e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

SENTIMENT ANALYSIS FOR PRODUCT REVIEW

Balpreet Kaur[*1], Kavya Srivastava[*2], Er. Priyanka[*3]

[*1,2]UG Student Of Department Of Information Technology, Shri Ramswaroop Memorial College Of Engineering And Management Lucknow, Uttar Pradesh, India.

[*3]Associate Professor, Department Of Information Technology, Shri Ramswaroop Memorial College Of Engineering And Management Lucknow, Uttar Pradesh, India

DOI : https://www.doi.org/10.56726/IRJMETS36776

ABSTRACT

Sentiment analysis is the technique of using data mining to study, view or analyse sentences in order to anticipate their emotional content using natural language processing . Text is divided into three categories for sentiment analysis: "Positive," "Negative," and "Neutral". It evaluates the data and assigns positive and negative labels to the "better" and "worse" sentiments. As a result, World Wide Web has developed into a vast source of user - or custom-generated raw data in recent years. Users can express their opinions and sentiments by using social media, e-commerce websites, and movie review sites like Facebook, Twitter, Amazon, and Flipkart. In the WWW, where millions of individuals share their opinions in their everyday interactions, whether on social media or in e-mails, which may be their feelings and thoughts about a specific subject. For any kind of decision-making process, whether positive or negative, these expanding raw data provide an extraordinarily high supply of knowledge. The field of sentiment analysis has emerged to automatically analyse such massive amounts of data. Sentiment analysis's primary goal is to classify online data by determining its polarity. Although sentiment analysis is a text-based process, it might be difficult to determine a sentence's precise polarity. This asserts that a superior solution must be found in order to outperform any prior method or strategy utilised to determine sentence polarity. Thus, there is a need for automated data analysis approaches to determine the polarity or emotion of a user or client. In addition to a novel strategy that is proposed in this research, a thorough survey of several sentiment analysis methods is included in this paper.

I. INTRODUCTION

There is a tremendous quantity of information, reviews, and Opinions are being stored as raw data on social media and e-services websites. to utilise those unprocessed data correct techniques are needed. Most techniques either concentrate on adjectives, adverbs, nouns, or verbs. Nonetheless, a current study has showed that the sentiment's use of adverbs and adjectives

Adjectives alone are not as effective as analysis. Nonetheless, no work has focused on all potential adverb and adjective combinations plus verbs. This essay provides a theoretical examination of various well-known approaches to or suggestions for sentiment analysis. the two are the following are the benefits and drawbacks of the strategies discussed: In the suggested strategy, new features are taken into consideration. the fresh approach incorporates document-level machine learning with a blend of verbs, adverbs, and adjectives. The subsequent adverbs and adjectives are taken into consideration for analysis. Adjectives-verbs, verb-adjectives, and verb-adverbs-adjectives besides verbs, adjectives, and adverbs. To derive results and conduct analysis, standard classifiers including Naive Bayes (NB), Linear Models, and Decision Trees are employed. The classification of sentiment analysis is covered in this section, which is followed by a thorough review of the sentiment analysis techniques now in use.

1.1 CLASSIFICATION OF SENTIMENT ANALYSIS

To derive results and conduct analysis, standard classifiers including Naive Bayes (NB), Linear Models, and Decision Trees are employed. The classification of sentiment analysis is covered in this section, which is followed by a thorough review of the sentiment analysis techniques now in use.

Theoretically, sentiment analysis can be carried out in two ways:

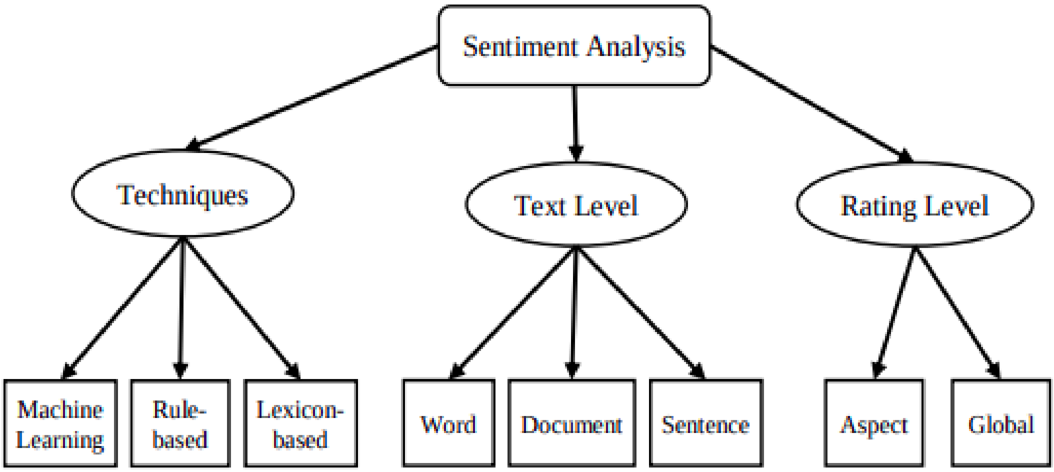

Fig.1. Categorization of Sentiment Analysis

Sentiment analysis can technically be performed in two ways:

1. Machine learning: Datasets must first be trained. Polarities are found using common machine methods.

2. Rule-based: Takes information from a dataset and attempts to rank it based on the opposite direction of concepts. There are various rules such as negation terms, idioms, dictionary polarity, emoticons, and so on.

3. Lexicon-based: By measuring the opinion and subjectivity of a review or comment and applying semantic orientation, it produces sentiment polarity (either positive or negative).

SA can be further divided into aspect level, document level, and sentence level based on the dataset structure or text level

The goal of document-level sentiment analysis is to categories a viewpoint or opinion as "good" or "negative." The entire document was viewed as a collection of information units. Sentiment analysis at the sentence level seeks to categories each sentence's expression of sentiment or opinion. Yet, because sentences are simply shorter than documents, there is no real distinction between classification at the document level and at the sentence level.

In order to determine if a sentence or document is "positive" or "negative," we must apply the document level approach. A product can be rated both at the aspect level and the overall level. Another SA classification is this one. Despite being the majority of e-commerce sites, movie review websites gauge the strength of public opinion globally.

II. EXISTING WORKS

People today are more social through social media, the internet, online shopping, etc. So, whether directly or indirectly, internet opinions and judgements are eventually receiving a lot of attention. The actual deal, though, is analysis or opinion mining. The review of a few current SA solutions is provided below. In Table, these techniques are also briefly summarised. When identifying features, nouns from dataset or reviews are extracted. Frequencies higher than the threshold frequency are kept else dismissed.

Product features and user reviews were gathered using OPINE, an unsupervised, web-based information extraction technique proposed by Propescu et al. It finds product features, opinions about those features, establishes the polarity of those opinions, and then ranks products in accordance with those rankings. Nouns are extracted from datasets or reviews for feature identification. Frequencies over the threshold frequency are preserved; otherwise, they are discarded. Exact features (presence of frequent features) are extracted using OPINE's feature assessor. To extract data, researchers have utilised a manual extraction rule. OPINE's advancement is its domain independence. Yet, because the OPINE system is not widely accessible, it does not find usage in daily life.

Sentence Analysis: Adverbs and adjectives are preferable to adjectives Benamara et al. presented Alone, a linguistic approach to sentiment analysis at the document level, in 2006. Adverb-adjective pairings and the degree of adverb intensity were first measured in this study using linguistic classifiers (using Scoring Methods).

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

The aforementioned scoring techniques employed here are Adverb First Scoring, Adjective Priority Scoring, and Variable Priority Scoring. All of these techniques have the same objective, which is to add a relative weight (in a variable, on a scale of 0 to 1) of the adverb score relative to the adjective score. This study seeks to identify the weight that most closely resembles how people allocate weight to different viewpoints. An analysis that best captures the emotions of readers must include 35% of adverbs together with adjectives, according to study on 200 news articles. Obtains a Pearson correlation (correlation between human sentiment and sentiment analysis algorithms) of approximately 0.47 (range between -1 and 1).

Although having a stronger Pearson correlation, this method only took into account a small number of datasets.

Opinion Digger, one of the answers to sentiment analysis, was introduced by Moghaddam and Ester. This form of unsupervised machine learning operates at the sentence level. standard rating criteria and product attributes are compared and correlated (used in Amazon, Snapdeal, flipkart1 etc). The two submethods in this proposal are separated. Initially, the incoming data is broken up into sentences. Repeated nouns are classified as aspects in the sentences. If an aspect (repeated nouns) pattern emerges, it is recorded. Second, elements are rated according to a scale (where 4 denotes "Excellent," 3 "Average," etc.) and are then given the appropriate labels of "Good," "Average," and "Poor". Its outstanding performance in product rating at aspect level with a loss of just 0.49 is a significant benefit. It was compared with very few techniques because of its demanding criteria and known data to rate, which are two of its biggest limitations. Thus, there aren't enough performance comparisons.

Table.1. Comparison Table

Method	Year of proposal	Classification	Text Level	Prediction Accuracy	Pros	Cons
OPINE	2005	Unsupervised rule-based approach	Word	87%	Domain independent	Difficulty in availing OPINE system, thus rare to get applied in real life.
Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone	2006	Linguistic approach	Document	Pearson correlation of 0.47	Adjectives are given more priority(adjectives expresses human sentiments better than adverbs alone)	None
Opinion Digger	2010	Unsupervised machine learning method	Sentence	51%	Rates product at aspect level	Requires rating guidelines to rate. Works only on known data.
Sentiment Classification Using Lexical Contextual Sentence Structure	2011	Rule based approach	Sentence	86%	Said to be domain independent [6]	Depends solely on wordNet
Interdependent Latent Dirichlet Allocation	2011	Probabilistic graphical model	Document	73%	Faster in comparing and correlating sentiment and rating	Correlation between identified clusters and feature or ratings are not explicit always[6]
A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating	2011	Machine Learning	Document	71% (in 3 categories) 46.9% (in 5 categories)	Automatic calculation of feature vector	Use of WordNet

Classification of Sentences from Internet Customer Reviews Khan et al. suggested using lexical contextual sentence structure. A semantic or rule-based (Dictionary Polarity) method of customer review analysis is presented in. First, each word is stored using the "POS" approach once input has been broken up into sentences. Second, the polarity of the supplied statement is determined depending on the context and sentence structure. Aspects are nouns that are created. The SentiWordNet concept of semantic score is utilised to categorise the sentence as positive or negative. 86% accuracy is produced. Said to be domain independent (review subject), which is a benefit, although the author only gathered limited data (about 3600). Its complete dependence on WordNet is a major issue .

A probabilistic graphical model for grading products at the aspect level is the Interdependent Latent Dirichlet Allocation proposed by Moghaddam and Ester. The majority of review websites use the number of stars as a rating system. The suggested approach performs the same thing by assuming that an aspect (feature) and its corresponding rating are interdependent. The goal of this approach is to produce and display multinomial distributions that translate cluster head words into aspects and reviews into ratings.

The representation of each discrete data point in the pool is a finite mixture over a set of latent variables. A rating accuracy of roughly 73% was discovered. Since graphical representation is

susceptible to flaws and mistakes in data representation, this method could not give the results that are expected.

De Albornoz et al. established a joint model of feature mining and sentiment analysis for product review rating. With all opinions taken into account at once, this machine learning system ranks products on a worldwide scale. There are basically four steps that make up this method. Important elements of the paper or study are initially highlighted. Second, features (aspects) in sentences are recognised. Next, the sentences' polarity and strength are determined. The final step is to rate products worldwide at the aspect level. Weights for features are determined automatically. The reviews have been represented by researchers using the Vector Feature Intensity Graph (VFIG) idea. The primary drawback of this work is the usage of WordNet, yet it nevertheless yields average prediction accuracy of 71% (3 categories) and 46.9%.

III. ALGORITHM FOR SENTIMENT ANALYSIS PROPOSED

The suggested algorithm for sentiment analysis is demonstrated in this section. According to Fig. 2, this proposed method is broken up into three steps.

- Data Filtration
- Training model
- Testing model

Here is a discussion of all stages' specific algorithms. The data filtration flow diagram, training model flow diagram, and testing model flow diagram are all shown in Figures 3 to 5.

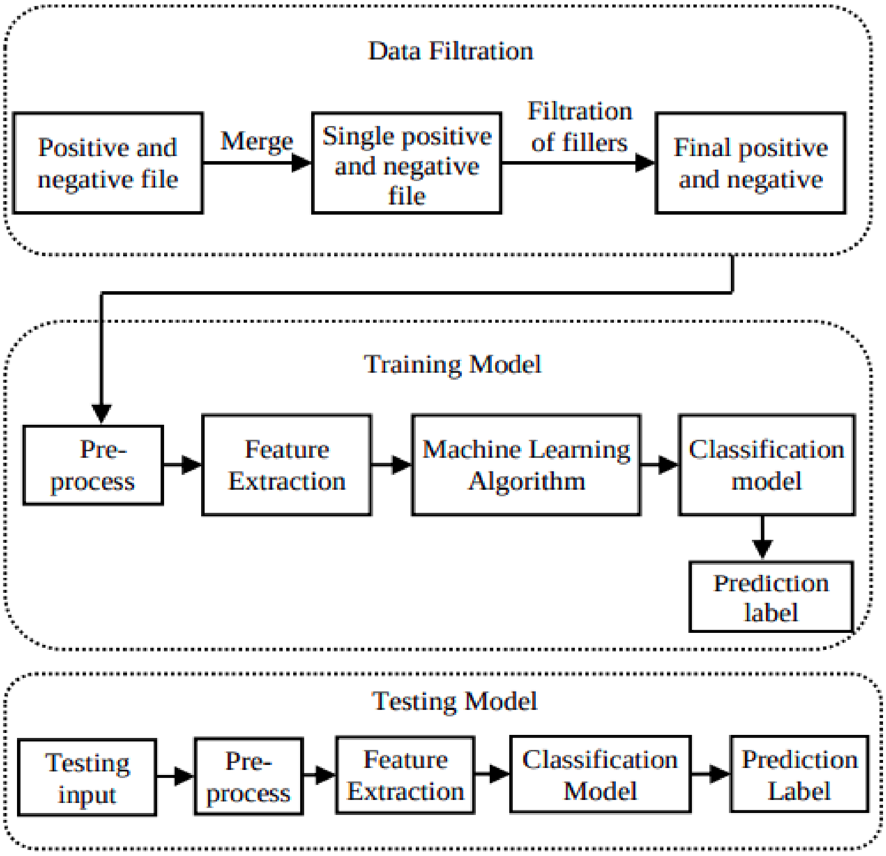

Fig.2. Illustration of the proposed model for sentiment analysis

Data Filtration: Data Filtration importing all positive and negative datasets from file and combining them into a single file. The data sets may contain lots of unwanted symbols, and number. These factors need to be corrected or solved to increase the efficiency. Therefore, in this process the unwanted symbols and number are removed.

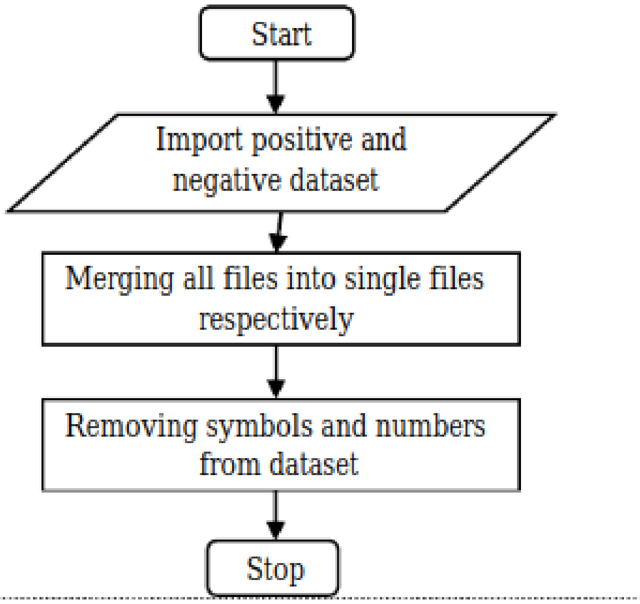

Fig.3 Flow Diagram For performing Data Filteration Algorithm

Training Model: Getting the datasets out of the file and extracting all the feature words, such as adjectives, adverbs, and verbs. Next, datasets are designated as "pos" for positive and "neg" for negative, respectively. Then, run a frequency distribution on the words that were collected, and choose 5000 words for training. Once more, random seed is used to shuffle the data in order to improve training. The labelled datasets in this instance are separated into training and testing percentiles of 70% and 30%, respectively. Dataset for training classification algorithms like the Naive Bayes method [5], the Linear Model algorithm [16], the SVM algorithm, and the Decision tree algorithm.

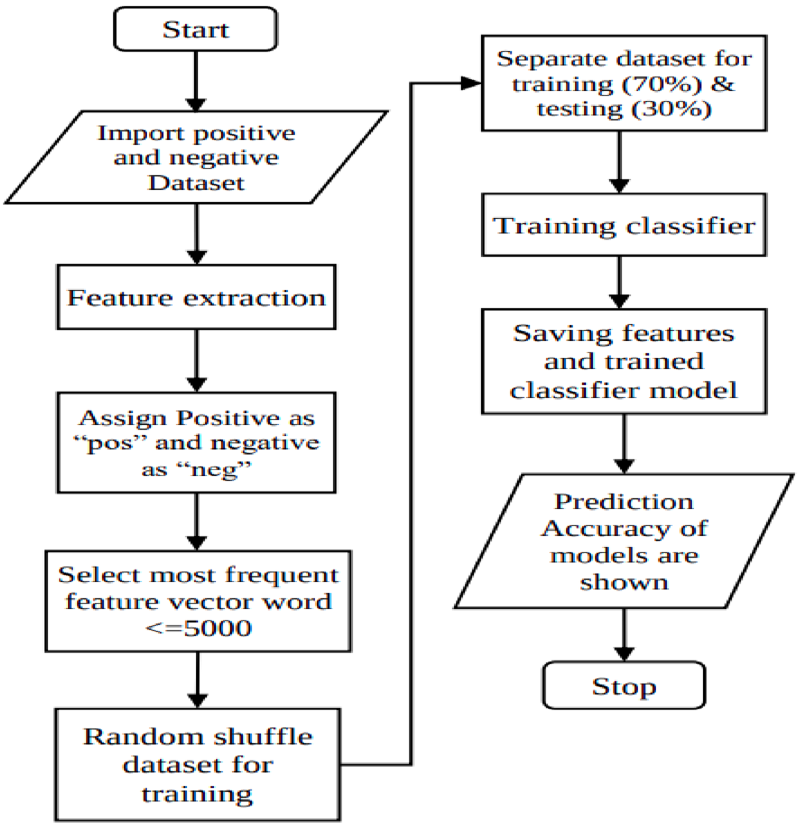

Fig.4. Schematic diagram for implementing machine learning algorithms

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

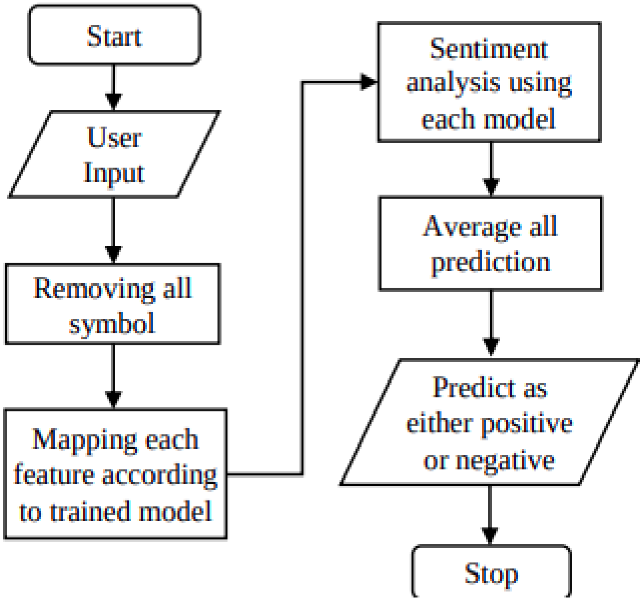

Fig.5. Diagram for testing the proposed model on datasets

Testing Model: Here user can test and analysis the respective model by performing preprocessing over the input data. The preprocessing contains the removal of the symbol and number. Mapping to user input using saved featured (based on training dataset). Then feed to saved model for prediction.

Algorithm 1: Data Filtration Algorithm

Step 1: Importing both positive and negative files and combining them into single file.

Step 2: Removal of punctuations and numbers from the dataset.

Step 3: Output (Filtered data)

Algorithm 2: Algorithm for Machine Learning Implementation

Step 1: Fetching text paragraph from dataset

Step 2: Feature Extraction phase: Extracting words corresponds to adjective, adverb and verb.

Step 3: All the positive sentences are labeled as "pos" and all the negative ones are labeled as "neg".

Step 4: Most frequent feature vector word is set to 5000 words.

Step 5: Random shuffling the dataset for training.

Step 6: Dividing dataset into 70% training and 30% testing dataset.

Step 7: Training dataset to classification algorithms like Naïve Bayes classification algorithm, Linear Model algorithm SVM algorithm

Step 8: Save the outputs of step 2, and step 7.

Step 9: Output (Representation of Accuracy of each model)

Algorithm 3: Proposed algorithm to perform Sentiment Analysis

Step 1: User Input.

Step 2: Preprocessing:

a. Removal of " ' " symbol from the text.

b. Mapping to user input using saved featured (based on training dataset).

Step 3: Feeding Mapped data to different model for sentiment analysis.

Step 4: Output (Averaging all the models).

IV. FEATURE EXTRACTION

The dimensionally reduction process of extracting informative and non-redundant values from a given dataset is called Feature Extraction. The Bag of Words model is used for creation of vocabulary after the cleaning up of the 50,000 reviews from the trained set and the frequency of occurrence of each of these words is calculated.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

The features obtained in this process are used to train the classifier. This action is performed by using sci-kit learns feature extraction module. This module extracts numerical features from the given movie or product reviews which are in text format in the following way:

1) Each string is converted into a unique 'token'.

2) Frequency of occurrence of each of these tokens is calculated.

3) Tokens are organized based on the frequency of occurrences.

With the possibility of obtaining a very large number of features while dealing with 50,000 reviews, one cannot use all the features that are extracted. A certain number of feature vectors need to be selected. Upon testing and experimenting, it was observed that selecting << 5000 or >> 5000 features was resulting in poor prediction accuracy. Therefore, a final array of 50,000 reviews in rows and 5000 features was created.

V. EXPERIMENTAL SETUP

The 50,000 movie reviews from the Stanford dataset, which is openly available, are used to evaluate the proposed sentiment analysis algorithm [15]. The provided dataset consists of 50,000 tagged movie reviews, of which 50% are favourable and 50% are unfavourable. 30% of the dataset is used for testing, and the remaining 70% is used for training. The basic design models shown in Figs. 3, 4, and 5, respectively, make it simple to illustrate this approach.

Dataset: The classifier works with two different types of datasets, the training dataset and the test dataset, in order to learn and predict.

Training Dataset: To categorise a particular test review as a good or negative review, it extracts features from the training dataset and creates a classification logic based on the features.

Test Dataset: The data set utilised to evaluate our algorithm is referred to as the test dataset. Our classifier has to be fed this test set in order to reliably classify reviews as positive ('pos') or negative ('neg').

Format of the Training Dataset: The training dataset consists of 50,000 reviews, with 25,000 of them being positive and 25,000 of them being unfavourable. The training dataset is offered in the layout shown in Fig. 6.

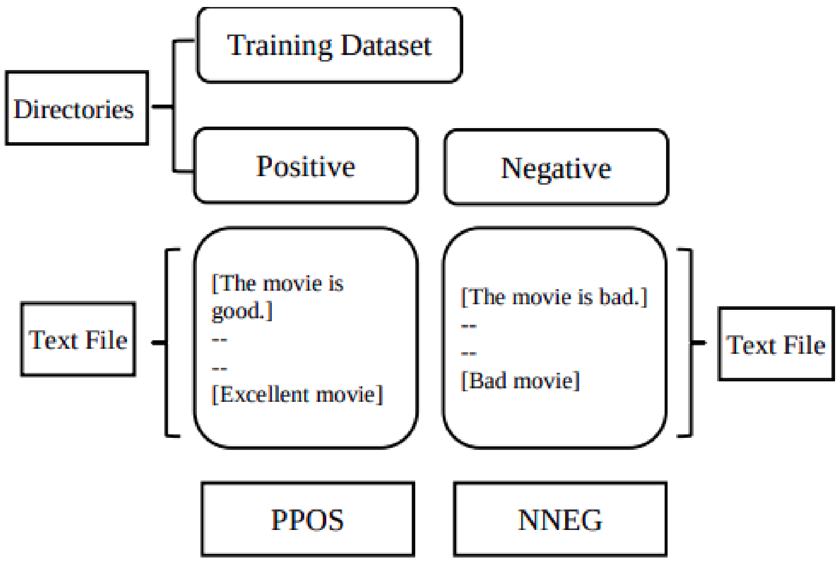

Fig.6. Demonstration of Dataset format used in the proposed algorithm

VI. EXPERIMENTAL RESULTS

Given that every sentence consists of a variety of components of speech, every possible combination will result in a unique accuracy rate. Table.2 and Fig.7 below provide comparisons of the accuracy of these speech segments using various classifier methods. Figures 7, 8, 9, and 10 illustrate numerous combinations of adverbs, adjectives, and verbs as well as various combinations of datasets. Table.3, Table.4, and Fig.7–Fig.10, respectively, demonstrate the execution times of the training and testing datasets together with the accuracy rates of various POS.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

Table.2. Performance Results of different classification models corresponding to different parts of speech (The best case considering each of the speech is styled in bold)

Parts of Speech Considered	Classifier			
	Naïve Bayes	Logistic Regression	Linear SVC	Decision Tree
Adjective	83.81764	**84.29752**	81.92482	82.15142
Verb	80.96507	**81.28499**	78.65902	78.04585
Adverb	79.49880	**81.43161**	79.89869	78.55238
Adjective + Verb	**89.85500**	88.60000	88.75000	87.87500
Adjective + Adverb	**89.85500**	**88.470000**	**88.66000**	83.82500
Verb + Adverb	**89.85500**	**87.255000**	86.95000	86.40500
Adjective + Adverb + Adverb	**89.85500**	**89.575000**	**89.36000**	87.78500

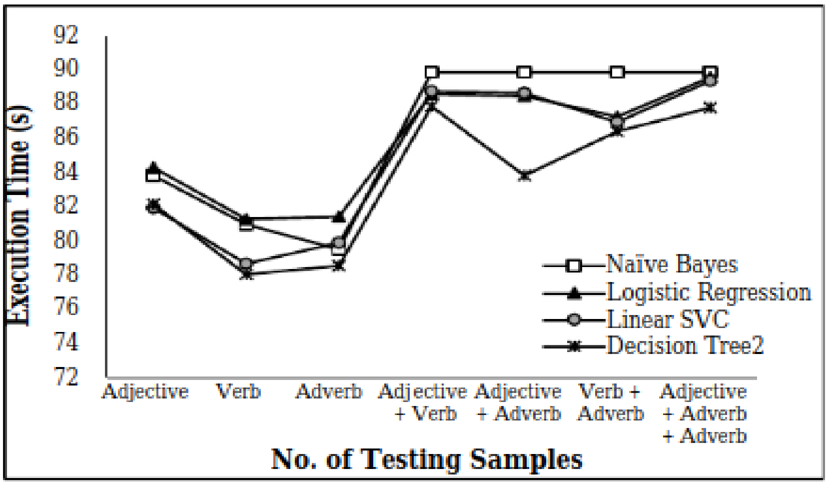

Fig.7. Graphical representation of classification model vs. parts of speech tested on Stanford Dataset

It is evident from the aforementioned Table.2 and Fig.6 that prediction accuracy varies for different POS over the same datasets and classifiers. When permutation is used, Nave Bayes provides an accuracy of 89.855% for the POS combinations adjective-verb, adjective-adverb, verb-adverb, and adjective-verb-adverb, which outperforms other classifiers and results in a straight line. LRC uses some logistic functions to generate prediction accuracy results that are more impressive than those produced by other classifiers, with values of 84.29752%, 81.28499%, 81.43161%, 88.47%, 87.255%, and 89.575% for reviews that contain POS combinations of adjective, verb, adverb, adjective-adverb, verb-adverb, and adjective-verbadverb, respectively. For reviews with adjective-adverb and adjective-verb-adverb combinations, LSVC yields prediction accuracy of 88.66% and 88.36%, respectively. Adjective-verb-adverb combinations perform better than any other adjective-adverb-verb combinations. The new strategy is therefore not only capable enough but also promises to be more effective than the current approaches [2] (86%, on average in Table.1). Figures 7 through 10 show numerous combinations of adverbs, adjectives, and verbs, as well as different combinations of datasets.

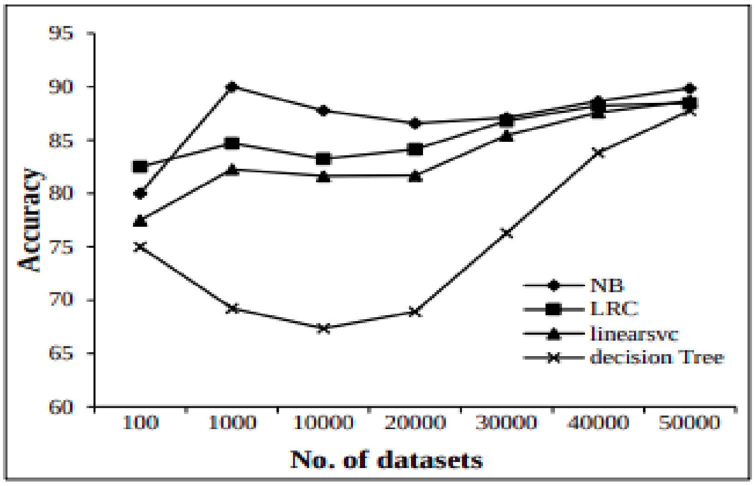

Fig.8. Adjective adverb

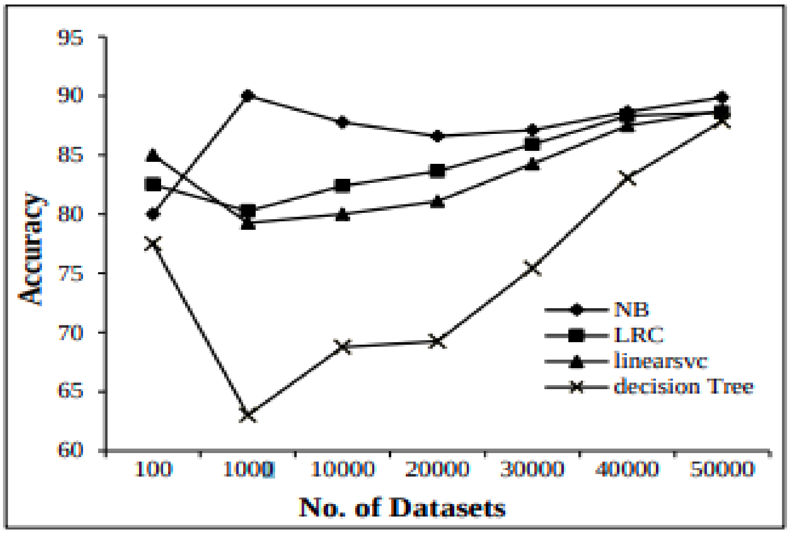

Fig.9. Adjective adverb

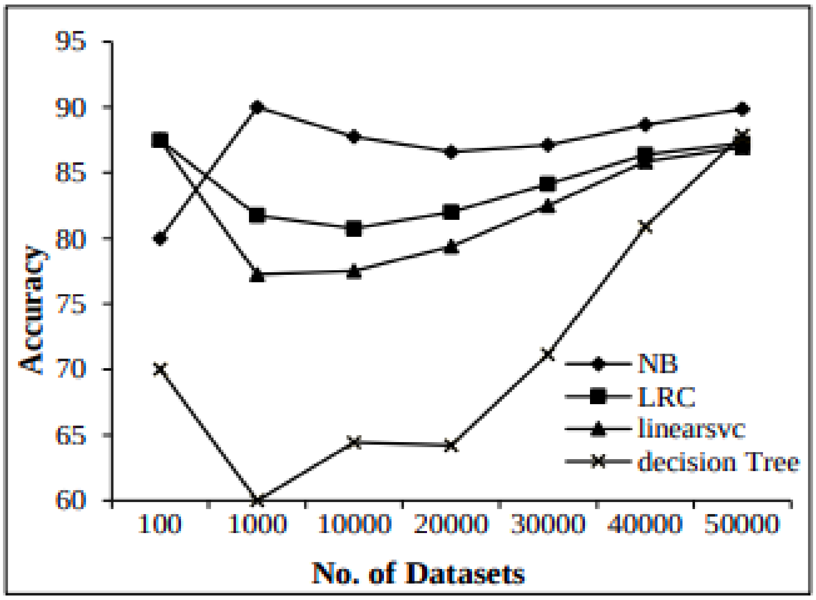

Fig.10. Verb Adverb

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

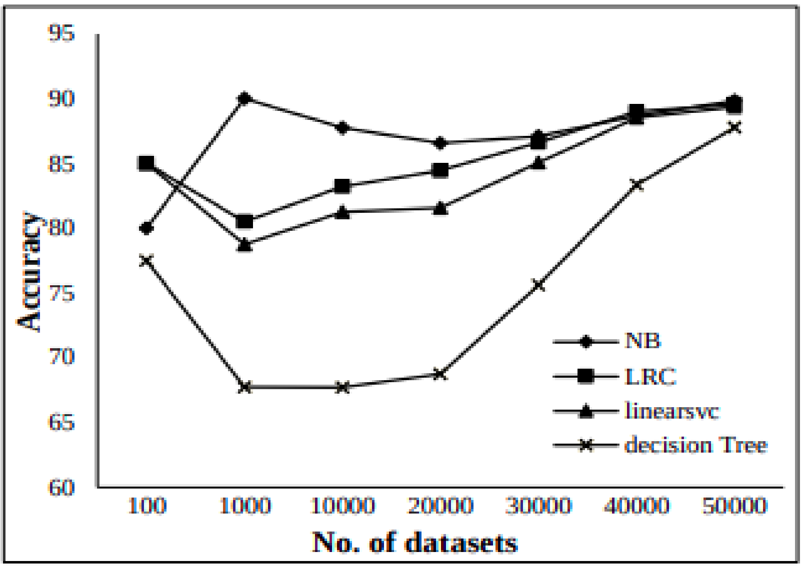

Fig.11. Adjective verb adverb

The Fig. 8–Fig. 11 mentioned above show the accuracy of the different dataset sizes as well as how the accuracy varies with dataset size.

Table.3. Tabular representation of execution time (in seconds) of training dataset corresponding to each classifier.

Representation of Execution Time				
Datasets	Naïve Bayes	Logistic Regression	Linear SVC	D-Tree
50000	26.7274	13.7237	17.4956	50.64934

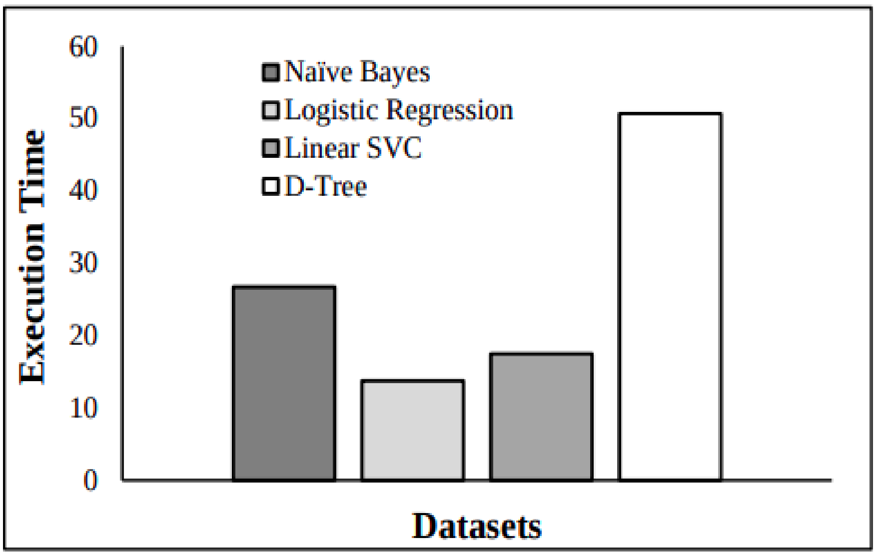

Fig.12. The graphical illustration illustrates the computation time for generating the dataset with several classifiers.

From Fig.12 The graph above illustrates the development time for the dataset for multiple classes across the same datasets.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:04/April-2023 Impact Factor- 7.868 www.irjmets.com

Table.4. Tabular representation of execution time (s) for testing dataset by different classifiers

Datasets	Naïve Bayes	Logistic Regression	Decision Tree	Linear SVC
1	0.0263	0.0698	0.0368	0.0254
2	0.0662	0.0551	0.0556	0.0543
3	0.0807	0.07022	0.1129	0.07233
4	0.1913	0.1188	0.1917	0.0972

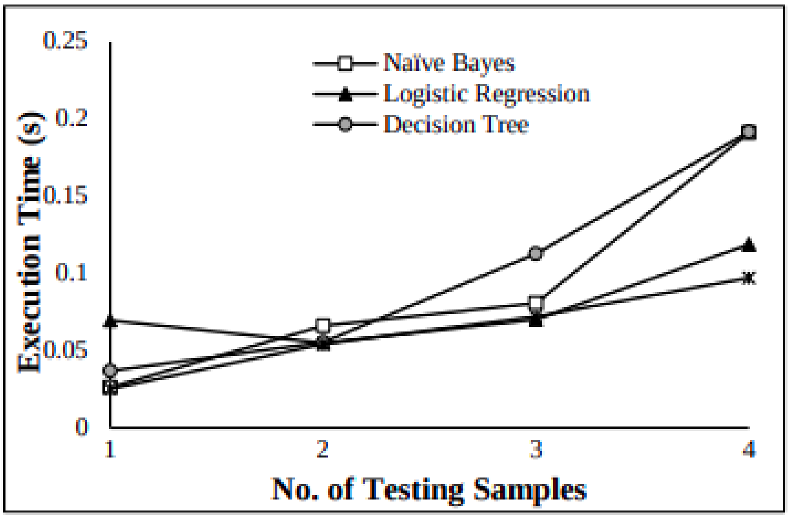

Fig.13. Graphical representation of execution testing dataset w.r.t different classifiers .

Tables 3 and 4 along with Figs. 11 and 12 showing the execution times for training and testing the dataset demonstrate how effective the suggested approach is. LRC requires the least amount of time to train a dataset of 50000, taking only 13.7237 seconds, while D-Tree requires the most time, taking 50.64934 seconds.

The four samples are tested by the linear SVC in roughly 0.0972 seconds, which is substantially faster and thus superior to other classifiers taken into consideration here. This new method promises to train and evaluate opinions more quickly than prior approaches since it improves the way sentiments are analysed with a higher accuracy rate.

VII. CONCLUSION

The Stanford Dataset, a well-known benchmark for natural language processing tasks, was used to evaluate the effectiveness of a newly proposed approach. The evaluation was conducted using six supervised classifiers, and after analyzing various parts of speech, it was observed that the combination of adjectives, adverbs, and verbs yielded the best results. This finding highlights the potential benefits of utilizing these three parts of speech in natural language processing tasks.

VIII. REFERENCES

[1] Reference from paper published on SENTIMENT ANALYSIS FOR PRODUCT REVIEW, by Najma Sultana, Pintu Kumar, Monika Rani Patra, Sourabh Chandra and S.K. Safikul Alam, 4Department of Computer Science and Engineering, Calcutta Institute of Technology, India 5Department of Computer Science and Technology, Raja Ranajit Kishore Government Polytechnic, India.

[2] Sentiment Analysis available at: 'https://ictactjournals.in/paper/IJSC_Vol_9_Iss_3_Paper_4_1913_1919.pdf'

[3] Understanding Sentiment Analysis: https://www.brandwatch.com/blog/understanding

[4] R.A. Hummel and S.W. Zucker, "On the Foundations of Relaxation Labeling Processes", Proceedings of International Conference on Computer Vision: Issues, Problems, Principles, and Paradigms, pp. 585-605, 1987.