International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Handling High Dimensional Word Patterns as Features by Ensemble Learning for Opinion Valuation from Twitter Streams

T. Jhansi Rani [*, 1], K. Anuradha[2]

Submitted: 06/06/2022 **Accepted:** 10/09/2022

Abstract: The stream of over billion tweets are often influence by ambiguity. Due to volume and ambiguity these tweets reflects high dimensionality. The curse of high dimensionality causes more false alarming in detection of sentiment polarity using supervised learning. Though the many of contemporary contributions portrayed novel ensemble classification strategies, limited to handle the volume of data constraints or ambiguity constrained. This manuscript endeavored to portray a novel ensemble classification model that uses fusion of diversified measures to find optimal features, and a novel clustering method fuzzy c-means clustering technique to handle the high dimensionality. The resultant clusters are further used as input training corpus for classification, such that each cluster is used as input training corpus for individual classifier. The experimental study has carried by multi label four fold cross validation. In order to scale the performance, the results obtained for cross validation metrics for proposed model titled "ELOV" and the contemporary contributions of ensemble models. The performance analysis projecting that the proposed model is outperforming the contemporary contributions.

Keywords: *Feature Optimization, Machine-Learning, KS-Test, Term-Occurrence, Naïve Bayes, Wilcoxon signed-rank, Fuzzy-c Means, Handling Dimensionality.*

1. Introduction

According to the Technorati firm, which is a social-media tracking organization, four out of every five internet users utilise social media in some form. It includes a microblogging site, a video-sharing site, a friendship network, and many more features. It has been observed that the World Wide Web has now entirely transformed into a corrected and more interactive or communicative web. It allows a large number of people to participate in a variety of ways. In truth, even folks who are new to web publishing processes are creating web content. However, the website's worth has been heavily influenced by the user, who determines the availability of data in it. Several websites allow users to express their own written ideas, beliefs, or experiences about a service or product as a review. The internet has been flooded with reviews or comments on a variety of goods such as hotel services, mobile phones, movie reviews, and many more. Furthermore, it is inspiring that these thoughts or reviews are not just about consumers' perspectives, but also a respected source.

[1]*Assistant Professor, CSE, Dept. of GITAM Deemed to be University, Hyderabad & Research Scholar, JNTU, Hyderabad, India.*
Email: jhansi.rani.t@gmail.com
[2] *Professor, CSE, GRIET, Hyderabad, India.*

** Corresponding Author Email jhansi.rani.t@gmail.com*

For example, if a user wants to book a hotel in any city [1], they may want to know the hotel review before purchasing a hotel. In addition, when a customer decides to acquire a certain type of digital camera, the purchaser reads the reviews provided by other users about the camera's characteristics and performance. It will aid consumers in obtaining additional and relevant information about the various items with a single click of the mouse. However, it also aids in reaching a more valuable conclusion. Instead of writing a review, consumers may want to write about their experience with a service or product in a blog. Nonetheless, in both cases, the data has been taken literally.

Some well-known sites, such as imdb.com and Carwale.com, are brimming with user evaluations of movies and automobiles. Users that provide reviews on this site come from a variety of backgrounds, including those who have recently purchased a product and those who have had a service experience. If we look at www.imb.com, an Internet Movie Database website, we will see some valuable material for those who are interested or attracted by movies. This website provides information on the release or development of a product in any area of the universe. Similarly, submitting reviews on blog sites displays a large number of user viewpoints. Despite the fact that blog posting is a complicated source for emotion analysis. It includes specific phrases that might be utilised to elicit emotion. The majority of the time, it contains connected or factual information that has not been evaluated or taken to the needed breadth. Nonetheless, we have an amazing supply of user ideas, views, or comments that should be leveraged for dynamic analysis or other beneficial studies.

As previously said, in this day and age, the majority of individuals use social networking websites to communicate or convey their ideas, feelings, or opinions. Furthermore, the opinions of others will always be influential. There are over 1.5 billion registered accounts on social websites, and billions of data, photographs, videos, or messages have been shared and sent. Social media has created a massive quantity of data.

Twitter is widely regarded as the leading social media website in the world, ranking fourth overall. In March of 2006, Jack Dorsey and his colleagues at the University of New York founded Twitter. Twitter's headquarters are in San Francisco, California, USA. Furthermore, Twitter has 500 million registered users, with 336 million active users.

Twitter has been dubbed a "miniaturised scale blogging destination," since it has evolved into a repository for a wide range of material. It is because the Twitter environment offers little write-ups, where people would regularly submit their comments or evaluations of things, or express their perspectives on current topics, and many more in daily life, as described in [2]. Opinions or reviews, on the other hand, might be favourable or negative. In this regard, sentiment analysis was developed to categorise tweets into negative or positive categories using machine-learning (ML) algorithms [3]. When there are both good and bad tweets, the more significant one should be considered. Tweets, in general, will include the usernames, hashtags, and emojis required for converting and processing into a definite structure. Furthermore, it must have extraction features such as unigrams and bigrams. We cannot rely on a single technique since it is ineffective. As a result, the most accurate strategy must be chosen [4]. In general, corporations research client responses and respond to clients using small-scale online end-points or destinations. Creating an invention to identify and alleviate the casual mood has been regarded as a serious issue. Many people are currently interacting via Instagram, Facebook, and Twitter [5]. Most of them use social media to express their thoughts and feelings about personalities, objects, or locations. For analysis, lexicon-based, AI, and hybrid methodologies were applied.

A quick procedure for extracting data from public or accessible information on interpersonal groups has been established. As mentioned in [6], investigating informal organizations such as Twitter might improve sentiment analysis precision and forecasts. Furthermore, Twitter has been regarded as one of the most important open sources for obtaining product knowledge. As a result, the focus of this work is on classifying a large number of tweets to determine which include negative feelings and which contain positive sentiments using a variety of ML techniques. In general, emotion mining has been done in three tiers, as detailed below:

- Document-level: In the case of document-level, the record can be organized completely as “neutral”, “negative,” or “positive” [6].
- Sentence-level: In the case of sentence-level, every sentence has been labeled as “fair-minded”, “positive,” or “negative”.

Feature and aspect level: In this level, records or sentences have been categorized into “non-fanatic”, “positive,” or “negative,” considering the substantial parts of documents or sentences and usually called to be perspective level evaluation clustering as stated in [6].

2. Related Work

Sentiment analysis based on the unsupervised document has been proposed in [7] to determine the text documents’ sentiment orientation depending on their polarities. The documents have been classified as negative and positive [7], [8], and sentiment words need to be extracted from the collections of documents and categorized as per polarities. Moreover, techniques based on an unsupervised dictionary have been used to identify negation. WordNet has been used to determine the expressions of an opinion, antonyms, and their synonyms [7]. Movie reviews have been gathered to use as input such that polarity sentiment related to documents has been detected. Now, the system or approach has categorized each of them in the form of impartial, positive as well as negative and generated the summary of the output, representing the overall amount of nonpartisan, negative as well as positive documents. Hence, the system has generated the summary report, which assisted decision-makers. The higher amount of opinion vocabularies present in the document states the sentiment polarity of that particular document.

Sentiment extraction at document level [9] has been proposed and focused on 3 phases. In the primary phase, the dataset comprises documents comprising opinions mined from the internet automatically. Next, extracts the adjectives of sentiment polarity negative and positive from the given training corpus. Finally, novel document test sets have been categorized depending on collected adjective sets in the second phase. Several simulations or empirical studies have been carried out on real data, and the model proposed in [9] attained 0.717 and 0.622 of F1-score to attain the positive and negative documents, respectively.

The work in [10] addressed the sentiment reviews classification issue regarding the products written or stated in the Chinese language. Their model has been dependent on unsupervised classification capable of teaching itself by enhancing the seed of vocabulary. Primarily, it included the single-word, which was labeled to be positive. Furthermore, a primary seed has been retrained iteratively for the classification of sentiment. The criterion of opinion density has been later used for measuring the sentiments ratio of the document. Simulations exhibited that the trained classifier achieved 87% of the F1-score to detect sentiment polarity after 20 iterations.

The contribution [11] endeavored to categorize the reviews as per their polarity by utilizing supervised-learning algorithms like Random-forest, SVM, linear discriminant-analysis, and NB (Naïve Bayes). To attain this, the projected model incorporated four steps. The initial step is a preprocessing step conducted for eradicating the special characters, stop-words and numeric. Secondly, reviews of text have been translated into numeric-matrix.

In the third step, vectors generated have been used as inputs for four diversified classifiers. Moreover, the outcomes have been attained subsequently by the two datasets classification. Later, several metrics like classification accuracy, f-measure, recall, and precision were calculated to measure the projected model performance. For the IMDB datasets and polarity, one of the classifiers, random forest, has performed better when compared to other types of classifiers.

The work [12] implemented the SVM to 3 diversified datasets for categorizing the reviews of the document. Various n-gram strategies have been used for measuring the SVM impact in documents classification. Furthermore, three weighting models have been used by the researchers for generating the feature-vectors called TFIDF (Term-frequencies and inverse of document frequencies), Term-Occurrence, and Binary Occurrence. Various simulations have been carried out for evaluating the probable numerous n-grams and weighting methods together. For dataset Taboada, the effective minimal false alarming has been attained by integrating SVM with trigram and TFIDF. For pang corpus, effective outcomes have been attained by utilizing trigram and binary occurrence. Exhibited that among the many of contemporary classifiers, the SVM attained the minimal false alarming while integrated with bigram and TFIDF [12]. The work [13] reviewed existing processes of opinion mining like vocabulary-based techniques and ML techniques. By using diversified ML algorithms such as SVM, NB, and Max Entropy, additionally, the work [13] explained the general complexities and used Twitter sentiment analysis.

The work [14] proposed a model for tweets sentiment classification. The concept behind it is cumulative feedback automatically. Moreover, the sentiment issue has been treated as binary classification, categorized tweets into negative and positive. The training data comprises tweets with emotions gathered depending on the supervision model, which has been projected by [15]. The work [14] used Twitter API for tweets extraction, which incorporated emotions.

Moreover, these have been utilized for recognizing tweets as either positive or negative. Repeated tweets and retweeted posts have been eradicated. Moreover, tweets comprising negative and positive emotions have been filtered. Several classifiers like SVM, NB, and MaxEnt have been used for tweets classification. Diversified features have been extracted like Bigrams, Unigrams, and unigrams with POS and unigrams with bigrams. Moreover, the effective outcomes have been attained by classifier MaxEnt in conjunction with bigrams and unigrams features that attained 83% accuracy compared with the NB classifier and attained 82.7% accuracy.

The work [16] projected a supervised model to classify tweets extracted from twitter trends, which denotes that the SVM is having more decision accuracy. Moreover, simulation tried to integrate PCA (principal component analysis) and SVM to lower the dimensionality of the feature. Also, hybrid, bigram, and unigram feature extraction models have been used. It has been exhibited that PCA integration with SVM with Hybrid feature selection might assist in lowering the dimensions of feature, and outcomes attained for classification accuracy is 92%.

The work [17] proposed a model for sentiment polarity extraction from Twitter data. In this, extracted features are words comprising of emoticons and n-grams. The empirical study exhibited that the performance of SVM is better when compared to NB (Naïve Bayes). SVM has been considered an effective model with a mix of unigram feature-extraction is attaining an 81% of precision accuracy and 74% of recall accuracy.

The work [18] devised an architecture known as opinion miner, which automatically examined and identified the social media message's sentiment. The tweets annotated have been integrated for analysis, and these messages of framework that comprises of feelings have been extracted along with their determined polarities. To attain this, the sentiment polarity labels neutral, decisive have been identified for each tweet given for sentiment polarity assessment [18] The work [19] used emoticons and Twitter API for gathering the positive as well as negative labeled sentiments, same as stated in [14]. The analysis of sentiment of given tweets has been considered to be multiple labels listed as positive, negative, and neutral. Moreover, statistical linguistic analysis has been done on gathered training data corpus using frequency of terms (words). The training corpus has been used for building the classifier, and simulations have been carried out on MNB (multinomial NB), CRF (Conditional random fields), and SVM, which have trained by varied features optimization strategies. The parts-of-speech (PoS tags) and n-grams used to train the MNB generated effective performance in simulations. The work [20] proposed a combination of semantic with unigram as well as PoS features. The semantic features were the concepts, which summarize the entities minded from the Twitter data. Also, features extracted have been utilized for measuring the relation of entity sets enhanced through respective sentiment polarity.

Moreover, it is prominent that including semantic features would improve the accuracy in tweet's sentiment polarity assessment. The work [20] used tweets of three training corpuses to extract semantic features. Furthermore, the experiments denoting that the classifier NB has been utilized along with semantic features extracted. The outcomes exhibited that semantic features result in enhancements in identifying sentiments compared to PoS and unigram features. However, for OMPD and HCR datasets, the sentiment-topic model performed more effectively than a semantic model. For HCR, the contemporary attained 68.15 of F1 score compared to 66.10 of F1score attained by semantic model. In terms of the OMD dataset, 72.80 of F1 score has been attained by utilizing sentiment-topic model when compared to 77.85 of F1 score attained by semantic model.

Diversified types of features have been extracted [21] to increase the sentiment classification accuracy. The unigram features have been proposed as a baseline, while the words have been taken as independent features. The features of domain-specific have also been incorporated, like the number of retweets. To mine the concepts in tweets, DB media has been used for mining, and these features are called DBpedia features. The term WordNet has been used for recognizing the synonyms of adjectives, adverbs, verbs, and nouns. SentiWordNet has been used for measuring the words frequencies of negative and positive labeled tweets respectively. The empirical study exhibited that accumulating adjectives, DBpedia, and SentiWordNet feature results in minor changes or enhancements inaccuracy of NB and SVM classifiers. The ratio of small enhancements is nearly 4% and 2% for NB and SVM in respective order.

The work [22] engaged feature selection based on chi-square and gained metrics for selecting the informative features after lemmatization and steaming processes. Moreover, simulations that involve the feature selection metrics by an SVM classifier result in an enhancement over the earlier contributions. Additionally, the work [23] examined the information gain impact as a feature selection criterion for ranking the semantic and unigrams features. It has been concluded that classifier performance can be approved even when choosing few diversified sentiment-topic features by utilizing information gain.

In contribution [24], ensemble classifiers with several Twitter sentiment models have been used to increase the effectiveness and performance of categorizing the polarity of the tweet. Their approach has integrated skip-gram scorer, linguistic resources-based model, ranking algorithm, and Word2Vec. It has been more prominent to focus that their projected ensemble model is dependent on voting techniques. To assess the projected model, the TASS competition training data have been selected. The outcomes from the simulation study exhibited that slight enhancement has been attained with the ensemble model compared to skip-gram and ranking algorithm models. Macro-F1score attained earlier was 62.8% compared to 61.60% of macro F1 score attained by the last combination.

The research work in [25] exhibited that the ensemble approach might generate better accuracy of emotion classifiers compared with single-classifiers. Also, they have integrated the lexicon and BOW features in terms of ensemble classification and carried out simulations exhibiting that whenever the extracted features have been utilized in integration or mix with these features, there is an enhancement in the classification accuracy. The combination of stacking models, SVM and SentiStrength, by using majority voting generated 86.05% of the F-score, considered the highest or maximum score.

The work [26] projected a system for categorizing the tweets depending on majority voting of 3 classifiers such as LR, SVM, and NB. The gathered tweets have been divided into two sets: testing and training sets. The individual classifiers have received a similar training set for recording their decisions. The final decision has been generated by the ensemble model depending on the majority of the highest number of votes gathered from classifiers. Information gain has been considered a significant aspect of this contribution to lowering the feature vectors' dimensionality. An empirical study has been conducted to investigate the information gain impact on classifier accuracy. The outcomes envisioned the enhancements in classification accuracy after dimensionality of feature vector has been lowered by utilizing information-gain. Therefore, information gains clearly showed the enhancements in the accuracy of classification in the overall dataset.

As per [27], collecting a huge amount of unlabeled data from the social networks was a simple task; nevertheless, identifying these sentiment labels is expensive. Hence, it was required to use models related to unsupervised methods of sentiment polarity assessment. It is obvious to notice the potentiality of the unsupervised models due to the phenomenal increase of unlabeled opinion data of social media. Besides, they anticipated word-level sentiment polarity indicators for identifying the post's polarity and to get the polarity of the world nearer to emotional indicators at the word level. In the simulation study, OMD and STS (Stanford twitter sentiment) datasets were used. The framework of ESSA attained 0.726, 0.692 of accuracy for corpus of tweets STS and the OMD in respective order. The outcomes exhibited the advantages of ESSA when compared with other strategies.

The work [28] presented the real-time framework for identifying the sentiment polarity of the given tweet's corpus. Besides, unsupervised strategy has used for exploring the tweets and identifying the polarity. Moreover, this classification model utilized an approach based on the dictionary for identifying the tweeted opinions polarity and their framework [28] comprised of several modules. Tweets were collected by utilizing the tweets acquisition module, which was linked to Twitter API for retrieving the tweets utilizing queries. Besides, text was tokenized by utilizing different modules. Later, token standardization, verbal correction as well as syntactic correctness are the several phases in the module of tweet processing. The opinion analysis approach has been introduced by various researchers for computing sentiment polarity of words, emoticons, and the mean value of the sentiment polarity. In this, simulations have been carried out depending on a dataset called SeMEval for measuring the framework quality. For dataset SemEval-2013, the projected system attained 0.559 accuracies compared to the SSA-UO system, whose accuracy is 0.50 [29]. The framework proposed in contribution [28] attained 0.533 of accuracy when compared with decision accuracy 0.539 exhibited for the tweet's corpus SemEval-2016 by the experiments carried by research group of GTI.

The work [30] exploited a lexicon-based model for predicting the emotional intensity level to make estimations. Their model was suitable for identifying subjective texts stating their opinion and for the classification of sentiment polarity to predict the sentiment polarity is negative or positive. The proposed lexicon-based approach attained an F1score of 86.5, 80.6 & 76.2 for Twitter, Myspace as well as Digg datasets, where these are outperforming when compared to all other supervised classifiers.

The work in [31] implemented a system based on vocabulary for the classification of sentiment that classified tweets as negative, unbiased, or positive. Moreover, this system has been discriminated against and ranked slang used in the tweets. In this, the empirical outcomes exhibited that the projected architecture performed better when compared to contemporary architectures, attaining a precision of 92% in double-characterization. In the case of multi-class clustering, it attained 87% of precision. The architecture has been required for improving the accuracy in scenarios of negative cases and reviewing unbiased cases. The work [32] projected an enhanced sentiment classification based on a lexicon that included a rule-based classifier. Moreover, the objective is to lower the data sparseness and enhance the sentiment classification accuracy.

The other contemporary contributions, "Use of Novel Ensemble Machine Learning Approach for Social Media Sentiment Analysis (EMLA) [33]." and "Twitter sentiment analysis using an ensemble majority vote classifier (EMVC) [34]." have endeavored to perform opinion valuation using the ensemble classification approach. The earlier contribution [34] has fused the multiple classifiers, which word patterns have trained. The latter one [33] of these two contributions has proposed a weighted majority rule ensemble classifier, which has also ensembled diversified classifiers like the earlier one. In addition, to address the curse of high dimensionality, diversified standard features have been collected in the context of each classifier used in the ensemble model. However, these two contemporary contributions evincing the above 90% accuracy, the high dimensionality of the features haven't been addressed. Moreover, usage of multiple classifiers is the primary factor of these models, which often termed hybridization of the multiple classifiers to address the high volume of training corpus.

Concerning to address the constraints noticed in these contemporary models, this manuscript portrayed a novel ensemble classification model, which trained by the optimal word patterns selected by fusion of diversity assessment methods to handle the curse of dimensionality.

3. Methods and Materials

It is necessary to transform the collection of tweets with different labels representing the sentiment polarity, including "not a review," into a pattern of words sequence as a record. Each resultant record has a final column that corresponds with an emotion polarity label. Additionally, tagged data with a certain sentiment polarity have been separated into several groups. In order to deal with the curse of dimensionality, each opinion record is split up into many clusters. In addition, the labels associated with these clusters have been used to train an algorithm that optimizes pattern of words sequence as features. Each record that belongs to a cluster and the resulting column of records with contrasting opinions must undergo the fusion of distribution diversity assessment measures. The resultant column must be evaluated to determine if it is optimal for the resulting cluster of the records based on the observed diversity. Word patterns have also been constructed from optimal cluster features to be utilised in training the random-forest classifier [35], [36]. Fig. 1 is a visual representation of the block diagram.

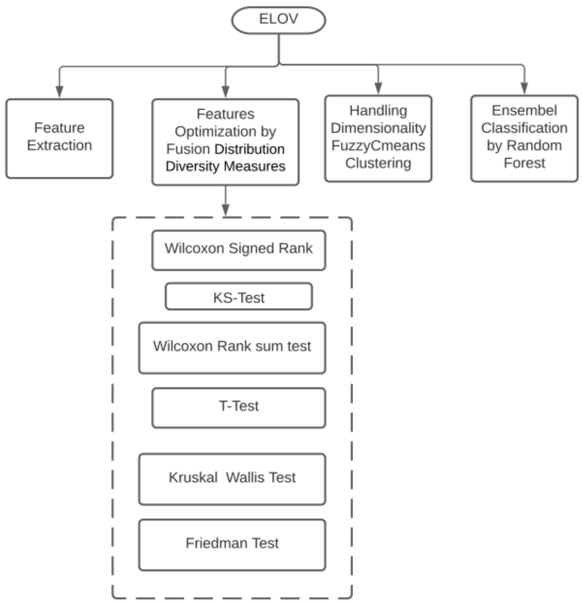

Figure 1- The block diagram represented ELOV

3.1 Distribution Diversity Measures

In this section, diversified distribution diversity evaluation models have been fused for performing optimum fusion feature selection. The models such as KS-test [37], Dual tailed t-test [38], Friedman test [39], Wilcoxon-rank-sum test [40], Wilcoxon signed-rank (WSR) [41] and Kruskal-Wallis-test [42] have been used for performing fusion of different evaluation measures for optimum features selection.

3.1.1 Wilcoxon signed-rank

The WSR-test [43], [44], [45] is superior than the sign test for testing one-sample as well as two-sample paired data because it takes into account the magnitudes as well as the signs of the differences.

The gist of the WSR-test is as follows:

Select an assurance level for both the null and alternative hypotheses. The median population difference between the matched data is taken as zero under the null hypothesis. If it isn't, that's the counterfactual.

Finally, the test statistic must be calculated. To do this, we will first calculate the disparities among the paired data samples, then rank the differences based on their magnitude alone (i.e., disregarding their sign), and then sum the rankings of the positive as well as negative disparities respectively. At last, the test statistic will be determined by picking the smallest of these positive and negative rank totals.

Next, the test statistic is compared to a predetermined threshold value. The lack of variation across samples can be inferred whereas if test result is smaller than the cutoff value.

3.1.2 KS-Test

To check if two samples are drawn from the same distribution, statisticians use the two-sample Kolmogorov-Smirnov test [37].

Let's assume that Sample $\boldsymbol{A}$ has a size of m and its cumulative distribution function $\boldsymbol{F(x)}$ is observed, whereas Sample $\boldsymbol{B}$ has a size of n and its cumulative distribution function $\boldsymbol{G(x)}$ is observed.

$$D(m,n) = \max\left(\left|F(x) - G(x)\right|, x\right)$$

The expression $D(m,n,\alpha)$ denotes the critical value, which is estimated by,

$$D(m,n,\alpha) = c(\alpha) * \sqrt{(m+n)*(m*n)^{-1}}$$

$if\left(d\left(m,n\right) > d\left(m,n,\alpha\right)\right)$ This condition denotes that diversity between two samples is true that the data are not normally distributed, hence returns 1, and otherwise returns 0.

The inverse Kolmogorov distribution at α (denoted by $c\left(\alpha\right)$) can be computed using the formula

$$D(m,n,\alpha) = KINV(\alpha) * SQRT((m+n)/(m*n))$$

In the Kolmogorov Distribution, where KINV is specified. In the last columns of the Kolmogorov-Smirnov Table, the values of $c\left(\alpha\right)$ are the numerators.

3.1.3 Wilcoxon rank-sum test

The well-recognized non-parametric test for comparing the results among two independent sets is the Wilcoxon rank-sum test (WRST) [40]. Sometimes, this WRST is also known as the Wilcoxon Rank Sum test, which has been utilized for testing whether two samples have been derived from a similar population. Some of the examiners understand this test by comparing medians among two populations. Also, recall, which is a parametric test used for comparing the means among independent sets.

On the other hand, the following is how the two-sided as well as null research hypothesis for a non-parametric test has been reported: For the same reason that "returns 0," the two distributions cannot be considered "different.", else returns 1.

This is a two-sided test, and the null hypothesis indicates that sample distributions are not identical to those tested in the opposite way. When studying a population, the onesided research hypothesis is used if the study's focus hinges on the growth or decline of one subset.

In addition, the test technique combines data from two independent samples into a single data set while retaining track of the original data source for each observation. Arrange them in a subsequent time in increasing order from 1 to $(n1+n2)$.

$wrs\_test(v_1, v_2)$ // Begin

Sort both vectors v_1, v_2 in ascending order of the values.

$$U(v_1) = 0$$

$$U(v_2) = 0$$

$\forall_{i=1}^{|v_1|} \{e_i \exists e_i \in v_1\}$ Begin //For each element of the vector v_1

$\forall_{j=1}^{|v_2|} \{e_j \exists e_j \in v_2\}$ Begin //For each element of the vector v_2

$$\begin{Bmatrix} (U(v_1) += 1) \exists (e_j < e_i) \\ (U(v_2) += 1) \exists (e_i < e_j) \end{Bmatrix}$$

End

End

$$U = \begin{Bmatrix} U(v_1) \exists (U(v_1) < U(v_2)) \\ U(v_2) \exists (U(v_2) < U(v_1)) \end{Bmatrix}$$

Find the d-critic of the given vectors v_1, v_2 having sizes $|v_1|, |v_2|$ from the U-table at a given distance threshold $d\tau(0.01, 0.05, or\ 0.1)$

Given two vectors $v1\, and\, v2$ of sizes $|v1|\, and\, |v2|$, determine the d-critic using U-table at the specified distance threshold $d\tau(0.1, 0.05, or\ 0.01)$.

$$return \begin{Bmatrix} 1 \exists (U > dc) \\ 0 \end{Bmatrix}$$

End

3.1.4 Friedman Test

As in [39], Friedman's test is a non-parametric test for detecting trends over several iterations of a treatment. In addition, there is a test that does not assume any particular data distribution. When the data distribution is unknown, it is often used instead of the ANOVA test. An extension of the sign-test used when there are many treatments, the Friedman test is widely used.

In the chosen data $\{x_{ij}\}_{n\times k}$, constituting matrix combination of blocks being represented by n rows, and treatments being depicted as k columns, the estimation of the ranks within each block is assessed using the single observation pertaining to intersection of treatment and block respectively. In the instances of values in tie, the respective ranks shall be averaged and assigned accordingly. The data in the subsequent new matrix shall be replaced as matrix $\{r_{ij}\}_{n\times k}$ where the entry r_{ij} is the rank of x_{ij} within block i.

Followed by the identification of the values are carried out based on $\bar{r}_{.j} = \frac{1}{n}\sum_{i=1}^{n} r_{ij}$

Test statistic is emphasized using the Eq 3

$$Q = \frac{12n}{k(k+1)} \sum_{j=1}^{k} \left(\bar{r}_{.j} - \frac{k+1}{2} \right)^2 \quad \text{... (Eq 1)}$$

It has been noted that Q values need not essentially be adjusted to the tied values over the data.

As a resulting outcome of the earlier steps, if n or k is large as depicted in parenthesis $(i.e.,\ n > 15\ or\ k > 4)$, probability distribution of Q can be approximated using chi-squared distribution. In the case wherein p-value is based on $P\left(\chi^2_{k-1} \geq Q\right)$. In cases when n or k are tiny, the chi-square approximation becomes inaccurate, and the p-value must be calculated using Friedman test tables instead. If p is especially salient, then many contrastive analyses have been conducted.

3.2 Preprocessing

The twitter corpus T should be cleaned up by removing records that don't have a label. Furthermore, each record $\{r \exists r \in T\}$ will be broken into sentences, with each phrase labelled with the parent record's sentiment polarity. Therefore, a set S of sentences have been generated. The preprocessing phase then splits each sentence $\{s \exists s \in S\}$ of the tweet $\{r \exists r \in T\}$ into set of words listed as vector vt_s. Remove the stop words from each of resultant vector, followed by the "ing" and "ed" versions of the remaining words. Each sentence's words vector vt_s has collected it into a set VT.

3.3 Handling Dimensionality

Reduce overfitting dimensionality with fuzzy c-means, according to [44], [45], [46]. The input data is divided into many tuples using the fuzzy C model so that each tuple contains associated records (less variability or dimensionality). The FC-Means can partition each record.

By giving each point a 0 to 100 percent membership in each cluster centre, fuzzy logic groups multidimensional data. Compared to hard-threshold clustering, where each point is labelled, this is more effective. Based on distance, this technique determines the membership of data points that correspond to cluster centres. Data is more likely to belong there if it is closer to the cluster centre. Each data point's total population should equal one.

We can segment fuzzy data using unsupervised clustering. The parameter m controls the fuzziness of the algorithm. The classes are blurred by high m values, and all items belong to every cluster. The results of optimization depend on m. Different m decisions lead to various partitions.

The fuzzy c-means process flow is listed below:

- Assume that there are c fixed clusters.
- Initialization: Determine the likelihood that each data point dp belongs to a certain cluster c, P (point dp has label $c\,|\,dp\,|$, c).
- Recalculate the cluster centroid as the weighted-centroid using the probability of membership for each data point in the iteration

$$\mu_c(m+1) = \left(\sum_{i=1}^{c} dp_i * P(\mu_c \mid dp_i)^b\right) * \left(\sum_{i=1}^{c} P(\mu_c \mid dp_i)^b\right)^{-1}$$

Continue iterating until convergence or up to the amount of iterations that the user specifies.

3.4 Feature Optimization

It has been assumed that a significant portion of the predicted model would be devoted to perfecting the model's features. Since set C contains the word pattern records to be used in the training phase, set $C = \{c_{angr}, c_{dsg}, c_{sad}, c_{joy}, c_{neut}, c_{sur}, c_{fear}\}$ must be separated so that each set represents the pattern of word sequences as records for one sentiment polarity (positive $(angr)$, negative (dsg), neutral (sad), as well as not-a-review (joy)). Each entry's salient aspects are the word configurations of the related tweet on Twitter. Further, FC-Means is used in the optimization procedure to create several clusters from the original $C = \{c_{angr}, c_{dsg}, c_{sad}, c_{joy}\}$ sets of records (see section 3.3). Each sentiment polarity's resulting cluster contains records that only appear in that cluster and n number of records (where $0 <= n$) from the other clusters, as depicted in the following mathematical notation.

$$\left\{\begin{array}{l} Cl_{em} = \{em_1, em_2, ..em_i, em_{i+1}, ..., em_s\} \exists \\ em \in \{angr, dsg, sad, joy\} \wedge \\ |\, em_i \cap em_j \,| \geq 0 \wedge i \neq j \end{array}\right\} \quad \text{...(Eq 1)}$$

The Eq 8 key phase of feature optimization is finding each cluster's optimal features of all sentiment polarity labels perform as follows.

One of the two dimensional matrix corresponds to each emotion polarity label l in the set $\{angr, dsg, sad, joy\}$, while another corresponds to each of the FC-mean clusters $\{em_i \exists i = 1, 2, ..., s\}$ of the opinion polarities l.

Each row em_i in the matrix represents the word patterns of varying sizes l that make up the characteristics of a tweet with a certain emotional valence. Each column in the related matrix displays the predicted word-patterns $vf_a = \{v_1, v_2, ..., v_{|c|}\}$ of a feature variable $\{f_a \exists a = 1, 2, ..., |r|\}$ over all of the records. Here, $|r|, |c|$ denotes the total rows as well as columns in the related matrix.

In order to determine which opinion polarity label cluster em_i is most accurate, a feature-optimization method selects the most relevant feature-attribute f_a. If the given diversity threshold $d\tau$ is greater than the mean diversity $d^a_{p_i}$ of the values vf_a from the values projected to the corresponding feature variable f_a across all clusters of the other sentiment polarity labels, then the other sentiment polarity labels are more likely to be positive or negative. Estimates of diversity for all feature attributes are now calculated by fusing the diversity measures considered (see section 3.2). Subsequent description depicts the algorithm used to optimize the features:

//Determining the best feature variables for the clusters "em_i" with sentiment polarity label records.//

- Do the following for each class of sentiment polarity l in the collection of records $\{angr, dsg, sad, joy\}$.

- Execute the following for each cluster " $\{em_i \exists i = 1,2,...,s\}$ " of the opinion polarity l .

$$\overset{|Fa|}{\underset{j=1}{\forall}} \{ f_j \exists f_j \in Fa \}$$

$$if\,(distribution\_diversity(f_j, vf_j, emC) > d\tau)$$

$$ofa(c_i) \leftarrow f_i$$

$distribution\_diversity(f, v, emC)$ /`/ //The values for feature-attribute f , which represent the final feature, have been obtained from the input cluster as well as the clusters of the other opinion polarity labels.

Begin

$ds = 0$ //distribution diversity measure

$\overset{|emC|}{\underset{i=1}{\forall}} \{em_i \exists em_i \in emC\}$ Begin // each of the clusters em

$fs = 0$ // fusion score

$vf \leftarrow em_i(f)$ // This expression represents the values that will be assigned to the feature property f for the em_i cluster.

$$fs+= \left\{ 1\exists \begin{pmatrix} wrsTest(v,vf) > 0 \vee \\ ksTest(v,vf) > 0 \vee \\ wsrTest(v,vf) > 0 \vee \\ fmTest(v,vf) > 0 \end{pmatrix} \right\}$$

//updates by the response of all suggested distribution diversity assessment measures

$$ds+= \left\{ 1\exists \left(\left(1 - \frac{1}{fs} \right) \geq 0.5 \right) \right\}$$

//If fusion metrics show that there is a lot of variety, then this expression will update the distance stat.

End

Return $\dfrac{ds}{|emC|}$ // returns ratio of distribution diversity measure

End // denotes the completion of the function $distribution\_diversity(f, v, emC)$

- Find all conceivable subsets " ngf " of the ideal feature variables " F " for each cluster of the opinion polarity labels.
- The total number of distinct subsets from the specified set of feature variables is denoted by the notation " $|ngf| = 2^{|F|-1}$ "

Find the distinct pattern values and predicted frequencies for each item in the set " ngf " in the resultant cluster of records

3.5 The classifier

One popular machine learning technique, called Random Forest (RF) [35], takes the decisions made by several decision trees (DT) and averages them into a single conclusion. Its accessibility and versatility in handling equally classification as well as regression issues have contributed to its meteoric rise in popularity.

By combining bagging with feature randomness, the RF algorithm generates a set of decision trees that are independent of one another. A randomly chosen subset of the features that often termed as either "feature-bagging" or "random-subspace approach," which provides minimal correlation across decision trees. This is where DT diverge significantly from RF. When making their selections, RFs focus on a subset of features, whereas DTs take into account all potential feature splits. Overfitting, bias, as well as general variance can be mitigated by considering all possible variations in the data. Therefore, our forecasts will be more reliable.

There are three key hyper parameters for random forest algorithms that must be established before training can begin. Examples of such factors are node size, tree depth, and feature sampling. The RF classifier may then be utilised to address any outstanding issues with your data's regression or classification.

Each tree in the ensemble of the random forest method is built using a sample of data selected from the data used in training phase with substitution, known as the bootstrap sample. One-third of the training sample is removed and stored as test data. Feature bagging then injects another kind of randomization into the mix, broadening the scope of the dataset while simultaneously decreasing inter-tree correlation. The manner in which the forecast is arrived at varies with the nature of the underlying problem. When doing a classification job, the majority vote, or the most common categorical variable, will be used to predict a class, whereas when performing the task of regression, the single DT will be averaged. Once the training is complete, the test data is utilised for cross-validation.

Details of the random forest algorithm's execution

Random Forest starts by selecting n random records from the dataset of k records.

In the second stage, separate decision trees are built for each sample.

The process flow of the Random Forest (RF)

Start the $\boldsymbol{RF}$ Algorithm.
$\boldsymbol{N}$ stands for the total nodes in the input.
The total modified characteristics is indicated by the symbol $\boldsymbol{M}$.
The trees that must be developed in total is indicated by the symbol $\boldsymbol{D}$.
The class with the most votes is indicated by the notation output $\boldsymbol{V}$.
Do the following ***while*** the stop requirement is false.
The Bootstrap sample was selected at random from the data used in training phase shown as $\boldsymbol{D}$.
The building of the tree $"\boldsymbol{T_i}"$ for sample $"\boldsymbol{A}"$ is to be completed using the subsection of rules below.

- Choose $"\boldsymbol{m}"$ characteristics from the collection $"\boldsymbol{M}"$ where $"\boldsymbol{m}" \ll "\boldsymbol{M}"$.
- The best split point between the characteristics $"\boldsymbol{m}"$ of the node is then determined.
- The best split technique must be used to divide the node into two sub-nodes.
- Up until the necessary number of nodes are reached for execution, the three phases outlined above will be repeated. The procedure must be done several times in order to advance.

The "***while***" condition has ended.
The number of constructed trees is the outcome.
Every construct should have the sample applied to it starting at the root node.
It is necessary to assign the data-sample to the appropriate leaf-node class.
Analyze the final trees' total votes and judgments.
The output will be the vote class having highest rating.
Stop the $\boldsymbol{RF}$ algorithm.

As a third step, the results of each decision tree are presented.

The fourth step involves taking into account the results of the classification or regression based on the majority vote or the average.

To vote randomly in a forest. Take the fruit basket in the following illustration as an example of data. In this step, we choose $'\boldsymbol{n}'$ random samples of fruit-basket and build a $\boldsymbol{DT}$ for each one. Decision - making is based on a simple majority. When asked to choose between an apple and a banana, most decision trees choose the former, so we go with that.

4. Experimental study

This section compares the proposed model to existing solutions in the literature and focuses on its practical use. In this part, we will discuss the dataset, the modifications to the software requirements, and the system circumstances, all of which are crucial to the performance study but are measured by cross validation metrics.

The model is run using the Python [47], [48], [49] programming language, and the code is created with PyCharm [46]. Model experiment hardware requirements include a 7th-generation Intel processor CPU, 16GB of RAM, and 2TB of storage space.

4.1. The Data

Table 1: Data from input records annotated with sentiment labels

Label ID	Positive	negative	Neutral	Mismatch
Total Records	56250	39094	28312	26344
Cluster Count	5	4	2	3
Cluster ID⇓	Count of Records in Respective Clusters⇓			
1	21938	13292	17553	12118
2	13500	15247	15855	8167
3	10125	17983	0	11064
4	14625	15638	0	0
5	19125	0	0	0

The Twitter dataset Charlottesville [50], [51] has 150000 tweets such that each tweet is a record. Further labeled the records as positive, negative, neutral, and mismatch through human annotation. To reduce the high dimensionality, the projected fuzzy c means [46] clustering technique has been applied. The statistics of the number of clusters framed are listed in the following table 1

One hundred fifty thousand tweets were employed in the controlled experiment on the two extremes of public opinion. Table-1 below depicts the dataset's statistical characteristics. Word patterns with varying degrees of positive or negative connotation are the framework for the range of emotions represented in the input data.

4.2. **Data Processing**

The Twitter tweets' input data are converted into word patterns, one for each sentiment polarity. Each tweet in the input stream is interpreted as a data record consisting of a vector space of word patterns. By the end of the data analysis, we have four datasets, one for each of the four polarity labels of opinion we considered.

4.3. **Performance Analysis**

The performance analysis has been done through multi-labelled 4-fold cross-validation. The results obtained for cross-validation metrics Ensemble Learning for Opinion Valuation (ELOV) have been compared to the cross-validation metric values of the contemporary models EMLA [33] and EMOC [34], which have been obtained from the experimental study performed using contemporary models on the same dataset. A detailed description of the comparative study is projected in the following exploration of the cross-validation metrics.

4.3.1. **Measures**

The confusion-matrix is used to produce the measure to scale the effectiveness of the classifiers, where the confusion-matrix provides the results for incorrect and accurate instances found for each event class. Thus, a range of specified measures that are statistically deliberated and employed in classifier comparative analysis is achievable.

- TP denotes the true positives, which are the set of records truly identified as positive from the records given for label prediction.
- TN denotes the true negatives, which are a subset of the records given for label prediction that are really categorized as negative.
- FP stands for false positives, a collection of records that were incorrectly categorized as positive based on the data provided for label prediction.
- FN stands for False Negatives, a collection of records that were incorrectly categorized as negative based on the records provided for label prediction.

These TP, FP, TN, and FN are crucial for estimating the many performance metrics used in performance analysis, where the scope is determined by the classifier quality chosen. Among the essential metrics employed in analysis are specificity, accuracy, sensitivity, receiver operating characteristic (ROC) value, and rate of detection. To address the estimations discussed previously, it is necessary to measure the rate of detection and the receiver operating characteristic (ROC) curve (receiver operating characteristics). Additionally, the rate of detection, commonly referred to as AC (accuracy), more precisely specifies the classified situations, as defined in the Eq 10.

$$AC = (TP + TN) * (TP + TN + FP + FN)^{-1} \quad \text{... (Eq 2)}$$

In compared to a 'Gold Standard' or criteria, specificity and sensitivity quantitatively characterize the correctness of a test that reports the occurrence of a condition.

- The proportion of people that have the ailment (as determined by the 'Gold Standard') who had a positive response on this test is referred to as sensitivity (True Positive Rate) (see Eq 11).
- The fraction of people not having the ailment (as determined by the 'Gold Standard') who had a negative response on this test is referred to as specificity (True Negative Rate) (see Eq 12).

Sensitivity is an indicator of how much a check can identify true positives, while specificity is indeed a measure of how much a check can recognize true negatives in a clinical diagnosis. There is always a barter among specificity and sensitivity in all screening and diagnostic testing, with higher sensitivities implying fewer specificities and conversely.

Precision (also known as positive predictive value) is the percentage of relevant examples found among the recovered instances (see Eq 13).

$$\textbf{True Positive Rate}(\textbf{sensitivity}) = TP * (TP + TN)^{-1} \quad \text{... (Eq 3)}$$

$$\textbf{True Negative Rate}(\textbf{Specificity}) = TN * (TN + FP)^{-1} \quad \text{... (Eq 4)}$$

$$\textbf{Positive Predective value(Precision)} = TP * (TP + FP)^{-1} \quad \text{... (Eq 5)}$$

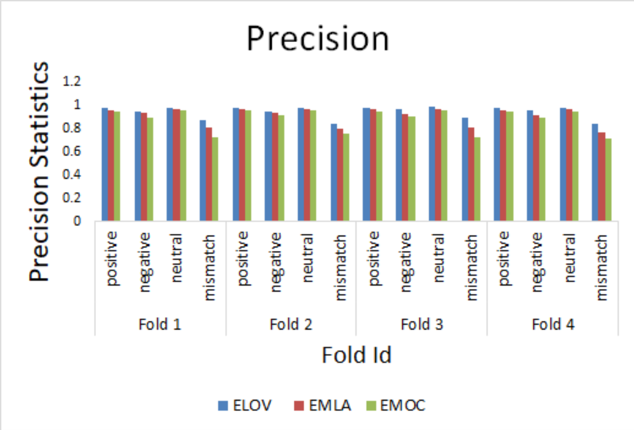

Figure 2: Precision rate related performance of the models across the four-fold cross-validation

In assessment of the sentiment polarity conditions, across all the three models compared as ELOC, EMLA and EMOC, it is imperative from the figurative representation in figure 2, that the ELOV model proposed in this manuscript has delivered significant performance in comparison to the other solutions, across the four folds tested in the experimental study.

Metric specificity is the other critical factor observed for performance analysis across the four-folds for all the three models compared in the analysis. Technically referred as TNR (True-Negative Rate), it is signified as the ration for TNs, wherein the sum of FPs and TNs are estimated as integral to the process. In terms of measuring the performance of the model ELOV, and other such contemporary models compared in the study, the performance of the model as depicted in the figure 3 refers to how various models have fared in the experimental study. From the inputs assessed, it is evident that the solution proposed in this manuscript as ELOV, has fared well in comparison to the other models.

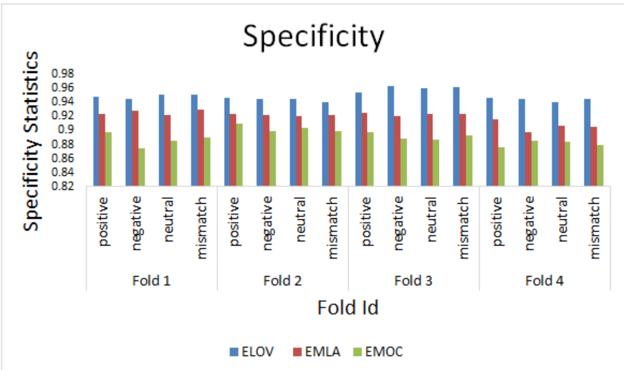

Figure 3: Performance of Models for the TNR Specificity across diversity thresholds

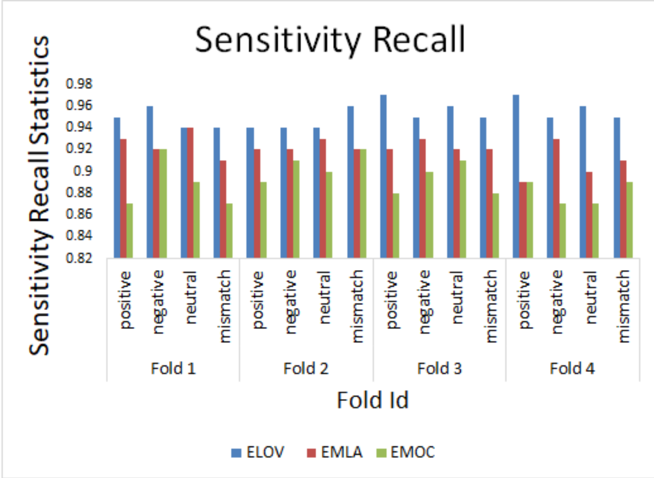

Figure 4: Recall/sensitivity (label level accuracy) statistics of both ELOV, EMLA, and EMOC observed for diversified labels

Recall is the other critical metric referring to the ratio of true positions in comparison to the aggregation of true positives and false negatives. The graph is plotted amongst the conditions of sensitivity, and divergent labels. The performance of all the three models applied for the comparative study as a plotted review is depicted in the figure 4 below. Performance of ELOV is seen superior to the other metrics, and in comparison, to the contemporary values of EMLA and EMOC models.

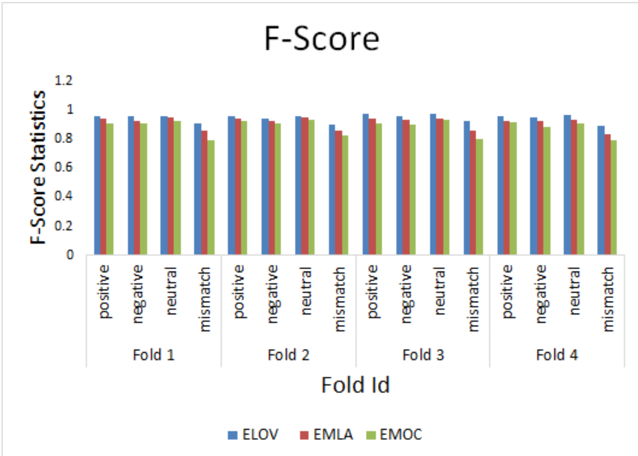

Figure 5: F-Score statistics related to ELOV, EMLA, and EMOC Performance Analysis

The graphs represented above indicates the performance of the models in terms of assessing the F-Score conditions wherein the harmonic mean related to each of the precision and sensitivity is assessed in the system. The key statistics discussed in the figure 5 above refers to performance of the models with respective to the F-score patterns. The proposed model of this manuscript ELOV has shown superior performance in comparison to the other models EMLA and EMOC.

Figure 6 indicates the performance of the models related to the accuracy levels in the detection of the records more effectively. Performance of ELOV is superior in comparison to the other models EMLA and EMOC, across the four folds, and it is evident that the model can be more pragmatic for implementation in the real-time environment.

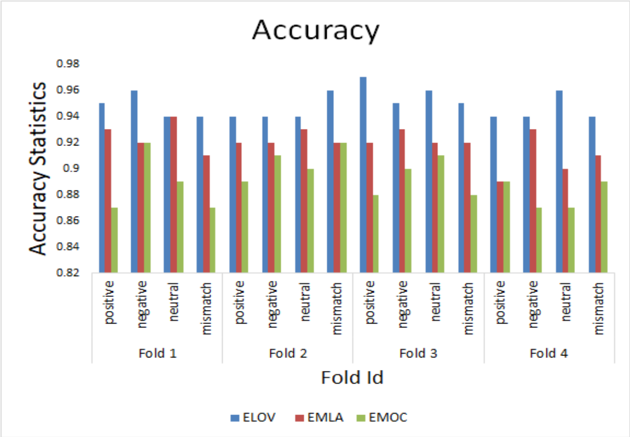

Figure 6: Accuracy related performance analysis for ELOV, EMLA & EMOC

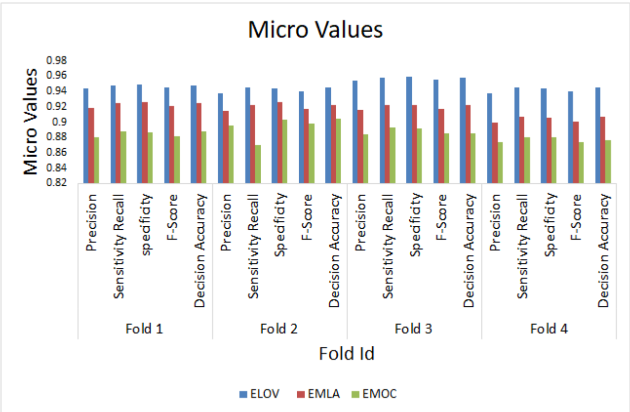

Figure 7: Figurative representation of a comparison for the metrics related to ELOV and contemporary models EMLA and EMOC

Figure 7 representation of the holistic performance of the models across all the four folds refers to the potential performance of each of the individual model and in comparison, to the proposed model ELOV. While the ELOV stands a superior in terms of performance, the other model that can be seen faring well is the EMLA, in comparison to EMOC. While the performance conditions and features chosen for analysis could have certain levels of influence on the outcome, for the chosen scenario, across the models compared for the study, ELOV stands top performing.

5. Conclusion

An ensemble classification to detect the sentiment polarity has portrayed in this manuscript. The main objective of the proposed model is to handle the curse of high dimensionality in the given training corpus to predict the sentiment polarity of multiple labels (positive, negative, neutral, and mismatch). In this regard the given training corpus in to multiple sets using fuzzy c-means clustering algorithm. Further detects the optimal word patterns of each cluster using fusion of diversity assessment measures. Further, these optimal word patterns of each cluster has been used to train the corresponding classifier assigned to that cluster. Further detects the sentiment polarity of multiple labels using “ensemble majority vote classification. The future research shall focus on other properties of tweets such as emoticons to include along with word patterns.

References

[1] Bing Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Hu- man Language Technologies 5 (2012), no. 1, 1-167.

[2] Hamdan, Yasir Babiker. "Faultless Decision Making for False Information in Online: A Systematic Approach." Journal of Soft Computing Paradigm (JSCP) 2.04 (2020): 226-235.

[3] Pooja Kumari, Shikha Singh, Devika More, and Dakshata Talpade, “Sentiment Analysis of Tweets”, IJSTE - International Journal of Science Technology & Engineering, ISSN: 2349-784X, Volume: 1, Issue: 10, pp: 130-134, 2015.

[4] Asst. Prof. A Kowcika, Aditi Gupta, Karthik Sondhi, Nishit Shivhre, and Raunaq Kumar, “Sentiment Analysis for Social Media”, International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume: 3, Issue: 7, 2013.

[5] Ali Hasan, Sana Moin, Ahmad Karim, and Shahaboddin Shamshirband, “Machine Learning-Based Sentiment Analysis for Twitter Accounts”, Mathematical and Computational Applications, ISSN: 2297-8747, Volume: 21, Issue: 1, 2016.

[6] Kadhim, R. R., and M. Y. Kamil. “Evaluation of Machine Learning Models for Breast Cancer Diagnosis Via Histogram of Oriented Gradients Method and Histopathology Images”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 10, no. 4, Apr. 2022, pp. 36-42, doi:10.17762/ijritcc.v10i4.5532.

[7] Rasika Wagh, and Payal Punde, “Survey on Sentiment Analysis using Twitter Dataset” 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018) IEEE Conference, ISBN: 978-1-5386-0965-1,2018.

[8] R. Sharma, S. Nigam, and R. Jain, “Opinion mining of movie reviews at document level,” arXiv preprint arXiv: 1408.3829, 2014.

[9] R. Sharma, S. Nigam, and R. Jain, “Polarity detection at sentence level,” International Journal of Computer Applications, vol. 86, no. 11, 2014.

[10] Sally Fouad Shady. (2021). Approaches to Teaching a Biomaterials Laboratory Course Online. Journal of Online Engineering Education, 12(1), 01–05. Retrieved from http://onlineengineeringeducation.com/index.php/joee/article/view/43

[11] A. Harb, M. Plantié, G. Dray, M. Roche, F. Trousset, and P. Poncelet, “Web Opinion Mining: How to extract opinions from blogs?” in Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology, 2008: ACM, pp. 211-217.

[12] T. Zagibalov and J. Carroll, “Unsupervised classification of sentiment and objectivity in Chinese text,” in Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I, 2008.

[13] A. Tripathy and S. K. Rath, “Classification of sentiment of reviews using supervised machine learning techniques,” International Journal of Rough Sets and Data Analysis (IJRSDA), vol. 4, no. 1, pp. 56-74, 2017.

[14] M. R. Saleh, M. T. Martín-Valdivia, A. Montejo-Ráez, and L. UreñaLópez, "Experiments with SVM to classify opinions in different domains," Expert Systems with Applications, vol. 38, no. 12, pp. 14799- 14804, 2011.

[15] V. Kharde and P. Sonawane, “Sentiment analysis of twitter data: A survey of techniques,” arXiv preprint arXiv: 1601.06971, 2016.

[16] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” CS224N Project Report, Stanford, vol. 1, no. 2009, p. 12, 2009.

[17] J. Read, “Using emoticons to reduce dependency in machine learning techniques for sentiment classification,” in Proceedings of the ACL student research workshop, 2005: Association for Computational Linguistics, pp. 43-48.

[18] M. Anjaria and R. M. R. Guddeti, “Influence factor based opinion mining of Twitter data using supervised learning,” in 2014 Sixth International Conference on Communication Systems and Networks (COMSNETS), 2014, pp. 1-8.

[19] A. Barhan and A. Shakhomirov, “Methods for Sentiment Analysis of twitter messages,” in 12th Conference of FRUCT Association, 2012.

[20] P.-W. Liang and B.-R. Dai, “Opinion mining on social media data,” in Mobile Data Management (MDM), 2013 IEEE 14th International Conference on, 2013, and vol. 2: IEEE, pp. 91-96.

[21] A. Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining,” in LREc, 2010, vol. 10, no. 2010.

[22] Chaudhary, D. S. . (2022). Analysis of Concept of Big Data Process, Strategies, Adoption and Implementation. International Journal on Future Revolution in Computer Science &Amp; Communication Engineering, 8(1), 05–08. https://doi.org/10.17762/ijfrcsce.v8i1.2065

[23] H. Saif, Y. He, and H. Alani, “Semantic sentiment analysis of twitter,” in International semantic web conference, 2012: Springer, pp. 508-524.

[24] H. Hamdan, F. Béchet, and P. Bellot, “Experiments with DBpedia, WordNet and SentiWordNet as resources for sentiment analysis in micro-blogging,” in Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), 2013, vol. 2, pp. 455-459.

[25] F. Akba, A. Uçan, E. A. Sezer, and H. Sever, “Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews,” in 8th European Conference on Data Mining, 2014, vol. 191, pp. 180-184.

[26] Sharma, V. N., & Hans, D. A. . (2022). A Study to Reconnoitering the dynamics of Talent Management Procedure at Hotels in Jharkhand. International Journal of New Practices in Management and Engineering, 11(01), 41–46. https://doi.org/10.17762/ijnpme.v11i01.172

[27] H. Saif, Y. He, and H. Alani, “Alleviating data sparsity for twitter sentiment analysis,” 2012: CEUR Workshop Proceedings (CEUR-WS. org).

[28] Kabisha, M. S., Rahim, K. A., Khaliluzzaman, M., & Khan, S. I. (2022). Face and Hand Gesture Recognition Based Person Identification System using Convolutional Neural Network. International Journal of Intelligent Systems and Applications in Engineering, 10(1), 105–115. https://doi.org/10.18201/ijisae.2022.273

[29] E. Martınez-Cámara, Y. Gutiérrez-Vázquez, J. Fernández, A. MontejoRáez, and R. Munoz-Guillena, "Ensemble classifier for Twitter Sentiment Analysis," 2015.

[30] T. Chalothom and J. Ellman, “Simple Approaches of Sentiment Analysis via Ensemble Learning,” Berlin, Heidelberg, 2015: Springer Berlin Heidelberg, pp. 631-639.

[31] M. M. Fouad, T. F. Gharib, and A. S. Mashat, “Efficient Twitter Sentiment Analysis System with Feature Selection and lassifier Ensemble,” in International Conference on Advanced Machine Learning Technologies and Applications, 2018: Springer, pp. 516-527.

[32] X. Hu, J. Tang, H. Gao, and H. Liu, “Unsupervised sentiment analysis with emotional signals,” in Proceedings of the 22nd international conference on World Wide Web, 2013: ACM, pp. 607-618.

[33] N. Azzouza, K. Akli-Astouati, A. Oussalah, and S. A. Bachir, “A realtime Twitter sentiment analysis using an

unsupervised method," in Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics, 2017: ACM, p. 15.
[34] R. Ortega, A. Fonseca, and A. Montoyo, "SSA-UO: unsupervised Twitter sentiment analysis," in Second joint conference on lexical and computational semantics (* SEM), 2013, vol. 2, pp. 501-507.
[35] G. Paltoglou and M. Thelwall, "Twitter, MySpace, Digg: Unsupervised sentiment analysis in social media," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 3, no. 4, p. 66, 2012.
[36] F. M. Kundi, A. Khan, S. Ahmad, and M. Z. Asghar, "Lexicon-based sentiment analysis in the social web," Journal of Basic and Applied Scientific Research, vol. 4, no. 6, pp. 238-48, 2014.
[37] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim, and I. A. Khan, "Lexiconenhanced sentiment analysis framework using rule-based classification scheme," PloS one, vol. 12, no. 2, p. e0171649, 2017.
[38] Nazeer, Ishrat, et al. "Use of Novel Ensemble Machine Learning Approach for Social Media Sentiment Analysis." Analyzing Global Social Media Consumption. IGI Global, 2021. 16-28.
[39] Abbas, Alaa Khudhair, et al. "Twitter sentiment analysis using an ensemble majority vote classifier." Journal of Southwest Jiaotong University 55.1 (2020).
[40] Matsuki, Kazunaga, Victor Kuperman, and Julie A. Van Dyke. "The Random Forests statistical technique: An examination of its value for the study of reading." Scientific Studies of Reading 20.1 (2016): 20-33.
[41] Yang XS, Deb S. Cuckoo search: recent advances and applications. Neural Computing and Applications. 2014 Jan 1; 24(1):169-74.
[42] Ghasemi, Asghar, and Saleh Zahediasl. "Normality tests for statistical analysis: a guide for non-statisticians." International journal of endocrinology and metabolism 10.2 (2012): 486.
[43] Budak, H. and TAS ̧abat, S.E. "A modified t-score for feature selection", Anadolu U niversitesi € Bilim Ve Teknoloji Dergisi A-Uygulamalı Bilimlerve Muhendislik €, Vol. 17 No. 5, 2016, pp. 845-852.
[44] https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php
[45] McKnight, Patrick E., and Julius Najab. "Mann Whitney U Test." The Corsini encyclopedia of psychology (2010): 1-1.
[46] Rey, Denise, and Markus Neuhäuser. "Wilcoxon-signed-rank test." International encyclopedia of statistical science.Springer Berlin Heidelberg, 2011. 1658-1659.
[47] https://www.statisticssolutions.com/kruskal-wallis-test/
[48] T-table. http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf, (2017).
[49] Rudra Kumar, M., and Vinit Kumar Gunjan. "Peer Level Credit Rating: An Extended Plugin for Credit Scoring Framework." ICCCE 2021. Springer, Singapore, 2022. 1227-1237
[50] Rudra Kumar, M., Rashmi Pathak, and Vinit Kumar Gunjan. "Machine Learning-Based Project Resource Allocation Fitment Analysis System (ML-PRAFS)." Computational Intelligence in Machine Learning. Springer, Singapore, 2022. 1-14
[51] M. Shasidhar, V.Sudheer Raja, B. Vijay Kumar, "MRI Brain Image Segmentation Using Modified Fuzzy C-Means Clustering Algorithm", IEEE International Conference on Communication Systems and Network Technologies, 2011.
[52] Madapuri, Rudra Kumar, and P. C. Mahesh. "HBS-CRA: scaling impact of change request towards fault proneness: defining a heuristic and biases scale (HBS) of change request artifacts (CRA)." Cluster Computing 22.5 (2019): 11591-11599
[53] Chalapathi, M. M., et al. "Ensemble Learning by High-Dimensional Acoustic Features for Emotion Recognition from Speech Audio Signal." Security and Communication Networks 2022 (2022).
[54] Python. (n.d.). Retrieved from https://www.python.org/downloads/
[55] Pycharm. (n.d.). Retrieved from https://www.jetbrains.com/pycharm/download/
[56] https://www.kaggle.com/vincela9/charlottesville-on-twitter/version/1/metadata