International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Automatic Sentiment Analysis Model Creation using Multi Kernel Improved Extreme Learning Machine

Srinidhi B. S., Suchithra R.

Abstract*: Recent research activities related to opinion mining, sentiment analysis and emotion detection from natural language texts are all under the umbrella of affective computation. There is now a huge amount of textual information on social media (for example, forums, blogs, and social media) about consumers' ideas about buying products and service experiences. Sentiment analysis or opinion mining is part of an investigation that analyzes people's thoughts and feelings from written text available online. In this paper, this work present a comprehensive experiment to evaluate the effectiveness of psychological and linguistic features in emotion classification. In this scheme, we used five broad categories of LIWC (namely, psychological processes, linguistic processes, punctuation, spoken categories and personal concerns) as feature sets. Five types of LIWCs and their group combinations were considered in the experimental analysis. To understand the predictive performance of various aspects of the engineering scheme, five controlled learning algorithms (namely, Naïve Bayes, support vector machines, Extreme Learning Machine, Kernel Extreme Learning Machine, Multi Kernel Extreme Learning Machine) and proposed Multi Kernel Improved Extreme Learning Machine (MKIELM) are used. Experimental results show that the ensemble feature sets provides a higher predictive effect than the individual set..

Keywords: Natural Language Processing (NLP), Emotion Recognition, social media analysis, Sentiment Analysis, psychological feature sets, Twitter, ELM and MKIELM.

I. INTRODUCTION

The huge amount of information that can be obtained with the significant growth of social media and micro blogging platforms can be an important source of decision-making for products, services and policies [1-2]. Twitter is a popular and fast-growing platform where people can send short messages (referred to as tweets) in a limit of 140.

Twitter enables users to communicate effectively. User-generated content on Twitter provides a useful source of information for researchers and practitioners [3]. Information received from Twitter can serve as an important source of information for a wide range of applications, including event detection, epidemiology, news and crisis management [4-6]. Emotional analysis is an important area of research in natural language processing that aims to determine the direction of Emotion in material objects. Emotional analysis can be used to obtain information about new products and services. It can be further applied to identify the positive and negative aspects of a particular product or service [7]. Emotional analysis methods can be divided into two groups: the acoustic-based approach and the machine-based approach.

In addition, emotional analysis can be performed in different ways, depending on the level of detail. Depending on the level of detail, sentiment analysis methods are grouped into three categories: sentiment analysis at the file level, sentence level and view [8]. Emotional analysis can be modeled as problems with text classification. In machine learning-based emotion analysis, managed classification algorithms (such as the Naive Bayes algorithm, support vector machines, k-nearest neighbor algorithm and logistic regression) can be used to target emotion.

Machine learning sentiment analysis schemes include pre-processing of data, extracting functions, and selecting and training classification algorithms that are controlled with labeled datasets. In order to obtain a scheme of high predictive classification efficiency, the extraction of traits is a necessary task [9]. LIWC (Language Investigation and Word Counting) is a research program for text analysis to extract psychometric features from text documents. Characteristics related to psychological, linguistic, social, and cultural aspects may be important for emotional analysis [10]. To this end, we present a psychological approach to the analysis of emotions on Twitter. In this document, we used five major LIWC categories (namely, linguistic processes, psychological processes, personal concerns, spoken categories and punctuation) as feature sets. Five types of LIWCs and their group combinations were considered in the experimental analysis. To investigate the anticipation of the different functions that constitute the scheme of work, this introduces the CSCNNN classification method.

The rest of the article is organized as follows. Section 2 presents the work related to emotional analysis. Section 3 presents the methodology of the study and Section 4 presents the experimental procedures and the actual results. Section 5 describes concluding remarks.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
B. S. Srinidhi*, Scholar, Department of MCA, Jain Deemed to be University, Bangalore, India.
R. Suchithra, Director, Department of MCA, Jain Deemed to be University, Bangalore, India.

II. RELATED WORK

In recent years, several proposals have presented studies based on two main methods of polarization: Semantic Orientation and Machine learning. The Semantic Orientation method (SO) uses lexical resources such as acoustics that are automatically generated or semi-automated [13], [14], [15]. These lexicons extend the featured WordNet database. Two obvious examples of this lexicons are WordNet-Affect [14] and SentiWordNet [13]. Many suggestions based on sentimental encyclopedias have been featured in the literature. For example, in [16], an acoustic approach was proposed to extract emotions from the text. The method is based on The Semantic Orientation CALculator (SO-CAL), that uses the dictionaries of words annotated with their semantic orientation (polarity and strength), and includes both reinforcement and negative. [[12] proposed an innovative approach that uses new web solutions to enhance the results obtained with traditional language processing techniques and website sentiment and technological analysis processes. Specifically, the proposed method is based on three steps: 1) an ontology based mechanism for feature identification, 2) a technique to assign a polarity to each feature based on SentiWordNet and 3) a new approach for opinion mining based on vector analysis. [17] presented a principle-based method for determining the specificity of sentence comments. SentiWordNet was used to calculate the total emotional score of each sentence. Moreover, the results suggest that SentiWordNet can be used as an important resource for emotion classification tasks. Most opinion polls are related to English. In fact, the major drawback of the sentiment-based approach is the lack of sentimental vocabulary in non-English languages, which makes it difficult to apply in other languages.

Despite the numbers mentioned above, there are some interesting works that provide a focused encyclopedia for languages such as Spanish [15], [18] German [19], Dutch [20] and Arabic [21]. In addition, due to the limitations mentioned above, some proposals use alternative methods. The first alternative approach uses a number of lexical resources to analyze WordNet-based emotions (independent of target language such as English) together with lexical databases based on WordNet and specific for target or multilingual languages [22]. The second alternative method uses a different emotional glossary than the target language. Therefore, automatic hull translation is performed before the polarization phase. However, this method depends on the availability and reliability of the available automated translator [23] [24]. On the other hand, there are a number of suggestions [25] [26] [27] [28] based on psychoanalysis tools such as lava, a tool that provides vocabulary in several languages: Spanish, English, French and German, among others. . These suggestions form a framework based on two dynamic orientations ("positive emotion" and "negative emotion") of 76 categories of LIWC. These categories include terms such as love, good, good, bad, bad, bad, bad, and worse. The LIWC dictionary has been used in a variety of studies, including languages such as English [25], [27], [28], Portuguese [26] and Spanish [29]. In addition, with regard to machine learning methods, it is worth noting that this method often relies on the assigned classification method. These methods use data collection to train classifier algorithms. Machine learning techniques commonly used in the sentiment polarity classification are Support Vector Machine (SVN), Naive Bayes (NB), Maximum Entropy (MaxEnt), among others. For example, in [30] they compared three controlled machine learning methods (Naïve Bayes, SVM and the character based N-gram model) for emotion classification. This comparison is made on user reviews obtained from travel blogs. The authors conclude that a well-trained machine learning algorithm can provide a good idea of the sentiment polarity classification on the travel domain. [31] examined whether different settings (n-gram size, corpus size, number of sentiment classes, balanced vs. unbalanced corpus, various domains) affect the accuracy of machine learning algorithms. Naïve Bayes, Decision Tree, and Support Vector Machines were considered.

A. Submission of the paper

Author (s) can send paper in the given email address of the journal. There are two email address. It is compulsory to send paper in both email address.

B. Plagiarism Check

Good quality plagiarism software/ tool (Turnitin / iThenticate) will be used to check similarity that would not be more than 20% including reference section. In the case of exclusion of references, it should be less than 5%.

C. Quality Check

All submitted paper should be cutting edge, result oriented, original paper and under the scope of the journal that should belong to the engineering and technology area. In the paper title, there should not be word 'Overview/brief/ Introduction, Review, Case study/ Study, Survey, Approach, Comparative, Analysis, Comparative Investigation, Investigation'.

D. Paper Submission Criteria

Any one author cannot submit more than 05 papers for the same volume/issue. The authors of the accepted manuscripts will be given a copyright form and the form should accompany your final submission. It is noted that:

- Each author profile along with photo (min 100 word) has been included in the final paper.
- Final paper is prepared as per journal the template.
- Contents of the paper are fine and satisfactory. Author (s) can make rectification in the final paper but after the final submission to the journal, rectification is not possible.

III. PROPOSED APPROACH

This section describes the process of collecting data sets, processes the engineering schemas for displaying datasets, classification algorithms, and combines the training methods used in experimental analysis. The outline of our proposed work is shown in Figure 1. The proposed work consists of four modules. They are

1. Dataset Collection
2. Data Preprocessing
3. Feature Extraction
4. Sentiment Analysis Model Generation.

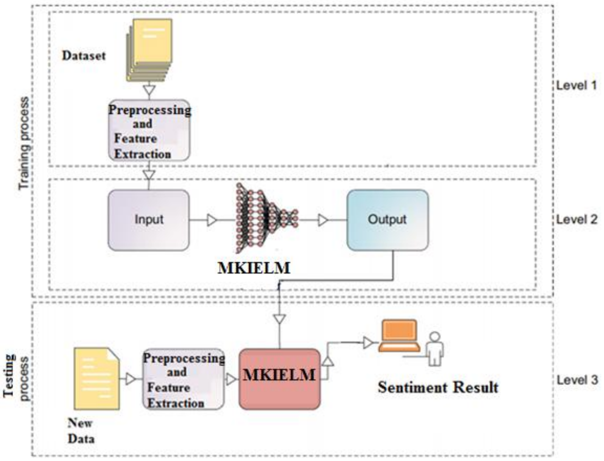


Fig. 1. Outline of the Proposed Work

A. Dataset Collection

To assess the predictive performance of the psychosocial and linguistic features of emotion analysis, we analyzed Twitter posts that contained positive, negative, and neutral emotions. In compiling the database, we adopted the framework presented in [11]. We used Twitter4J, an open source Java library, to use Twitter Streaming API, to collect tweets. Each tweet is labeled by a class of labels, either positive, negative or neutral. After collecting tweets, automatic filtering is performed to remove inappropriate and unnecessary tweets (copy and duplicate). So we got a collection of 6,218 negative, 4891 positive and 4252 tweets. To achieve a balanced corpus, our final dataset contains 4,200 tweets of negative tweets and 4,200 positive tweets.

B. Preprocessing

Due to the precarious and informal nature of Twitter posts, pre-tweeting is necessary to eliminate specific issues (such as unnecessary repetition and text abuses) [12]. In the pre-treatment phase, we adopted the framework presented in [32]. The pre-processing phase is primarily aimed at eliminating unnecessary symbols or sequences that are not valuable for emotion classification. To this end, the following tasks are performed for each tweet [32]:

• Exclude mentions and replies from other users' tweets, represented by a string that begins with "@".
• Eliminating URLs (such as strings with "http: //").
• Eliminate the "#" character

C. Feature Extraction

After the database has been processed, the next step is to create a feature matrix. Before feature extraction, tokens are first applied to pre-processed data. Tokenization is simply the process of dividing a sentence into words. Then stop. Stop words a frequently used word and is so common that they lose their semantic meaning. Words like “of, are, the, it, is” are just a few examples of the word stop. There are two common methods of removing words, and both are clear.

One way is to count all event words and assign a numeric value for the number and get rid of any terms/words that occur more than the specified value. Another way is to have a predefined word break list that can be drawn from a short list of stop words, which can be removed from the list of tokens/tokenized sentences. n our work, we have implemented both methods to remove stop words.

Finally, the process of stemming is continued. The purpose of is to reduce the reflective form and sometimes relate the derivative of the word to a common basic form. Stemming is easy way to reduce conditions to their roots, just set a rule for cutting some characters at the end of a word and hope they get the best of them. And then the features of the matrix are created.

In this section, we examine different psychological sets of emotional analysis. In this scheme, we used LIWC (linguistic investigation and word counting) to extract psychometric features from the dataset. LIWC has been successfully used in many areas of the language, including the identification of ambiguities and satirical findings [33, 34]. LIWC is a textual analysis program for identifying emotional, cognitive, and structural aspects of speech and speech patterns. The first version of the LIWC was launched in 1993 and the last version was released in 2015 [35].

The LIWC dictionary contains approximately 6400 words, words, and emoticons. Each dictionary entry contains one or more types of words or hyphens. For a particular word that occurs in the text, ratings for each category or dictionary are increased. This category can be categorized into five major categories: psychological processes, linguistic processes, punctuation, spoken categories and personal concerns. In Table 1, main LIWC sets and categories are listed.

Table I. Main liwc sets and categories

Feature Set	Categories
Linguistic Processes	Word count, total pronouns, personal pronouns, articles, prepositions, auxiliary verbs, adverbs, conjunctions
Psychological Processes	Affective processes, positive emotion, negative emotion, social processes, cognitive processes, perceptual processes
Personal Concerns	Work, leisure, home, money
Spoken Categories	Assent, Non-fluencies, fillers
Punctuation	Total punctuation, periods, commas, colons, semicolons, question marks, exclamation marks, dashes

As can be seen from the categories listed in Table 1, language processing includes grammatical information such as word count, pronouns, personal pronouns, articles, prepositions and auxiliary verbs. Psychological processes include psychological information such as affective, positive, and negative emotions. Personal concerns include information such as work, entertainment, home and money. Speech types include information about spoken languages.
Finally, punctuation sets include punctuation marks such as punctuation marks, punctuation marks, question marks, dashes, exclamation points. Based on the five key LIWC features listed above, the target word or phrase is searched through the LIWC dictionary. Each word is assigned to one or more subdirectories.

Retrieval Number: E6842018520/2020©BEIESP
DOI:10.35940/ijrte.E6842.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

D. Sentiment Analysis Model Generation

To build a high-efficiency Sentiment Analysis model by Multi kernal Improved Extreme Learning Machine (MKIELM). ELM has the advantage of fast training speed and high generalization performance, but ELM also has the disadvantage of bad robustness. To solve this problem the kernel method is combined with extreme learning machine and produce forward extreme learning machine with kernel. This method is called Kernel Extreme Learning Machine with Kernel(KELM).

Extreme Learning Machine with Kernel

In KELM, the outputs of the hidden layer of ELM can be regarded as the nonlinear mapping of samples. When the mapping is an unknown, we can construct the kernel function instead of HH^T.

$$HH^T = \begin{bmatrix} K(x_1,x_1) & \cdots & K(x_1,x_N) \\ \vdots & \ddots & \vdots \\ K(x_N,x_1) & \cdots & K(x_N,x_N) \end{bmatrix}$$

$$h(x)\, H^T = \begin{bmatrix} K(x_1,x_1) \\ \vdots \\ K(x_1,x_N) \end{bmatrix} \quad (1)$$

The most popular kernel of KELM in use is the Gaussian kernel $K(x_i, x_j) = \exp(-\|x_i - x_j\|/\delta)$, where δ is the kernel parameter. Thus, the output weight matrix OV in KELM can be expressed as (4.13) and the Classification of formula of KELM can be expressed as (4.14):

$$OV = \left(\frac{I}{C} + HH^T\right)^{-1} . \ Y \quad (2)$$

$$f(x) = h(x)\, H^T OV = \begin{bmatrix} K(x_1,x_1) \\ \vdots \\ K(x_1,x_N) \end{bmatrix} \left(\frac{I}{C} + HH^T\right)^{-1} . \ Y \quad (3)$$

where C is a scale parameter which adjusts experiential risk and structural risk. Even though KELM has shown promising results because it chooses weights and biases of hidden nodes efficiently and obtains the output weights and biases analytically. In most cases, KELM is fast and presents good generalization, but the stability and generalization performance still can be improved. So instead of gaussian kernel, Bayesian "sum of kernels" model is introduced in KELM and named as Extreme Learning Machine with Multiple Kernels (MK-ELM).

Extreme Learning Machine with Multiple Kernels (MK-ELM)

In this method, the kernel function is represented by a weighted sum of kernel functions, which is constructed by a prior knowledge. The "sum-of-kernels" model explores the additive effects through the linear combination of the different generating functions. The goals using the Bayesian methods are obtaining a sparse representation with few kernels in the sum, and getting good kernel functions for regression and classification problems. This method optimizes the kernel function using a weighted sum of kernel functions by a prior knowledge. MK-ELM can be summarized in below algorithm

MK-ELM Algorithm

Input: Given a training set, a validation data, a set of kernel functions, the number of hidden nodes.
Training phase:
Step 1: Randomly assign the weights of input layer and the biases of output layer . Step 2: Randomly generate the initial parameter for the kernel function.
Step 3: For k=1,2......L
a) Calculate the hidden layer output matrix using the initial kernels.
b) Calculate the output weight
c) Calculate the error function using the network. And apply empirical Bayesian to approximate the marginal. Then update the parameter estimation of the kernel function.
Step 4: Using the Multiple Kernels to train Extreme Learning Machine.

Multi Kernel Improved ELM

Even though MK-ELM works better than KELM and ELM, it sums all of the kernel and produce weighted sum value. But the existing analysis proved that the improved learning produces very satisfactory result in classification approach. So this work introduce the improved learning in MK-KLM by not doing additive process of multiple kernel. Instead of adding the kernel use all the kernels separately in classification process. Finally the all the classification values are fused by AVG fusion approach to produce the Deep classification. So in Multi kernel Improved ELM the gaussian kernel, Bayesian kernel, Laplican kernel, Gaussian Radial Basis kernel and polynomial kernel are used separately and then multiple kernel ELM approach is proceeded to classify the features value. Finally all the classified values are fused by applying the average function to produce the final classified result. This approach produce high value because using improved learning.

Multi Kernel Improved ELM Algorithm

Input: Given a training set, a validation data, a set of kernel functions, the number of hidden nodes.
Training phase:
Step 1: Randomly assign the weights of input layer and the biases of output layer .
Step 2: Randomly generate the initial parameter for the kernel function.
Step 3: For k=1,2......L
a) Calculate the hidden layer output matrix using the initial kernels.
b) Calculate the output weight
c) Calculate the error function using the network. And apply five kernel function separately to approximate the marginal. Then update the parameter estimation of the kernel function.
Step 4: Using the Multiple Kernels to train Extreme Learning Machine.
Step 5: Apply average fusion approach to fuse all the classification value.

Retrieval Number: E6842018520/2020©BEIESP
DOI:10.35940/ijrte.E6842.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

IV. RESULT AND ANALYSIS

A. Efficiency Parameters

To assess the efficiency of the proposed ontology constructing process, several efficiency metrics are available. This paper employs the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate to analyses the efficiency.

Detection Accuracy

Detection Accuracy is the measurement system, which measure the degree of closeness of measurement between the original labeled texts and the correctly labeled texts

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (4.1)$$

Where, TP – True Positive
FN – False Negative
TN – True Negative
FP – False Positive

Error Rate

Error Rate is the measurement system, which measure no of falsely recognised characters form the given input character images.

$$\text{Error Rate} = \frac{\text{No of Images of Falsely labeled texts}}{\text{Total No of texts}}$$
(4.2)

Precision Rate

The precision is the fraction of retrieved instances that are relevant to the find.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4.3)$$

Where, TP – True Positive
FP – False Positive

Recall Rate

The recall is the fraction of relevant instances that are retrieved according to the input image.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3.3.4)$$

Where, TP – True
FN – False Negative

Sensitivity

Sensitivity also called the true positive rate or the recall rate in some field's measures the proportion of actual positives.

$$\text{Sensitivity} = \frac{TP}{(TP + FN)}$$

where, TP – True Positive (equivalent with hit)
FN – False Negative (equivalent with miss)

Specificity

Specificity measures the proportion of negatives which are correctly identified such as the percentage.

$$\text{Specificity} = \frac{TN}{(FP + TN)}$$

where, TN – True Negative (equivalent with correct rejection)
FP – False Positive (equivalent with false alarm)

F-Measure

F-measure is the ratio of product of precision and recall to the sum of precision and recall. The f-measure can be calculated as,

$$F_m = (1 + \alpha) * \frac{\text{Precision} * \text{Recall}}{\alpha * (\text{Precision} * \text{Recall})}$$

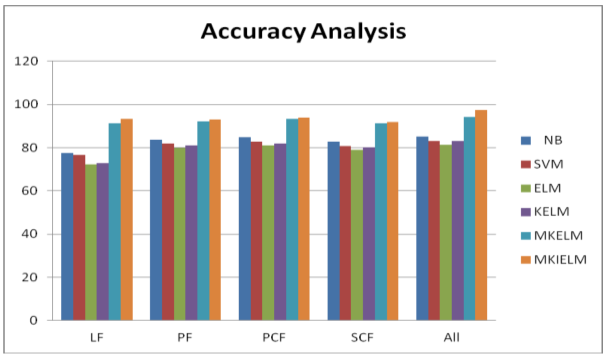
Experiment No #1 : Accuracy Analysis of Proposed Sentiment Analysis Model

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this sentiment analysis scheme, the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate measures are employed. It is shown in equation 4,5,6 and 7 correspondingly. Ideally, a excellent feature retrieval scheme is accepted to have a high Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure value. Table 1 lists the accuracy analysis of IDCNN.

Table 1: Detection Accuracy Analysis of Proposed Sentiment Analysis Model

Classifier						
Metrics	NB	SVM	ELM	KELM	MKELM	MKIELM
LF	77.34	76.72	72.29	72.79	91.11	93.21
PF	83.75	81.85	79.95	80.98	92.24	92.92
PCF	84.66	82.76	80.86	81.89	93.15	93.83
SCF	82.74	80.84	78.94	79.97	91.23	91.91
All	85.11	83.13	81.38	83.03	94.23	97.23

As observed from Table 1, the Accuracy of the MKIELM in range 93-97, which is superior than other methods. So the MKIELM classifier is considered to be the best for sentiment analysis. Fig.8 depicted the Detection Accuracy of classifier approaches.



As observed from above figure, the Accuracy of the MKIELM in range 93-97, which is superior than other method. So the MKIELM classifier are best for sentiment analysis.

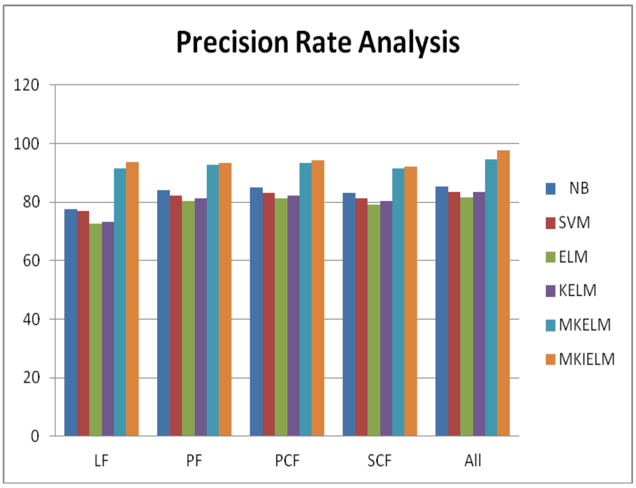
Experiment No #2 : Precision Rate Analysis of Proposed Sentiment Analysis Model

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this sentiment analysis scheme, the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate measures are employed. It is shown in equation 4,5,6 and 7 correspondingly. Ideally, a excellent feature retrieval scheme is accepted to have a high Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure value. Table 1 lists the Precision Rate analysis of MKIELM.

Table 2: Precision Rate Analysis of Proposed Sentiment Analysis Model

Classifier						
METRICS	NB	SVM	ELM	KELM	MKELM	MKIELM
LF	77.65	77.03	72.6	73.1	91.42	93.52
PF	84.06	82.16	80.26	81.29	92.55	93.23
PCF	84.97	83.07	81.17	82.2	93.46	94.14
SCF	83.05	81.15	79.25	80.28	91.54	92.22
All	85.42	83.44	81.69	83.34	94.54	97.54

As observed from Table 2, the Precision Rate of the MKIELM in range 92-96, which is superior than other method. So the MKIELM classifier is considered to be the best for sentiment analysis. Fig.8 depicted the Precision Rate measures of classifier approaches.



As observed from above figure, the Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure of the MKIELM in range 97-98, which is superior than other method. So the MKIELM classifier are best for sentiment creation.

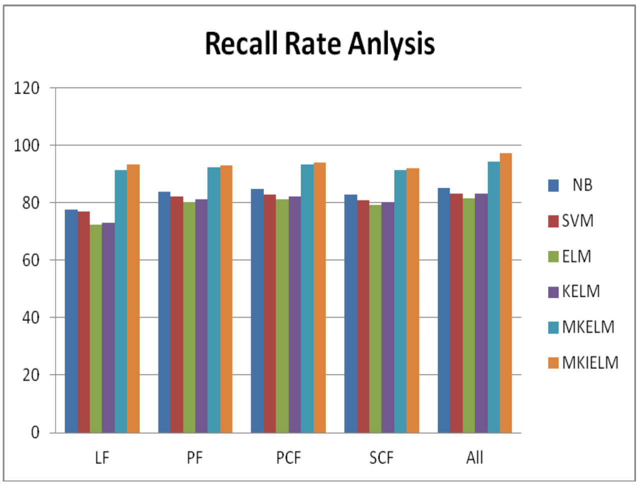
4.2.3 Experiment No #3 : Recall Rate Analysis of Proposed Sentiment Analysis Model

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this sentiment analysis scheme, the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate measures are employed. It is shown in equation 4,5,6 and 7 correspondingly. Ideally, a excellent feature retrieval scheme is accepted to have a high Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure value. Table 1 lists the Recall Rate analysis of MKIELM.

Table 3: Recall Rate Analysis of Proposed Sentiment Analysis Model

Classifier						
Metrics	NB	SVM	ELM	KELM	MKELM	MKIELM
LF	77.56	76.94	72.51	73.01	91.33	93.43
PF	83.97	82.07	80.17	81.2	92.46	93.14
PCF	84.88	82.98	81.08	82.11	93.37	94.05
SCF	82.96	81.06	79.16	80.19	91.45	92.13
All	85.33	83.35	81.6	83.25	94.45	97.45

As observed from Table 1, the Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure of the MKIELM in range 93-97, which is superior than other method. So the MKIELM classifier is considered to be the best for sentiment analysis. Fig.8 depicted the Recall Rate measures of classifier approaches.



As observed from above figure, the Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure of the MKIELM in range 93-97, which is superior than other method. So the MKIELM classifier are best for sentiment analysis.

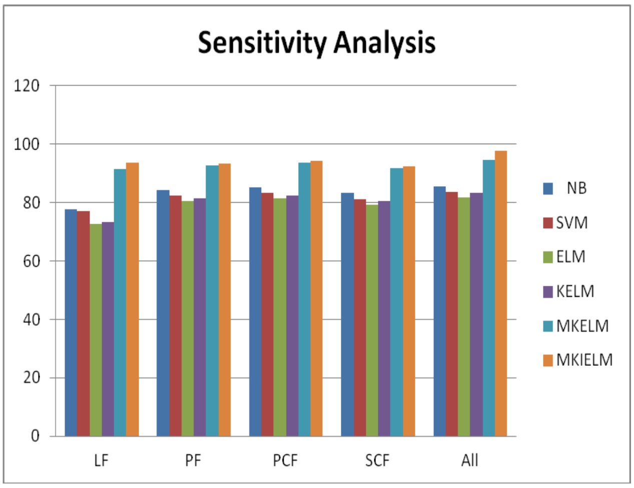
4.2.4 Experiment No #4 : Sensitivity Analysis of Proposed Sentiment Analysis Model

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this sentiment analysis scheme, the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate measures are employed. It is shown in equation 4,5,6 and 7 correspondingly. Ideally, a excellent feature retrieval scheme is accepted to have a high Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure value. Table 1 lists the accuracy analysis of MKIELM.

Table 4: Sensitivity Analysis of Proposed Sentiment Analysis Model

Classifier / Metrics	NB	SVM	ELM	KELM	MKELM	MKIELM
LF	77.75	77.13	72.7	73.2	91.52	93.62
PF	84.16	82.26	80.36	81.39	92.65	93.33
PCF	85.07	83.17	81.27	82.3	93.56	94.24
SCF	83.15	81.25	79.35	80.38	91.64	92.32
All	85.52	83.54	81.79	83.44	94.64	97.64

As observed from Table 1, the Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure of the MKIELM in range 93-97, which is superior than other method. So the MKIELM classifier is considered to be the best for sentiment analysis. Fig.8 depicted the Sensitivity measures of classifier approaches.



As observed from above figure, the Sensitivity of the MKIELM in range 93-97, which is superior than other method. So the MKIELM classifier are best for sentiment analysis.

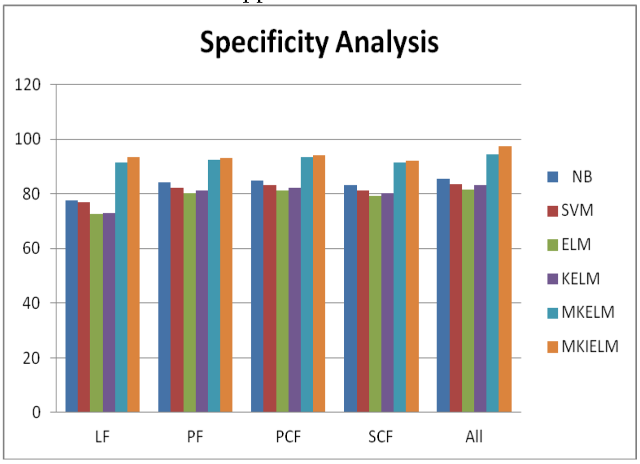
4.2.5 Experiment No #5 : Specificity Analysis of Proposed Sentiment Analysis Model

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this sentiment analysis scheme, the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate measures are employed. It is shown in equation 4,5,6 and 7 correspondingly. Ideally, a excellent feature retrieval scheme is accepted to have a high Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure value. Table 1 lists the Specificity analysis of MKIELM.

Table 5: Specificity Analysis of Proposed Sentiment Analysis Model

Classifier / Metrics	NB	SVM	ELM	KELM	MKELM	MKIELM
LF	77.67	77.05	72.62	73.12	91.44	93.54
PF	84.08	82.18	80.28	81.31	92.57	93.25
PCF	84.99	83.09	81.19	82.22	93.48	94.16
SCF	83.07	81.17	79.27	80.3	91.56	92.24
All	85.44	83.46	81.71	83.36	94.56	97.56

As observed from Table 1, the Specificity, F-Measure of the MKIELM in range 93-97, which is superior than other method. So the MKIELM classifier is considered to be the best for sentiment analysis. Fig.8 depicted the Specificity measures of classifier approaches.



As observed from above figure, the Specificity of the MKIELM in range 97-98, which is superior than other method. So the MKIELM classifier are best for sentiment analysis creation.

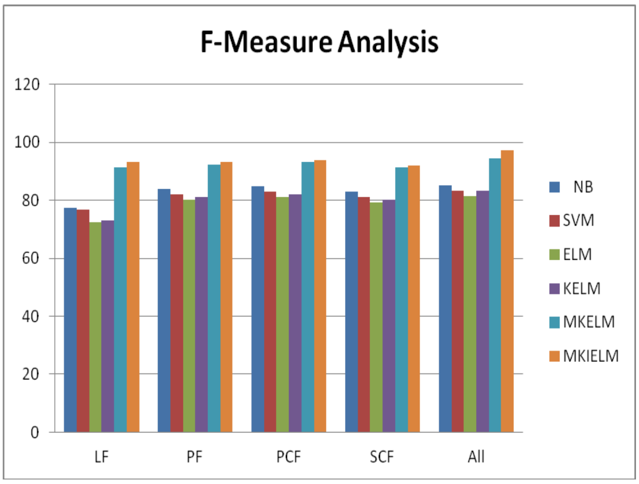
4.2.6 Experiment No #5 : F-Measure Analysis of Proposed Sentiment Analysis Model

In this experiment, we will assess the contribution of each classifier approaches which are employed in the work. To assess the efficiency of this sentiment analysis scheme, the Detection Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure and Error Rate measures are employed. It is shown in equation 4,5,6 and 7 correspondingly. Ideally, a excellent feature retrieval scheme is accepted to have a high Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure value. Table 1 lists the F-Measure analysis of MKIELM.

Table 6: F-Measure Analysis of Proposed Sentiment Analysis Model

Classifier						
Metrics	NB	SVM	ELM	KELM	MKELM	MKIELM
LF	77.49	76.87	72.44	72.94	91.26	93.36
PF	83.9	82	80.1	81.13	92.39	93.07
PCF	84.81	82.91	81.01	82.04	93.3	93.98
SCF	82.89	80.99	79.09	80.12	91.38	92.06
All	85.26	83.28	81.53	83.18	94.38	97.38

As observed from Table 1, the Accuracy, Precision Rate, Recall Rate, Sensitivity, Specificity, F-Measure of the MKIELM in range 94-98, which is superior than other method. So the MKIELM classifier is considered to be the best for sentiment analysis. Fig.8 depicted the F-Measure measures of classifier approaches.



As observed from above figure, the F-Measure of the MKIELM in range 94-98, which is superior than other method. So the MKIELM classifier are best for sentiment analysis.

V. CONCLUSION

In this document, we present the emotion classification method that uses the psychological and linguistic features that LIWC receives when analyzing emotions on Twitter. To this end, the five main types of LIWC (namely, linguistic processes, psychological processes, personal concerns, spoken categories and punctuation) and those combinations are taken into account. Experimental analysis with the ranking algorithm shows that the psychopathology set can provide encouraging results when analyzing the sentiment of Twitter data. Experimental analysis shows that the band is more than individual. For Twitter sentiment analysis, the highest predictive performance (98.31%) was achieved by combining the language and mental processes of MKIELM. Therefore, this proposed method and combination of language processes, psychological processes, perform best in emotion analysis.

REFERENCES

1. A. Onan, “Twitter mesajları üzerinde makine öğrenmesi yöntemlerine dayalı duygu analizi”, Yönetim Bilişim Sistemleri Dergisi, Vol. 3, No. 2, 2017, pp. 1-14.
2. A. Onan, S. Korukoğlu, and H. Bulut, “A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification”, Expert Systems with Applications, Vol.62, 2016, pp.1-16.
3. A.Onan, “A machine learning based approach to identify geo-location of Twitter users”, in Proceedings of the ICC 2017, UK, 2017, pp.1-7.
4. J. Mahmud, J. Nichols, and C. Drews, “Home location identification of twitter users”, ACM Transactions on Intelligent Systems and Technology, Vol. 5, No.3, 2014, pp.47.
5. Z. Cheng, J. Caverlee, and K.Lee, “You are where you tweet: a contentbased approach to geo-location twitter users”, in Proceedings of the 19th ACM International Conference on Information and Knowledge Management, USA, 2010, pp.759-768.
6. B.Hecht, L.Hong, B. Suh and E.D.Chi, “Tweets from Justin Bieber’s heart: the dynamics of the location field in user profiles”, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, USA, 2011, pp.237-246.
7. A. Onan and S. Korukoğlu, “Makine öğrenmesi yöntemlerinin görüş madenciliğinde kullanılması üzerine bir literatür araştırması”, Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, Vol. 22, No. 2, 2016, pp. 111-122.
8. W. Medhat, A. Hassan and H. Korashy, “Sentiment analysis algorithms and applications: a survey”, Ain Shams Engineering Journal, Vol. 5, No. 4, 2014, pp. 1093-1113.
9. A. Onan and S. Korukoğlu, “A feature selection model based on genetic rank aggregation for text sentiment classification”, Journal of Information Science, Vol. 43, No.1, 2017, pp.25-38.
10. M.P. Salas-Zarate, E.Lopez-Lopez, R.Valencia-Garcia, N. Gilles, A.Almela and G.Alor-Hernandez, “A study on LIWC categories for opinion mining in Spanish reviews”, Journal of Information Science, Vol.40, No.6, 2014, pp.749-760.
11. M.Salas-Zarate, M.A. Paredes-Valverde, M.A.Rodriguez-Garcia, R.Valencia-Garcia and G.Alor-Hernandez, “Automatic detection of satire in Twitter: a psycholinguistic-based approach”, KnowledgeBased Systems, Vol.128, 2017, pp.20-33.
12. J.M.Cotelo, F.L.Cruz, J.A.Troyano and F.J.Ortega, “A modular approach for lexical normalization applied to Spanish tweets”, Expert Systems with Applications, Vol. 42, No.10, 2015,pp. 4743-4754.
13. A. Esuli and F. Sebastiani, ‘SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining’, in In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC’06, 2006, pp. 417–422.
14. R. Valitutti, ‘WordNet-Affect: an Affective Extension of WordNet’, in In Proceedings of the 4th International Conference on Language Resources and Evaluation, 2004, pp. 1083– 1086.
15. F. L. Cruz, J. A. Troyano, B. Pontes, and F. J. Ortega, ‘ML-SentiCon: Un lexicón multilingüe de polaridades semánticas a nivel de lemas’, Procesamiento del Lenguaje Natural, vol. 53, pp. 113–120, 2014.
16. M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, ‘Lexicon-Based Methods for Sentiment Analysis’, Comput. Linguist., vol. 37, no. 2, pp. 267–307, Apr. 2011.

Retrieval Number: E6842018520/2020©BEIESP
DOI:10.35940/ijrte.E6842.038620
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

17. M. Ghosh and K. Animesh, 'Unsupervised Linguistic Approach for Sentiment Classification from Online Reviews Using SentiWordNet 3.0.', International Journal of Engineering Research and Technology, vol. 2, no. 9, 2013.
18. V. Perez-Rosas, C. Banea, and M. Rada, 'Learning Sentiment Lexicons in Spanish', LREC, 2012.
19. S. Clematide and K. Manfred, 'Evaluation and extension of a polarity lexicon for German', presented at the Proceedings of the First Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, 2010, pp. 7–13.
20. I. Maks and P. Vossen, 'Different approaches to automatic polarity annotation at synset level', presented at the Proceedings of the First International Workshop on Lexical Resources, 2011, pp. 62–69.
21. A. Mageed, Muhammad, and M. Diab, 'Toward building a large-scale Arabic sentiment lexicon', presented at the Proceedings of the 6th International Global WordNet Conference, 2012, pp. 18–22.
22. I. Dehdarbehbahani, A. Shakery, and H. Faili, 'Semi-supervised word polarity identification in resource-lean languages', Neural Netw., vol. 58, pp. 50–59, Oct. 2014.
23. M.-T. Martín-Valdivia, E. Martínez-Cámara, J.-M. Perea-Ortega, and L. A. Ureña-López, 'Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches', Expert Syst. Appl., vol. 40, no. 10, pp. 3934–3942, Aug. 2013.
24. A. Balahur and M. Turchi, 'Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis', Comput. Speech Lang., vol. 28, no. 1, pp. 56–75, Jan. 2014.
25. R. Hsu, B. See, and A. Wu, 'Machine Learning for Sentiment Analysis on the Experience Project', 2010.
26. P. P. B. Filho, T. A. Pardo, and S. M. Alusio, 'An evaluation of the brazilianportugueseliwc dictionary for sentiment analysis', presented at the In 9th Brazilian Symposium in Information and Human Language Technology, Fortaleza, Ceara, 2013.
27. P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha, 'Comparing and Combining Sentiment Analysis Methods', in Proceedings of the First ACM Conference on Online Social Networks, New York, NY, USA, 2013, pp. 27–38.
28. C. J. Hutto and E. Gilbert, 'Vader: A parsimonious rule-based model for sentiment analysis of social media text', presented at the Eighth International AAAI Conference on Weblogs and Social Media, 2014.
29. M. del P. Salas-Zárate, E. López-López, R. Valencia-García, N. Aussenac-Gilles, Á. Almela, and G. Alor-Hernández, 'A study on LIWC categories for opinion mining in Spanish reviews', J. Inf. Sci., vol. 40, no. 6, pp. 749–760, Dec. 2014.
30. Q. Ye, Z. Zhang, and R. Law, 'Sentiment classification of online reviews to travel destinations by supervised machine learning approaches', Expert Syst. Appl., vol. 36, no. 3, Part 2, pp. 6527–6535, Apr. 2009.
31. G. Sidorov, S. Miranda-Jiménez, F. Viveros-Jiménez, A. Gelbukh, N. Castro-Sánchez, F. Velásquez, I. Díaz-Rangel, S. Suárez-Guerra, A. Treviño, and J. Gordon, 'Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets', in Advances in Artificial Intelligence, I. Batyrshin and M. G. Mendoza, Eds. Springer Berlin Heidelberg, 2013, pp. 1– 14.
32. E.Kontopoulos, C.Berberidis, T.Dergiades and N.Bassiliades, "Ontolog-based sentiment analysis of twitter posts", Expert Systems with Applications, Vol.40, No.10, 2013, pp.4065-4074.
33. R.Justo, T.Corcoran, S.M.Lukin, M.Walker and M.I.Torres, "Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web", Knowledge-Based Systems, Vol. 69, 2014, pp.124-133.
34. S.Skalicky and S.Crossley, "A statistical analysis of satirical Amazon.com product reviews", European Journal of Humour Research, Vol.2, 2015, pp.66-85.
35. J.W.Pennebaker, R.L.Boyd, K.Jordan and K.Blackburn, "The development and psychometric properties of LIWC 2015"

AUTHORS PROFILE

B. S. Srinidhi, is a research scholar in Jain University. He is working on the text classification and Sentiment analysis. He is also a web developer with .Net technologies His interests are in the area of natural language processing and Emotional Intelligence , He find it very interesting to discover patterns in the data relating to Emotional behavior of people.

R. Suchithra, is the director of MCA department in Jain Deemed to be University, Bangalore . She has 16 years of experience in teaching and research .She has published more than 24 papers in reputed journals and her area of specialization is on cloud computing , machine learning and data science . She is guiding research scholars in the PhD programme