were considered separately. Whole information retained in this technique along with sentiment words. For classification Support Vector Machine classifier was used for classification of text and this supervised machine learning method was compared with popular rule-based sentiment dictionary and TF-IDF weighting techniques.

Authors in their paper [16], suggested a dependency parsing method to be used along with opinion association migration and familiar distance for the sentiment analysis of small content in tweets. Their experimental results shows that the method used by them was accurate and structure of the sentiment also play important role in text classification. Furthermore, it demonstrated how each phrase contributes differently to the short text sentiment computation, depending on the connection between the modifier and the emotion term.

The accuracy of the model used in the earlier research was a significant limitation of that work. They examined the tweets by using several of methods of supervised machine learning. In this research work, approach is relatively comparable to the one that was used before; nevertheless, there has been a significant shift in the assessment parameters of the model, which emerges as the most significant limitation; hence, there is a necessity for a new model that has a higher level of accuracy.

III. MACHINE LEARNING METHODS

A. Naïve Bayes:

In Naïve Bayes classifier Thomas Bayes' theorem is basic form factor. The Naive Bayes classification method operates on the presumption that each feature is absolutely unrelated to every other feature and that they also have no bearing on one another. The Naive Bayes method is simple to construct and operates on the fundamental idea of probability. It is helpful in situations involving very big datasets. [2] Since it

assumes that the presence of one characteristic is unrelated to the occurrence of other features, the Naive Bayes method is known termed "naive". Calculating the conditional probability that an item with a feature set belongs to a specific class is the primary goal of the Naive Bayes method. For Naive Bayes, we can write the conditional probability as

$$P(X|y_j) = \prod_{i=1}^m P(x_i|y_j)$$

Where X denotes a feature vector and y_j is the class label.

Results are anticipated by comparing the posterior probabilities, which are determined for all feature classes.

B. Support Vector Machine

It is extensively used supervised machine learning method. This method was invented by Vladimir N. Vapnik and Alexey Ya. SVM is used for regression and classification tasks. This method's goal is to locate a hyperplane in N -dimensional area that accurately categorizes the dataset by maximising the distance between the two classes. Classification accuracy could be increased by optimizing the decision boundary in SVM. By including some more characteristics as greater dimensions, C Parameter, Multiple Classes, and Kernel Trick, Support Vector Machine method can be made even more effective. SVM makes use of a discriminative function that can be expressed as

$$g(X) = w^T \phi(X) + b$$

Features are represented by the vector ' X ', while weights and biases are represented by the ' w ' and ' b ' respectively. Non linear mapping is represented by ϕ . This method can be used to distinguish between linear and non-linear space.[6] SVM can be divided into two subclasses, those are Linear Support Vector Machine (LSVM) and Nonlinear Support Vector Machines (NLSVM). The Linear Support Vector Machine (LSVM) classifies data using two-dimensional datasets that may be linearly separated. Datasets for nonlinear support vector machines (NLSVM) cannot be separated.

C. KNN

One of the easiest machine learning methods is the KNN algorithm. This method is very easy to understand and works really well in actuality. KNN algorithm does not make any assumption on the data being passed to the algorithm and because of this property it is known as method for non-parametric learning. This method uses labeled data for training and testing as it falls under the category of supervised machine learning. In knn we used minkowski as a distance metric. Minkowski distance can be expressed as

$$\text{Minkowski Distance} = \sum_{i=1}^n |x_i - y_i|^{1/p}$$

KNN is used for labeled data means it falls under the category of supervised learning technique. KNN is used because it is simple and easy to implement. This method uses the Euclidean distance formula to classify data points. Category of data points is assigned where more number of neighbors of same category are there.

D. Decision Tree

The algorithm, decision tree is a type of supervised machine learning which is used in classification tasks. Decision tree could be used for regression and as well as classification problems. It is a flowchart type of algorithm in which each internal node resembles a test on feature and branch nodes shows the outcome while terminal nodes represent a class label. For splitting of decision tree, we used Gini index. Gini index determines how often a feature is incorrectly classified when chosen randomly. In mathematics, the Gini Index is represented as the difference between the total squared probability of each class and one.

$$\text{Gini Index} = 1 - \sum_{i=1}^n (P_i)^2$$

Here, P_i represents the likelihood that a given element belongs to a particular category. Decision tree works on continuous and categorical type of variables. The main idea behind the decision tree is that dataset features are used to create yes or no questions and tree is continually split the dataset until all data points belonging to same class. Goal is to split feature space and apply rule, until there are no more rules to apply or no data point is left.

IV. PRESENT WORK

In this research work, we built the machine learning model which predicts results using different algorithms (support vector machine (SVM), Naïve Bayes, k nearest neighbor (KNN) and decision Tree). Main goal of this research work was to find best algorithm for the task of sentiment analysis. A movie review dataset was used in which reviews were labeled as positive or negative. In this task we used supervised machine learning algorithms for the classification of reviews. Evaluation parameters used are accuracy and F1 score.

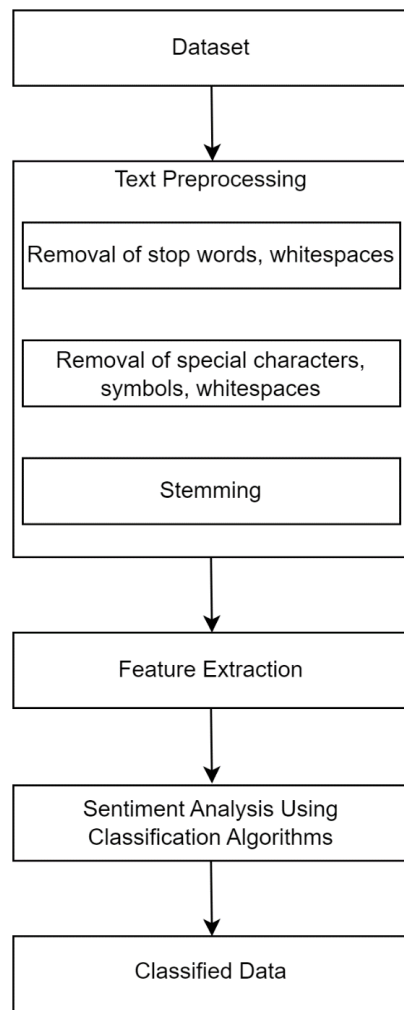


Fig. 2. Purposed Methodology

In the first step Dataset of movie review used which was collected from API available. Reviews are labeled as positive and negative according to the sentiment present.

In the second step, pre-processing of user review data is done. During this step, stop words, special characters, symbols, whitespaces, numbers, alphanumeric are removed from the text data. For the removal of stop words, we used NLTK library. NLTK facilitates stop word removal. NLTK's corpus module lists stop words for removal. By breaking down the text word by word and checking against the NLTK's list of stop words, we eliminated stop words from the phrase.

After that stemming is performed on the text. Stemming replaces the words written by user with root words of those words. Next step is of feature selection. In opinion mining, feature selection is technique to identify subset of features to reach objective. These objectives are to increase accuracy of the model, to reduce computational power required for this work and to lower down overfitting of the model. Further Feature selection can be divided into different categories, these categories are embedded techniques, filter techniques and wrapper techniques.

For the final phase various machine learning classification algorithms are used to categorize text data into positive or negative. In this experimental work, unigram feature extraction method is used to inspect labeled data.

To classify the reviews into positive or negative sentiment, we used KNN, SVM, Naïve Bayes, Decision Tree supervised machine learning algorithms. for this purpose, a model was built using python programming language. In this model, machine learning algorithms were trained using training data available and after that performance of each algorithm was evaluated using different metrics.

To perform the classification on data, we split the data into training and test dataset. The machine learning classifier is taught to make predictions using training data. Classifiers are evaluated on test datasets, and the resulting data utilized to make predictions.

A. Dataset details:

In this research work we used IMDB movie review dataset which is a binary sentiment classification dataset. This dataset consists of 50000 reviews. Review and label are two features in dataset. In this dataset, the labels have a sentiment that ranges from 0 (negative) to 1 (positive). We divided the dataset into training set for the training of the model and test set for the validation of the model.

B. Tf-idf vectorization:

Term frequency — Inverse document frequency (TFIDF) is a text vectorization method. It's a technique for determining how unique a given word is by comparing the frequency with which it appears in a document. This approach is widely used to convert text to a numerical representation that may be used to train a machine learning model for prediction. A tf-idf vectorizer is a method that takes into consideration two factors in the vectorization procedure which are term frequency and inverse document frequency. A phrase's TF-IDF is determined by multiplying its TF score by its IDF score. In this research paper we extracted features from text using Tf-idf method.

C. Evaluation Parameters

The accuracy metric indicates how many predictions were accurate. Test accuracy is the percentage of a test set that machine learning classification algorithm effectively classifies is called as accuracy of classifier.[12] This metric is often referred to as the classifier's overall identification rate in the research. The characteristics of the model, which determine how well the prediction is made, are listed below.

1) F1 score: The F-measure is used to quantify the reliability of a test. The F1-score takes the mean of the precision and recall statistics to produce a single composite statistic.

2) Training Accuracy: Training accuracy of a machine learning model is the accuracy in relation to the examples that were used to build it.

3) Validation Accuracy: It is also known as test accuracy; validation accuracy of a machine learning model is the accuracy in relation to examples that were not used to build it. [15].

V. RESULTS

For classification we used KNN, SVM, Naïve Bayes, Decision Tree classifiers. These machine learning classifiers were compared with each other. The research work makes use of library called Natural Language Toolkit. The NLTK library facilitates rapid development by providing a selection of machine learning models. In this research work we used jupyter notebook, python programming language version 3 or higher, Pandas Library, Scikit-learn library, re library, NumPy library, Matplotlib library

We performed several experiments using machine learning approaches after first preprocessing text data and using feature extraction strategies. features were extracted using Tf-idf vectorization.

There are four models we used to solve the sentiment analysis problem; however, the accompanying table makes it very evident that support vector machine provides the highest accuracy of 87.5%.

In the Naive Bayes algorithm, we used $\alpha=1.0$, $\text{fit_prior}=\text{True}$, $\text{class_prior}=\text{None}$ as the parameters. While in support vector machine classifier, parameter values are $\text{penalty}='l2'$, $\text{loss}='squared_hinge'$, $C=1.0$. parameter values in case of KNN algorithm for this experimental work were set as $\text{metric}='minkowski'$, $\text{n_neighbors}=5$, $\text{weights}='uniform'$, $\text{leaf_size}=30$. In decision tree classifier $\text{criterion}='gini'$, $\text{min_samples_split}=2$, $\text{min_samples_leaf}=1$.

TABLE I. ACCURACY COMPARISON OF CLASSIFIERS

Classifier	Accuracy	Accuracy from [2]
Multinomial NB	84.4%	52.8%
SVM	87.5%	72.7%
KNN	71.3%	50.5%
DT	71.1%	50.0%

In the Table 1, we compared the performance of the different machine learning classifiers. It is clear from the table 1 that our results are better than the research paper [2]. In research work [2], Support vector machine performed better with accuracy of 72.7%. SVM model developed by us showed accuracy of 87.5%.

F1 Score, Precision and Recall for each of the ML methods used to develop our model is detailed in the table below:

TABLE II. CLASSIFIER'S PRECISION, RECALL AND F1 SCORE COMPARISON

Classifier	Precision	Recall	F1 Score
Multinomial NB	0.84	0.84	0.84
SVC	0.88	0.88	0.88
KNN	0.71	0.71	0.71
DT	0.71	0.71	0.71

Figures 3 to 7 are the plot of receiver operating

characteristic curve. It is a graphical representation of the accuracy of a classification model across a range of threshold values. This curve is plotted between two parameters true positive rate and false positive rate. It is clear from the figure 4 have more area under the curve which indicates the better performance.

SVM algorithm performed better in this experimental set up because the differences across classes are reasonably large. We used optimal parameters in SVM. Performance of classifiers also depends upon pre-processing of text. Classifier performs better when there is less noise in the text.

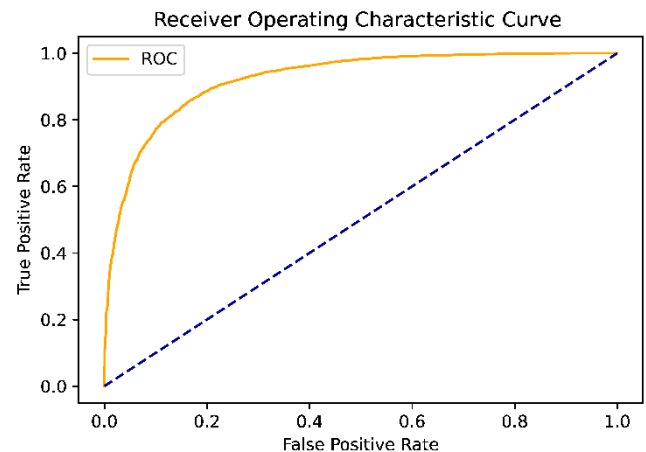


Fig. 3. ROC curves for Multinomial NB

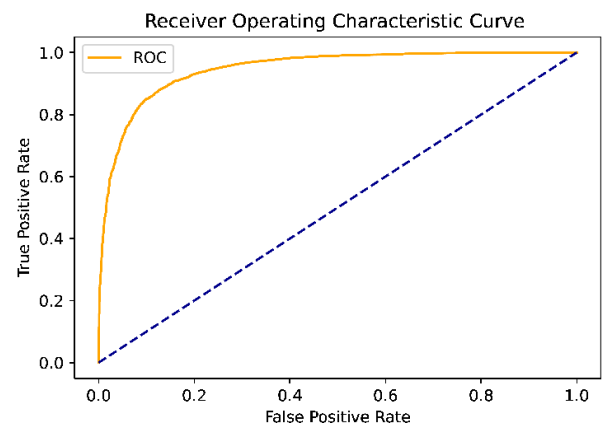


Fig. 4. ROC curves for support vector machine

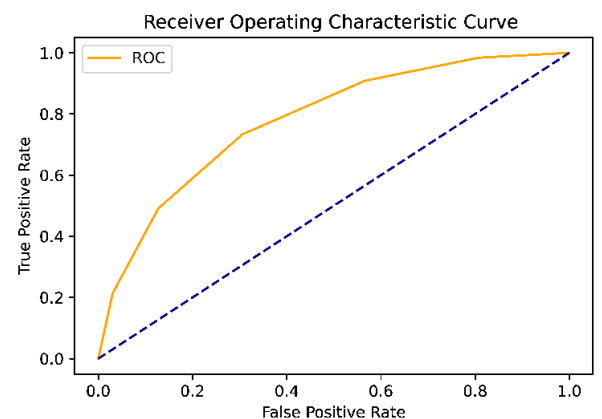


Fig. 5. ROC curves for KNN

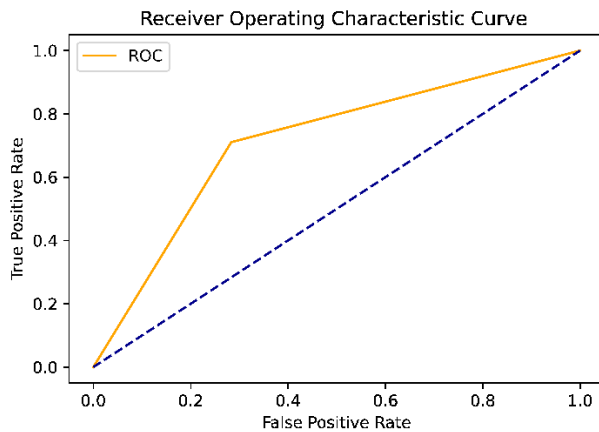


Fig. 6. ROC curves for Decision Tree

Based on the accuracy score by models in Table 1, bar chart is built as shown in figure below. Support Vector machine method have the highest accuracy, followed by NB, KNN and Decision Tree.

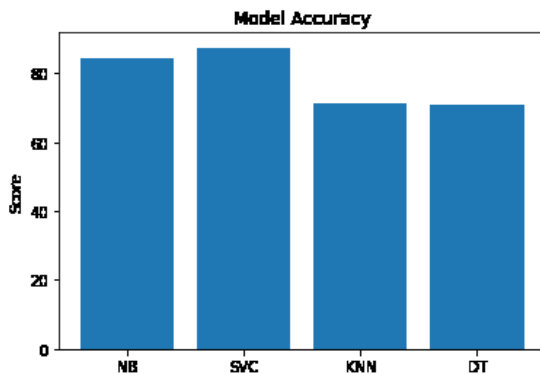


Fig. 7. Comparison of Algorithms

VI. CONCLUSION

All of the methods used in the research are evaluated using the given performance metrics, and the outcomes are reported. Comparisons are made between the available models, and the most effective one is identified for use in the categorization of opinion analysis data. Using supervised machine learning techniques, this research aimed to provide an understanding for sentiment analysis. Machine learning methods like KNN, SVM, Naïve Bayes, Decision Tree were used in this research work. According to results, SVM achieved accuracy of 87.5 % while decision tree algorithm showed lowest accuracy. To evaluate the model, we utilise the F1 score and the accuracy rate. Our results are better as compared to other research works mentioned. The study's goal is to classify a movie review data into either positive or negative opinion. When sentiment features are utilised in place of typical text categorization, the performance of text classification is enhanced.

REFERENCES

- [1] Yogesh Chandra; Antoreep Jana "Sentiment Analysis using Machine Learning and Deep Learning, 7th International Conference on Computing for Sustainable Global Development, ISBN:978-93-80544-38-0, 2020 pp 1-4.
- [2] R. Ragupathy, Lakshmana Phaneendra Maguluri, "Comparative analysis of machine learning algorithms on social media test", International Journal of Engineering & Technology, ISSN: 1793-8236, 2018, pp 284-290.
- [3] Kudakwashe Zvarevashe, Oludayo O. Olugbara, "A Framework for Sentiment Analysis with Opinion Mining of Hotel Reviews", Conference on Information Communications Technology and Society, ISBN:978-1-5386-1002-2, DOI: 10.1109/ICTAS.2018.8368746, 2018.
- [4] Eissa M.Alshari, Azreen Azman, "Effective Method for Sentiment Lexical Dictionary Enrichment based on Word2Vec for Sentiment Analysis" Fourth International Conference on Information Retrieval and Knowledge Management, 2018, pp 177-179.
- [5] Meylan Wongkar, Apriandy Angdresy, "Sentiment Analysis Using Naive Bayes Algorithm of The Data Crawler: Twitter", Fourth International Conference on Informatics and Computing (ICIC), ISBN:978-1-7281-2208-3, DOI: 10.1109/ICIC47613.2019.8985884, 2019.
- [6] Zeliha Ergul Aydin, Zehra Kamisli Ozturk, Zeynep İdil Erzurum Cicek, "Comparison of Machine Learning Algorithms on a Sentiment Analysis Application" Turkish Online Journal of Distance Education-TOJDE, ISSN 1302-6488, 2021, Volume: 22, pp 124-140.
- [7] Hamza, Mohammed; Gupta, Sawar (2022): A COMPARISON OF SENTIMENTAL ANALYSIS ALGORITHMS ON TWITTER DATA USING MACHINE LEARNING, TechRxiv, 2022, pp 1-9.
- [8] Victoria Ikor, Maria Sharmina, "Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers", Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2018, pp 95-98.
- [9] PIYUSH KUMAR SONI, RADHAKRISHNA RAMBOLA, "A Survey on Implicit Aspect Detection for Sentiment Analysis: Terminology, Issues, and Scope", VOLUME 10, 2022, pp 63932-63957.
- [10] Kiran Baktha, and B K Tripathy, "Investigation of Recurrent Neural Networks in the field of Sentiment Analysis" International Conference on Communication and Signal Processing, 2017, pp 2047-2050.
- [11] Chhinder Kaur, Anand Shama, "Social Issues Sentiment Analysis using Python", International Conference on Computing Communication and Security (ICCCS), ISBN:978-1-7281-9181-2, DOI: 10.1109/ICCCS49678.2020.9277251, 2020.
- [12] Tarun Anand, Vikrant Singh, "Survey Paper: Sentiment Analysis for Major Government Decisions", International Conference on Intelligent Engineering and Management (ICIEM), ISBN:978-1-7281-4098-8, DOI: 10.1109/ICIEM48762.2020.9160252, 2020.
- [13] HAO LIU, XI CHEN "A Study of the Application of Weight Distributing Method Combining Sentiment Dictionary and TF-IDF for Text Sentiment Analysis" IEEE Access, ISSN: 2169-3536, pp 32280 – 32289, 2022.
- [14] Avinash Surnar, Sunil Sonawane, "Review for Twitter Sentiment Analysis Using Various Methods", International Journal of Advanced Research in Computer Engineering & Technology, Volume 06, Issue 05, May 2017, ISSN: 2278 – 1323, 2017.
- [15] Satuluri Vanaja Meena Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data", International Conference on Inventive Research in Computing Applications (ICIRCA 2018), 2018, pp 1275-1279.
- [16] Jie Li; Lirong Qiu, "A Sentiment Analysis Method of Short Texts in Microblog", International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2017, pp 777-779.