

Word Sense Disambiguation Based Sentiment Classification Using Linear Kernel Learning Scheme

P. Ramya^{1,*} and B. Karthik²

¹Anna University, Chennai, 600025, Tamil Nadu, India

²Sona College of Technology, Salem, 636005, Tamilnadu, India

*Corresponding Author: P. Ramya. Email: ramya01phd@gmail.com

Received: 21 December 2021; Accepted: 24 February 2022

Abstract: Word Sense Disambiguation has been a trending topic of research in Natural Language Processing and Machine Learning. Mining core features and performing the text classification still exist as a challenging task. Here the features of the context such as neighboring words like adjective provide the evidence for classification using machine learning approach. This paper presented the text document classification that has wide applications in information retrieval, which uses movie review datasets. Here the document indexing based on controlled vocabulary, adjective, word sense disambiguation, generating hierarchical categorization of web pages, spam detection, topic labeling, web search, document summarization, etc. Here the kernel support vector machine learning algorithm helps to classify the text and feature extract is performed by cuckoo search optimization. Positive review and negative review of movie dataset is presented to get the better classification accuracy. Experimental results focused with context mining, feature analysis and classification. By comparing with the previous work, proposed work designed to achieve the efficient results. Overall design is performed with MATLAB 2020a tool.

Keywords: Text classification; word sense disambiguation; kernel support vector machine learning algorithm; cuckoo search optimization; feature extraction

1 Introduction

Word Sense Disambiguation (WSD) based text mining approaches are performed in variety of application like spam detection, email, topic labeling, web search, document summarization, etc. Text document mining uses the machine learning and deep learning approaches. Sentimental analysis and various data mining techniques are performed with novel machine learning approaches. In the Natural Language Processing (NLP), WSD has been depicted as the assignment which chooses the fitting significance to given word in content where this importance is discernible from other possible word [1]. The text document mining is to eliminate the noise in a number of text records, for example, the immaterial and repetitive features. Accordingly, the measurement decrease strategies have been reviewed to tackle this issue, including feature extraction and selection [2]. Feature extraction strategy separate the



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

characteristics from the new and low-dimensional element space that is changed from unique element space, for example, Principle Component Analysis, Linear Discriminate Analysis, etc. Research is to realize feature analysis techniques, which allot the element weightings that are frequently standardized in the reach of (0, 1) to each element, at that point rank the element weightings in slipping request [3]. The lexical or word sense uncertainties are the major data mining models, which is identified with the sense of words [4].

WSD is the issue of allotting the fitting significance (sense) to a given word in content where this importance is discourse from different faculties possibly owing to that word [5]. Information based frameworks are acquiring significance in the words errands of last Senseval occasions to label words with a low number of preparing models [6]. Most frameworks consolidate both information based and corpus-based methodologies when all-words information must be prepared. They gain proficiency with those words for which marked models are accessible; and, apply a separately for those words without enough preparing models [7]. WSD task emerges during the lexical scientific classification development [8]. Here by lexical scientific classification, the comprehend word sense and edges address hyponymy relations are determined. Scientific categorization structures of numerous semantic assets are lexical information bases, thesaurus, and ontology [9,10]. Such assets are broadly utilized: they are crucial as assortments of word sense portrayals just as a component of NLP undertakings, for example, fabricating a dataset of semantic likeness or as an apparatus for term speculation [11]. Text Document Classification (TDC) has an incredible importance in data handling and recovery where unstructured records are coordinated into pre-characterized classes [12]. Deep Neural Network (DNN) investigation in NLP interpretation is extremely delayed to prepare and it has been seen that for straightforward content characterization issues traditional ML approaches to give comparable outcomes with snappier computation time [13].

WSD techniques utilized for the overall corpora are appropriate for this sort of WSD errands. Consequences of applying an existing WSD strategy to the errand of hypernym disambiguation in a monolingual word reference definition [14]. The neural organizations entered the NLP. In this, the WSD dependent on Long Short Term Memory (LSTM) neural network presents a coarse model of sentences consecutively. This prepared to anticipate a covered word subsequent to perusing a sentence [15]. The word sense inserting vector to be utilized in WSD task is gotten from the inward portrayal of the word in the organization. To restrict the extent of the work we confined ourselves to only three models: Lesk model as a benchmark, Skip-gram in Word2Vec usage as a best in class WSD model and AdaGram as a cutting edge word sense enlistment model. This field turns out to be more challenging because of the way that many requesting and fascinating exploration issues actually exist in this field to settle. Slant based investigation of an archive is very hard to act in examination with theme based content characterization. The assessment words and assumptions consistently change with circumstances. Consequently, an assessment word can be considered as sure in one situation yet may get negative in some other condition. Research contributes to improve with real time application-based data processing and improved with feature analysis. In this proposed work, the kernel support vector machine classification and feature extraction of text is done by cuckoo search optimization to improve the overall opinion mining and also the prediction of emotions for reviews. Here the movie dataset is used to get the positive and negative review classification with efficient classification results.

This paper summarized as follows. Section II describes the literature survey of various related work. Section III presents the proposed work overview, algorithm and processing flows. Section IV discusses the experimental results and various parameter analyses. Section V decided with the design of text mining and result analysis; finally it presents the future enhancement approach discussion.

2 Literature Survey

Jin Zheng et al. (2019) has presented the bidirectional **Convolutional Neural Networks** with recurrent model for text classification. This combination of two models determines the feature as hybrid analysis. This directs on four datasets, including Yahoo Answers, Sogou News of the theme order, Yelp Reviews, and Douban Movies Top250 short surveys of the opinion examination. The text classification task is a significant application in regular language handling. The convolutional neural network and recurrent neural network have accomplished better outcomes for this assignment, yet the multi-class text arrangement and the fine-grained opinion investigation are as yet testing. The model joins the bidirectional LSTM and CNN with the consideration instrument and word2vec to accomplish the fine-grained text arrangement task. In this, the word2vec is to produce word vectors naturally and a bidirectional repetitive design to catch logical data and long haul reliance of sentences. Feature processing is not sufficient in this paper which results the poor classification accuracy.

Xiaohan et al. (2018) has presented the Phase-shift oscillator based feature analysis on text mining approach. The text mining applications majorly includes the choice of feature analysis on the content dataset, which are valuable for order. The strategies utilized Reuter-21578 dataset. Feature analysis ought to be eliminated to build the exactness and reduction of intricacy and running time; however it is regularly a costly cycle, and most existing techniques utilizing a basic channel to eliminate feature. Moreover, there is little exploration utilizing PSO to choose features for text grouping. This methodology utilizing a two-stage strategy for text feature determination, where with the features chose by four distinctive channel positioning strategies at the principal stage, more superfluous features are taken out by PSO to form the last element subset. PSO based optimization falls under local optima issue and also it is time consuming.

Hiram et al. (2019) has presented the WSD using DNN is designed for text mining approaches. The approaches dependent on neural network models to take the issue of disambiguation of the significance of words utilize a bunch of classifiers toward the end, which brings about a specialization in a solitary arrangement of words-those for which they were prepared. This makes difficult to apply the learned models to words not recently found in the preparation corpus. This tries to address a speculation of the issue of WSD to settle it through DNN without restricting the technique to a fixed arrangement of words, with an exhibition near the best in class, and a satisfactory computational expense. Various structures on multilayer perceptrons, Long Short-Term Memory and Gated Recurrent Units classifier models are designed. Various sources and measurements of embeddings were tried also. The principle assessment was performed on the Senseval 3 English Lexical Sample. To assess the application to a concealed arrangement of words, learned models are assessed in the totally inconspicuous expressions of an alternate corpus, beating the arbitrary pattern.

However, the proposed method is used for requiring the manually annotated training data for each word that required to disambiguate. Moreover, the manual annotation was very expensive, complex and also time consuming to apply for large volume of datasets.

Jibing et al. (2020) has proposed the deep learning with multi-label classification of text data is performed with graph transformer model. Conventional strategies for multi-mark text arrangement, especially profound learning, have accomplished astounding outcomes. Notwithstanding, the vast majority of these strategies use word2vec innovation to address consecutive content data, while disregarding the rationale and inner pecking order of the actual content. Here the Hierarchical Graph Transformer based deep learning model for enormous scope multi-label texting. Initially, the structure can epitomize the diverse semantics of the content and the associations between them. Here the multi-layer transformer structure with a multi-label consideration component uses the word, sentence, and chart

levels to completely get the features of the context. At last, various leveled relationship are dependent on the semantic distances of the marks. Analyzing the correlation between two variables is little difficult.

Muhammad et al. (2020) has proposed the CNN based text classification with single later analysis. Urdu is the most loved exploration language in South Asian dialects in light of its mind boggling morphology, interesting highlights, and absence of phonetic assets like standard datasets. When contrasted with short content, similar to semantics, long content order needs additional time and exertion due to huge jargon, more clamor, and repetitive data. AI and Deep Learning models have been generally utilized in content preparing. In spite of the significant constraints of Machine Learning models, as learn coordinated features are the most loved techniques for Urdu TDC. To the most amazing aspect our insight, it is the primary investigation of Urdu TDC utilizing DL model. Single-layer Multisize Filters Convolutional Neural Network (SMFCNN) for order and contrast its exhibition and sixteen ML standard models on three imbalanced datasets of different sizes. The multiscale CNN increases the time complexity and also consumes more storage space.

Hai et al. (2020) has presented the feature analysis based nonequilibrium Bi-LSTM is modeled for text mining. This model uses Chinese content arrangement dependent on an element upgraded nonequilibrium bidirectional long short term memory (Bi-LSTM) network. That breaks down Chinese content data top to bottom and improves the exactness of text grouping. To start with, the bidirectional encoder portrayals from transformers model to vectorize the first Chinese corpus and concentrate primer semantic features. At that point, a nonequilibrium Bi-LSTM network was applied to expand the heaviness of text data containing significant semantics and further improve the impacts of the critical features in Chinese content characterization. The Bi-LSTM process is very CPU and memory intensive.

Shajari et al. (2020) presented the forex prediction approach for word sense disambiguation application in which positive and negative sentiments are determined. This is designed for the news headlines and the prediction of market approach. This paper determines the proper sense of all significant words for the headlines of news application. Different word sense disambiguation approaches are proposed for evaluation such as similarity threshold, relevant gloss retrieval, verb normalization, etc. Similarly, proper sentences are determined for the important words for text contents that enhance the sentiment class prediction. Automatic classification of sentiments for the text values to the word sense disambiguation requires the process of it for big data based applications. Different study approaches are presented for word sense disambiguation such as Supervised, Semi-Supervised and Unsupervised. Sign language is presented to apply for sentiment analysis in which sense of the text is analyzed for big data public applications (Naranjo-Zeledon et al. (2019)). Author anuratha et al. (2020) has proposed the sentiment analysis for public mediums for the word sense disambiguation for Twitter applications. One of the real-time applications such as COVID-19 pandemic analysis is implemented by this system.

3 Proposed Method

The proposed text classification uses the linear kernel Support Vector Machine Learning algorithm. Here the feature extraction plays a vital role in this text mining. Selection of appropriate feature extraction technique gives the success of using a machine learning algorithm in text document classification and constructing an effective classifier model using machine learning algorithms help in predicting the labels of the input text documents. Here the movie review dataset is used for classification of positive and negative reviews by analyzing features.

The overall process flow of proposed work is given in the [Fig. 1](#). In this, the SVM, cuckoo search algorithm with WSD plays a vital role to improve the performance. The words from Movie review dataset accomplished the proposed work to classify the positive and negative values. Here the Word2vec takes as its information a huge corpus of text and produces a vector space with every novel word being

relegated a comparing vector in the space. Word vectors are situated in the vector space with the end goal that words share basic settings in the corpus are situated in closeness to each other in the space.

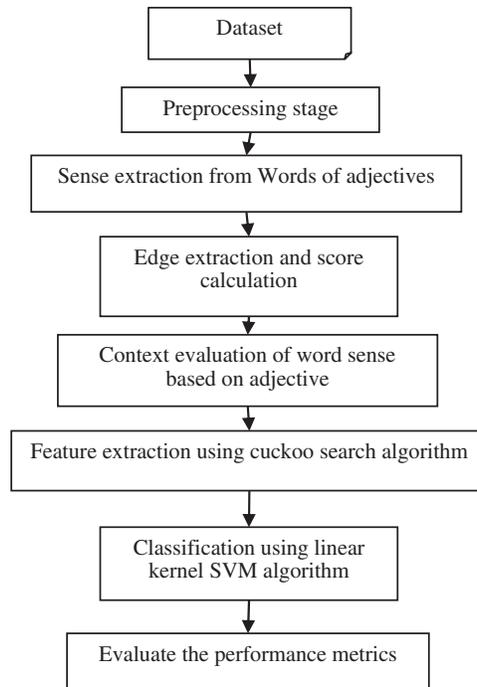


Figure 1: Process block of proposed work

In NLP, the issue of finding the proposed importance or "sense" of a word which is initiated by the utilization of that word in a specific setting is by and large set known as word sense disambiguation issue. This issue impacts different fields of characteristic language preparing, which includes semantic examination and machine interpretation. Here, WSD issue is displayed as a combinatorial improvement issue where the objective is to discover an arrangement of implications that augments the semantic significance among the focused on words that is adjective findings.

Feature extraction words are chosen from the setting to decide the correct feeling of the uncertain word from positive and negative review of movie datasets. Information based WSD as a rule chooses words in a specific length of dataset and it select the feature as adjective word of sentence. At that point, as indicated by the relatedness between feature words and each feeling of uncertain word, the sense with greatest relatedness is chosen as the correct sense.

3.1 Dataset

The experiment is conducted with IMDB dataset. It is classified into two categories. They are positive and negative categories. It consists of 1000 positive reviews and 1000 negative reviews of the movies. The determination of the dataset collection is finished and thorough and ought to be firmly identified with layer assessment. The dataset ought to be partitioned into two pieces of training and testing, which is directed standardization handling to stay away from Figureuring immersion wonder.

3.2 Preprocessing

Initial stage of text mining performs the preprocessing stage to reduce the noise of context by performing tokenization, normalization, stemming, stop word removal and overall noise removal.

1. Tokenization
2. Normalization
3. Stemming
4. Stop word removing
5. Noise removal

Here the tokenization performed with the raw data and this token of document ignores the punctuation, numeric and spaces. The semantic unit of WSD determines the similarities. Uniqueness of the words is performed to get the better text mining. The stop word removing process helps to remove the verbs and conjunction like less informative words on sentence. Stemming process performs the stop word to limit the features. The component choice stage fundamentally helps in refining text feature, which are considered as contribution for arrangement task. This determined unquestionably a useful errand thought about dependent on the test result. They have applied to match information of feature analysis strategy with SVM classifier in the space of movie dataset.

3.3 Kernel Support Vector Classification

The entire system is organized into four major modules namely, Preprocessing, Learning, Classification and Evaluation. The preprocessing stage involves the techniques and processes which completes task of text mining. The kernel SVM is formulated by the training and testing modules which indeed represents the learning and classification tasks. Finally the evaluation phase measures the efficiency and performance of the system. Hyperplane SVM with labeled classes is given in the Fig. 2. After the learning cycle, a model is made and kept in touch with a record for mining. For the testing stage, the learned model is perused with and the testing design name model sets are stacked with. At that point, it emphasizes over all the testing models, characterizes every model, composes the name to a document, finding the deficiency of this model, and afterward may assess the forecast and collect measurements.

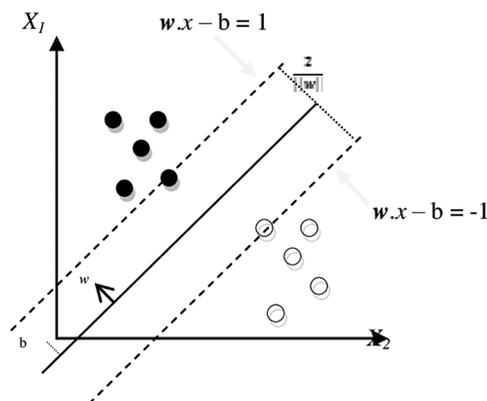


Figure 2: Hyperplane with two labeled classes

SVM is intended for paired arrangement issue, yet many analysts have chipped away at multi-class issue utilizing this legitimate procedure. By and large, messages and reports are unstructured informational indexes. In any case, these unstructured content arrangements should be changed over into an organized component space when utilizing numerical displaying as a feature of a classifier. To begin with, the information should be cleaned to discard superfluous characters and words. After the information has been cleaned, formal element extraction strategies can be applied. The classification label is given as,

$$(a, b)_1, \dots, (a, b)_n, a_n \in K, b_n \in [-1, +1] \tag{1}$$

The hyper plane of SVM determined with the vector function, constant determination and weight function, which is given as,

$$w.a + x = 0 \tag{2}$$

The SVM predicting the multivariate yields to perform directed learning by approximating planning utilizing named preparing tests. The SVM is retrained with the new preparing set, and this technique is preceded iteratively until maximum iteration reached. At every cycle, new semi labeled test reviews are added to the training set developed in the past emphasis, while the semi labeled tests that change their name are taken out from the new semi labeled set and moved back to the unlabeled set. The linear kernel SVM categorized with adjective feature of positive and negative classifications. This helps to the speed computation and improves the results.

3.4 Cuckoo Search Optimization

Feature extraction process of text mining is performed using cuckoo search optimization algorithm. Optimization strategy depends on population of arrangements or a global arrangement of overall performance. Population based calculations investigate a wide territory of a given pursuit space. Designing the harmony among misuse and investigation is important to arrive at the worldwide ideal arrangement. This model is executed utilizing the multitude knowledge search calculation of WSD. This calculation boosts the semantic relatedness and likeness relying upon different strategies that measure the previously mentioned models.

The cuckoo search calculation depends on the parasitic proliferation system of cuckoo bird and Levy flights search guideline. It is principally founded on the accompanying three ideal principles:

1. Each cuckoo lays one egg at one time and selects a home arbitrarily to bring forth it.
2. The best nest will be saved to the cutting edge in an arbitrarily chosen gathering of homes.
3. The number of homes accessible is fixed, and the likelihood that the host bird of a home will locate the outlandish bird's egg is $P_a [0,1]$. Objective function evaluation is given by,

$$f(a) = a_1 + a_2 \dots a_K \tag{3}$$

Fitness value of cuckoo search algorithm is given as,

$$f_{a,b} = f(\min)_b + \gamma_{ai}(f(\max)_b - f(\min)_b) \tag{4}$$

For the aspect of insight cuckoo search optimization has not been executed in the semantic disambiguation area.

Flow chart of cuckoo search optimization algorithm is given in Fig. 3. Here the Levy flight has a place with one kind of arbitrary walk, and the strolling step fulfills a steady circulation of tailed followed. In this type of cuckoo strolling, distance investigation and infrequent significant distance walk substitutes are determine. Levy flight in wise streamlining calculation can grow search scope, increment population variety and take it simpler to leap out of neighborhood optimum. In any case, it shows compelling execution on comparative spaces, like, grammatical feature labeling like adjective separation. This has propelled us to actualize Cuckoo search optimization (CSO) for the semantic disambiguation where it has normal errand with CSO labeling issue.

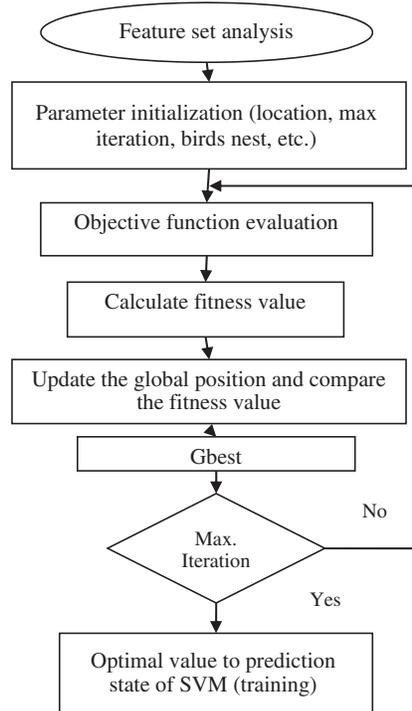


Figure 3: Process flow of cuckoo search algorithm

3.5 Performance Metrics

Performances are analyzed with classification stage of text mining with precision, recall, F measure and accuracy. The calculations are performed based on the accurate text with feature extracted text. The weight harmonics of the precision and recall shows the F score.

$$Precision = \frac{\text{relevant documents retrieved}}{\text{retrieved documents}} \quad (5)$$

$$Recall = \frac{\text{relevant documents retrieved}}{\text{relevant documents}} \quad (6)$$

$$sure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

FMea

where, TN and TP represent the true positive rate and true negative rate respectively. FN and FP represents the false positive and false negative respectively. By evaluation these results, the MATLAB 2020a tool helps to determine various analysis over proposed work. Here the accuracy and precision is the major factor of text mining and it obtained the better performance results.

4 Result and Discussion

Thus the performance classification of text mining based on movie dataset using linear kernel SVM with cuckoo search algorithm achieves the better performances. Here the feature extraction process improves the

performance of WSD, which uses the swarm intelligence based cuckoo search optimization algorithm. Tokenization is the way toward separating a flood of text into words, expressions, images, or some other significant components called tokens. The primary objective of this progression is to separate individual words in a sentence. Sentences can contain a combination of capitalized and lower case letters. Various sentences make up a content record. To decrease the issue space, the most widely recognized methodology is to diminish everything to bring down case. Confusion matrix determination is given below [Tab. 1](#).

Table 1: Confusion matrix

| Analysis | Relevant | Not Relevant |
|---------------|---------------------|---------------------|
| Retrieved | True Positive (TP) | False Positive (FP) |
| Not retrieved | False Negative (FN) | True Negative (TN) |

In SVM classifier, the negative review has a score of ≤ 4 out of 10, and a positive survey has a score of ≥ 7 out of 10. Neutral reviews were excluded from this dataset. SentiWordNet is one such dictionary which characterizes every equivalent set inside WordNet with assessment scores and direction. As NLP is vague, an accurate sense for a word in SentiWordNet should be arranged by the setting wherein the word happens. The SVM utilizes a vector machine to group information. SVM is especially appropriate for use with wide datasets, that is, those with countless indicator fields. This can utilize the default settings on the hub to create an essential model moderately rapidly, or you can utilize the Expert settings to try different things with various sorts of SVM models.

[Tab. 2](#) shows the proposed results comparison of overall performance with various methods M1, M2, M3, and M4. Hence, the proposed model is well-organized in performance metrics compared with other previous models. Notwithstanding, most review-based arrangement assignments remove wistful words from SentiWordNet without managing word sense disambiguation (WSD), however linear kernel SVM embrace the conclusion score of the primary sense or normal sense. SVM have been used to characterize records into positive and negative extremity (i.e., feeling) classes. Investigations show that we accomplish condition of workmanship execution in assumption investigation of standard movie dataset.

Table 2: Results comparison

| Model | Precision | Recall | F score | Accuracy |
|---|-----------|--------|---------|----------|
| BoW weighted by TF-IDF with SVD NB model -M1 | 0.81 | 0.80 | 0.79 | 0.80 |
| BoW weighted by TF-IDF with SVD Linear SVC Model -M2 | 0.70 | 0.70 | 0.69 | 0.70 |
| BoW weighted by TF-IDF with SVD LR model-M2 | 0.72 | 0.80 | 0.80 | 0.80 |
| Doc2Vec weighted by TFIDF with SVD NB model-M3 | 0.94 | 0.78 | 0.82 | 0.82 |
| Doc2Vec weighted by TFIDF With SVDLinear SVC model-M4 | 0.75 | 0.69 | 0.74 | 0.73 |
| Proposed work | 0.94 | 0.9 | 0.9 | 0.90 |

IMDb is perhaps the most well-known online information bases for motion pictures and characters, a stage where a large number of clients peruse and compose film surveys. This gives a huge and different dataset for assumption investigation. In this, to actualize diverse order models to foresee the supposition of IMDb surveys, as sure or negative, utilizing just content each survey contains. The objective is to

locate the model with the most elevated F1 score also, best speculation. We prepared various models utilizing numerous text features extraction and hyper-boundary settings for both the classifiers and the highlights, which might really touch the display altogether. The graphical analysis of overall performance comparison is given in the Fig. 4. The component vector at this point comprises predominantly of assessment words. A SVM classifier is applied on these vectors, which characterizes records in to positive and negative classes.

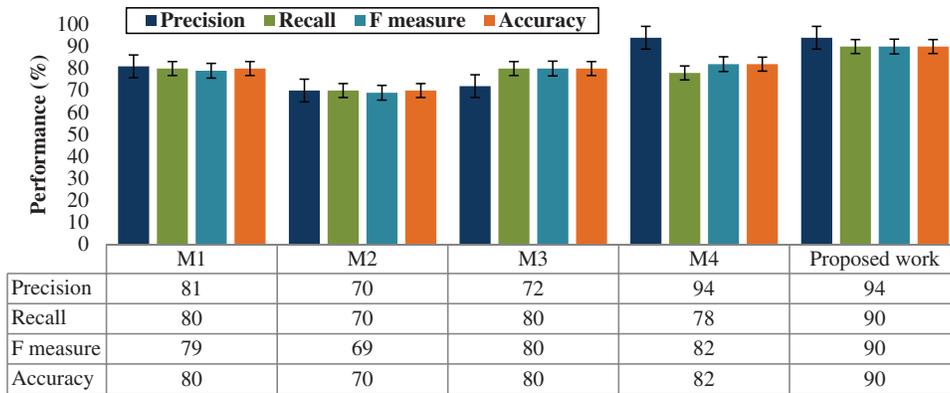


Figure 4: Performance results comparison analysis graph

For various datasets such as Senseval, Senseval-2, SemEval-07 and SemEval-15, the performance of the precision, recall, f-score and accuracy is evaluated and illustrated in Figs 5–8. The comparison shows that the proposed method has proved with the outstanding performance than the previous methods.

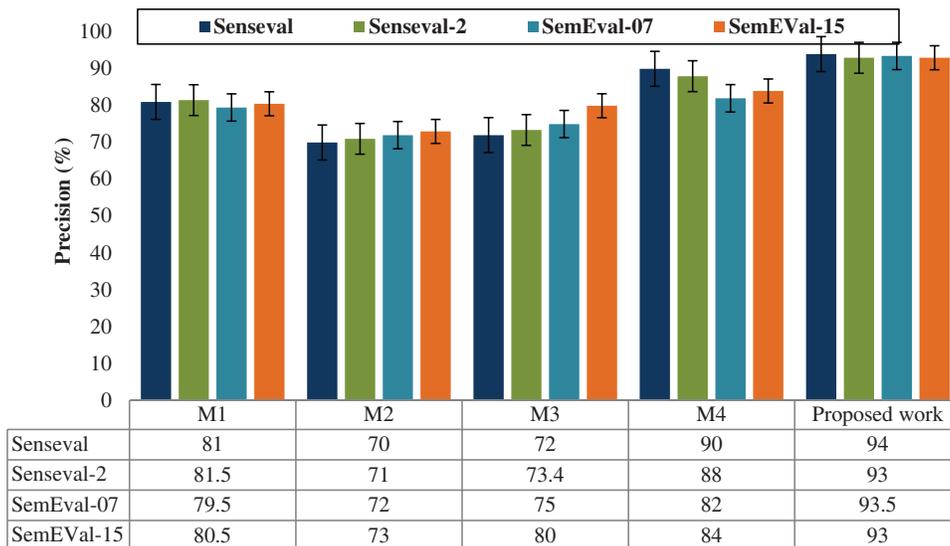


Figure 5: Performance results graph for precision

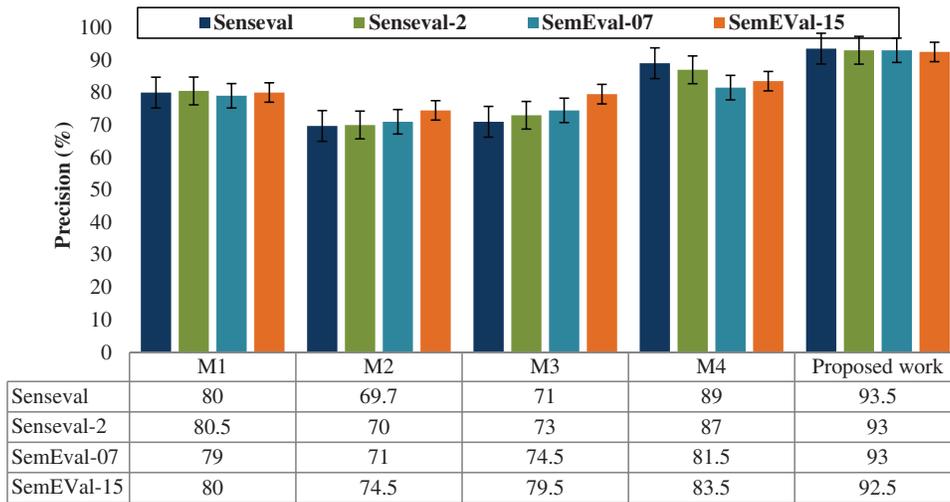


Figure 6: Performance results graph for recall

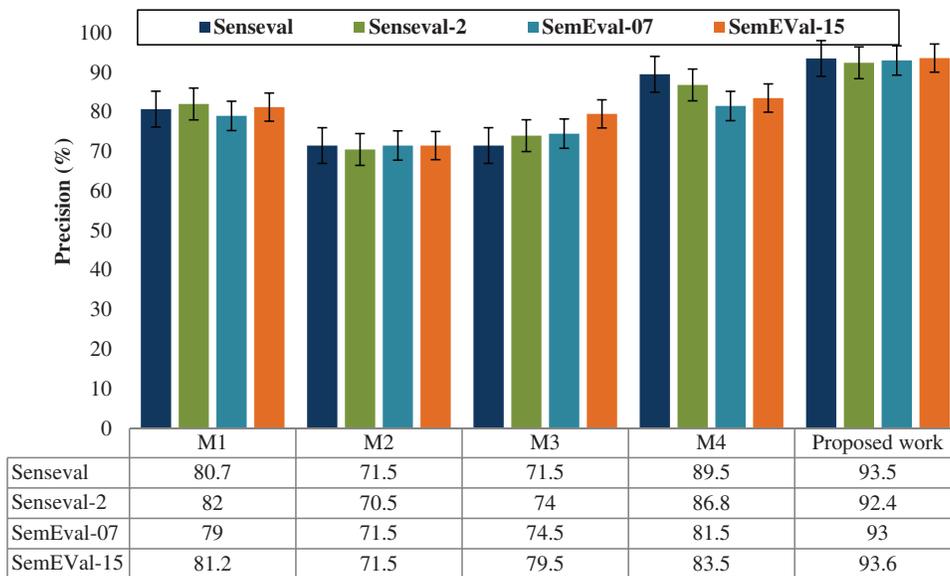


Figure 7: Performance results graph for f-score

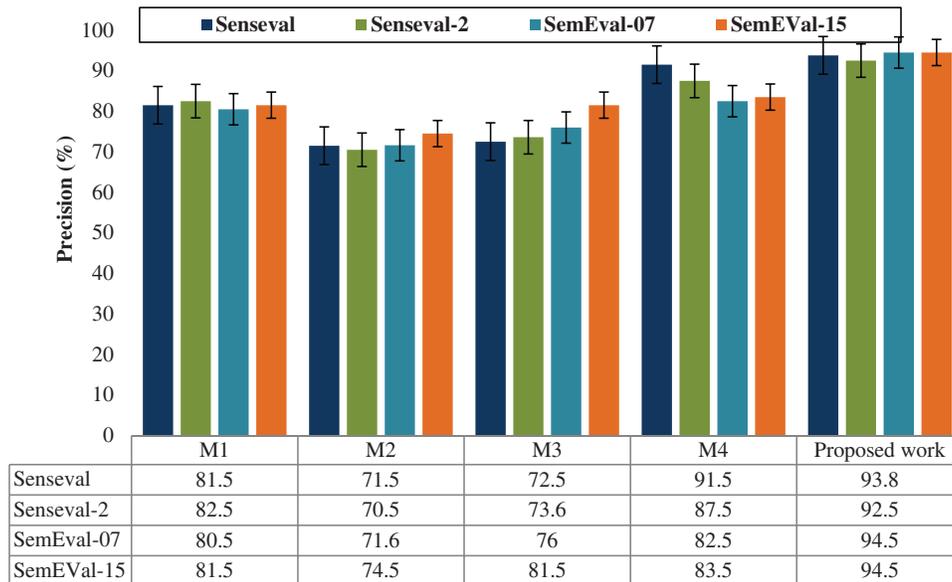


Figure 8: Performance results graph for accuracy

5 Conclusion and Future Scope

The proposed text document classification is performed effectively by optimized feature extraction and the classification approaches. Here the text data is normalized and removed by the noise as unwanted data and unwanted symbol then it is trained using linear kernel function and it is tested. Here the K-SVM is performed effectively on classification and the cuckoo search optimization approach helps to extract the feature. Experimental results show that the better analyzed results of text mining, feature analysis and the classification. By comparing with the proposed and existing approach, the proposed work achieves the better results than recent literatures. In future, the work may extend with big data analysis and various real-time applications.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] H. Liang, X. Sun, Y. Sun and Y. Gao, "Text feature extraction based on deep learning: A review," *EURASIP Journal on Wireless Communications and Networking*, no. 1, pp. 1–12, 2017.
- [2] J. Zheng and L. Zheng, "A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification," *IEEE Access*, vol. 7, pp. 106673–106685, 2019.
- [3] P. Karuna, Ukey and A. S. Alvi, "Text classification using support vector machine," *International Journal of Engineering Research and Technology*, vol. 1, no. 3, pp. 1–4, 2012.
- [4] M. Ikonomakis, S. Kotsiantis and V. Tampakas, "Text classification using machine learning techniques," *WSEAS Transactions on Computers*, vol. 4, no. 8, pp. 966–974, 2005.
- [5] J. Gong, Z. Teng, Q. Teng, H. Zhang, L. Du *et al.*, "Hierarchical graph transformer-based deep learning model for large-scale multi-label text classification," *IEEE Access*, vol. 8, pp. 30885–30896, 2020.

- [6] C. Silva, U. Lotric, B. Ribeiro and A. Dobnikar, "Distributed text classification with an ensemble kernel-based learning approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 3, pp. 287–297, 2010.
- [7] H. Huan, J. Yan, Y. Xie, Y. Chen, P. Li *et al.*, "Feature-enhanced nonequilibrium bidirectional long short-term memory model for Chinese text classification," *IEEE Access*, vol. 8, pp. 199629–199637, 2020.
- [8] V. Kuppili, M. Biswas, D. R. Edla, K. R. Prasad and J. S. Suri, "A mechanics-based similarity measure for text classification in machine learning paradigm," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 2, pp. 180–200, 2018.
- [9] S. Ahmad, M. Z. Asghar, F. M. Alotaibi and S. Khan, "Classification of poetry text into the emotional states using deep learning technique," *IEEE Access*, vol. 8, pp. 73865–73878, 2020.
- [10] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, A. Mehmood *et al.*, "Document-level text classification using single-layer multisize filters convolutional neural network," *IEEE Access*, vol. 8, pp. 42689–42707, 2020.
- [11] H. Calvo, A. P. Rocha-Ramirez, M. A. Moreno-Armendáriz and C. A. Duchanoy, "Toward universal word sense disambiguation using deep neural networks," *IEEE Access*, vol. 7, pp. 60264–60275, 2019.
- [12] Q. P. Nguyen, A. D. Vo, J. C. Shin and C. Y. Ock, "Effect of word sense disambiguation on neural machine translation: A case study in Korean," *IEEE Access*, vol. 6, pp. 38512–38523, 2018.
- [13] Z. Q. Wang, X. Sun, D. X. Zhang and X. Li, "An optimal SVM-based text classification algorithm," in *Int. Conf. on Machine Learning and Cybernetics*, Dalian, China, pp. 1378–1381, 2006.
- [14] Y. G. Lee, "A study on optimization of support vector machine classifier for word sense disambiguation," *Journal of Information Management*, vol. 42, no. 2, pp. 193–210, 2011.
- [15] M. Awad and R. Khanna, "Support vector machines for classification," in *Efficient Learning Machines*, pp. 39–66, Apress, Berkeley, CA, 2015.