

© 2016, TextRoad Publication

News Headlines Classification Using Probabilistic Approach

Mazhar Iqbal Rana¹, Dr. Shehzad Khalid², Fizza Abid³, Armughan Ali⁴, Mehr Yahya Durrani⁵ and Farhan Aadil⁶

^{1,2,3}Department of Computer and Software Engineering, Bahria University, Islamabad, Pakistan ^{4,5,6}Department of Computer Sciences, Comsats Institute of Information Technology, Attock, Pakistan

> Received: January7, 2016 Accepted: March 22, 2016

ABSTRACT

This paper is aimed at news classification on basis of their headlines. Researchers have worked a lot for carrying out news classification at full text level but work in the domain of news headlines classification exists in very limited ratio, Therefore, after analyzing variety of existing news classification methodologies, a probabilistic framework is presented in this paper for classifying news headlines. News headlines classification process is divided into three modules, headlines pre-processing module, probability learning module, and news headlines classification module. Based on availability of variety of headlines, probabilistic framework is designed, which classifies each news headline to its pre-defined category by calculating its maximum probability in that category. Work has been performed using bag of words approach where each headline is split into words and each word is given a certain probability. Furthermore, it is shown that proposed system gives better accuracy results as compared to existing headline classification systems. **KEYWORDS**: News classification, News headlines classification; Probabilistic model; Frequency statistics; Probability learning

INTRODUCTION

1

In the recent times, "Text Mining" has been an area under spotlight in the field of data mining, with the extensive engagement of numerous of researchers carrying out research in this specific field. Text Mining is the process of combing through extensive digitized text to find out useful and high-quality information. The importance of this field is evident through its utilization in different disciplines; for instance it has become a major part of machine learning discipline such that it utilizes knowledge discovery methods through which logical rules can be derived to classify text within desired category. This can be further elaborated with an example of the attainment of already labelled incoming news from various categories like sports, politics, entertainment etc. In addition, an effective classification model is prepared to classify an unseen text into its correct classification category, by interpreting the content of the input text. Two factors are essential for text classification namely, accuracy and efficiency. Where the former signifies the correctness of the classification of the text with its desired class (for example research related to medical and finance fields require high level of accuracy since a minor error may result in disastrous outcomes), the latter refers to the classification model's performance. One of the topics that have been extensively scrutinized in research is the news classification, where research on complete news has been an area of extensive interest as compared to precise news texts. According to some researchers the prime reason for this apathy is due notion that short news texts/messages though utilize fewer words to convey the message but do not facilitate with important data that could be utilized for further calculation and result in loss of accuracy and efficiency [8]. However, a number of other researchers have successfully conducted research on "news headline classification", obtaining adequate results. In the past, research experts have carried out numerous research on news classification, classification of emotions based on news headlines and so on; where most of them have proven that news headlines consist of less substantial data but on the other hand they also have claimed that with regard to the performance parameter, considering factors like: consumption of energy, time, accuracy, accurate classification, the precise texts/ news headlines give better results. Whereas, few of the short text classifications, which have taken place in past include: emotions classification on basis of news headlines, classification of short texts [8], financial news classification [5], automatic news headlines classification [22], news headlines classification using N-gram model [18], classification of news headlines for providing user centred e-newspaper

* Corresponding Author: Mazhar Iqbal Rana, Department of Computer and Software Engineering, Bahria University, Islamabad, Pakistan. Email: imazhar74@hotmail.com,

[3], emotions extraction from news headlines [15], and short news headlines classification of twitter social networks [4]. Furthermore, "complete news classification" has been an entrenched research area, in comparison to "news headlines classification", and this paper is focused on "news headlines classification" respectively. This paper introduces the probabilistic approach for news headlines classification, where each news headline will be classified into its pre-defined class.

Through the probabilistic approach, the classification will be done such that the probabilities of the words of news headlines are calculated, words in headlines are first presented as container of words and through statistical calculations, effective news headlines classification is performed. This is done by computing probability of words of all classes and the words with maximum probability are considered useful for a certain class. Furthermore, words with high frequency in the classes are derived out; and the evaluation is done using the standard news headline datasets. Following this brief gist of the paper, the remainder of the paper describes the vital topics of our research accordingly; section II gives a brief review of existing news classification techniques as well as feature selection techniques, section III details out the proposed probabilistic approach, section IV presents the effectiveness of proposed probabilistic news classifier and the experimental evaluations; and the last section, section V, presents the conclusion of the paper.

2. **Background and Realted Work.** The basic purpose of performing text mining is to extract high-quality information from text. This is achieved by training the classifier and performing the required operations for removing the noise. To carry out the entire process, mandatory classification process is to be followed, which comprises of three major steps i.e. text pre-processing, feature selection, and classifying news headlines. There is dearth of literature regarding news headlines classification, however, this is what we are majorly focussing on in this paper. Basic pre-processing of both full text classification as well as short text classification remains the same, but there are few changes in feature selection phase. The following discussion comprises of how researchers have contributed in each area of news headlines classification.

The basic objective of gathering information from scattered data is termed as data mining. The most important feature which has to be performed before feature selection is to segment the entire news headline into words. This is known as tokenization [16]. Many researchers have made contributions in various classification related activities like stop words removal where all those words, which are frequent in all classes, are eradicated. Many research scientists have used various modes of stop words removal. The most common of the stop words removal methods include Term Frequency Inverse Document Frequency [19, 24], Zipf's Law [23], Conceptual Based Removal and existing SMART list based removal [12], Automatic List Generation [19], and much more. Moreover, contributions are not only limited to this but several stemmers have been proposed by researchers in the past, and the most widely used amongst them is porter stemmer. Stemming is defined as reduction of word to its root. S-Stemmer [6] was proposed to conflate singular forms as well as plurals. Its major drawback is that it only handles and works for words greater than length three. In 1968, Lovins introduced a stemmer known as Lovins stemmer [6, 23]. The stemmer comprises of 29 rules and one can apply any of them to generate word stem. This Lovins stemming approach lost its importance when more comfortable and efficient stemmer i.e. Porter Stemmer was introduced in 1980. It is considered to be the best one up till now just because of its accurate and satisfied results. We, in this paper, will also be using the same Porter stemmer for carrying out stemming operations, which will later be discussed in section III. In 1990, Mr Paice introduced Paice/Husk stemmer for carrying out better news classification tasks. In Paice/Husk stemmer, rules are divided into five main steps. Three rules are mandatory to apply and remaining two rules are completely optional. For instance, one of the rules of Paice/Husk states that if word terminates with '>', this means that word still needs stemming and must be passed to next cycle but if word termination contains '.', it means that final stem of the word has been obtained [6,23].

News headlines or any type of short text classification gives better and improved results if and only if suitable and the most relevant features are selected. Filtering out the useful data from huge amount of data requires cautious scrutinization for which appropriate features are selected accordingly. This feature selection is performed in the case of news headlines, which is represented as a bag of words approach. News headlines and short text classification saves time and yields more precise and accurate results in comparison to full text classification. Many feature selection methods have been adopted in the past by researchers. The most common of these include Boolean weighting [16], Document Frequency [16,17], TF-IDF [14,16], Information Gain [9,16], Mutual Information [10], and lot more.

Anita used Boolean weighting [16] methodology for selecting features, which searches for those words that are either present or absent in given news headlines. In addition, literatures shows that Anita and Andreas

Hotho emphasized upon document frequency threshold method [10,16]. This outputs all those news documents, which consist of specific word at least once. This technique eradicates all those words completely, , which may have document frequencies less than a certain defined threshold value, hence, classifying news in a much better way [19]. Many other techniques such as Term-Frequency Inverse Document-Frequency [24] method, Information gain etc were used by researchers to classify full news stories as well as news headlines. TF-IDF [16,17] can also be used for stop words removal but it is mostly used for feature selection. Information Gain [1,9] predicts whether a word appears in a certain news document or not. Presence or absence of word allows researchers to select useful and the most informative features easily, hence, making news headlines classification more robust and reliable.

Equation for TF-IDF is as follows.

$$TF - IDF_{w} = TF_{w} * IDF_{w}$$
(1)

The very last phase of classification process is to create feature set and on basis of this set we classify news headlines along with accuracy computation. After feature selection, classification phase is important where unseen news is classified into their respective classes. Major aim of classification is to place documents in their respective classes. Classification in the past has been performed for various purposes i.e. Short Text Classification by Heb [8], News Headlines Classification by Pope [3], Xin, Deshmukh [22], Chen p [11], and Khushbu, and all others [2,4,13, 17,18]. Moreover, social blog messages classification of twitter by Caldera et al [4], and Financial news classification by Drury [5] on basis of news headlines was also performed. News classification was not only limited to English but numerous other languages like. Urdu by Ali & Ijaz [2] was also classified. Mark Pope [22] performed automated classification of online news headlines on basis of text by classifying news in accordance to its respective categories using KNN [13] and results were compared with full text classifications. Multi-value headlines classification [8] as well as full text news classification was performed by Andreas using Naive Bayes and results were compared. Poisson Model was proposed as a result of incorrect parameter assessment of Naïve Bayes and improved accuracy measure was reported [25]. Poisson Model was declared better than other classifiers by S. Ting et. al [25]. Human reader emotions [6,11] were classified and identified using SVM, news from the most visited social networking website i.e. Twitter were classified using SVM [4] in which Neural networks were used in combination with Naïve Bayes and this provided a new approach named as Knowledge Based Neural Networks [7] for handling noise data as well as outliers for classification of full text news. In medical science, Podgorelec, Kokol, Stiglic, & Rozman [21], decisions are to be taken rapidly and minor delay may lead to serious consequences. Performing conceptual analysis [20] and decision making proves appropriate for such situations and system may work better than before. For manipulating such situations, decision trees have helped researchers a lot where medical experts can make better examinations on basis of results they obtain and rules can be deducted on their basis

3. **Proposed News Headlines Classification System.** In this section, we present our proposed approach of classifying news based on their headlines. Our system comprises of three modules. In module 1, news headlines pre-processing is performed on training dataset. Pre-processed data is then passed to module 2 for generating trained model. In module 3, news headlines are classified on the basis of cumulative probability of each word in all classes. Accuracy results are presented at the end.

The proposed news classification system is outlined in Figure 1; furthermore, we have theoretically as well as experimentally explained our system in subsequent sections. Let D = (X, Y) be the news headlines dataset where $X = \{x1, x2, ..., xn\}$ is a dataset containing n headlines and $Y = \{y1, y2, ..., yn\}$ is a set containing the label information of corresponding headline in set X. Let $X^{(j)}$ represent the sample belonging to class j. $C = \{c1, c2, ..., cn\}$ is a set containing the classes of news headlines. (Jia et al., 2009)



Fig 1:. Proposed News Headlines Classification System

- *A) News Headline Preprocessing.* The procedure to pre-process the news headlines dataset comprises of the f ollowing steps.
- 1) A sample news headline x is selected from training data X.
- 2) After selecting a sample x, we tokenize and extract words on basis of delimiters. These delimiters include whitespace, comma, semi-colon etc.
- 3) After having tokenized sample x, we further sanitize it to remove non-letter characters, including punctuations, quotations, numbers, special characters etc.
- 4) Following the above, we remove the most occurring words from the sanitized word list of sample x, as stop words by comparing sanitized word tokens with generic list of stop words.

$$X_{fil} = \begin{cases} X_{fil} \cup x_i, & if(ch_{in} > 2) \lor (x_i \in S_L) \\ X_{fil} & Otherwise \end{cases}$$
(2)

where ch_{ln} is the defined character length for each word, X_{fil} is a filtered list with no stop words, and x_i is the number of news headlines samples in X.

- 5) After getting stop words free words from sample x, stem the words using existing Porter Algorithm [6,23].
- 6) After stemming, train data X is pre-processed and is updated as \overline{X} .
- 7) Repeat steps 1-6 for all the samples x in X.
- B) Generating a Probabilistic Model of News Categories In this section, methodology for generating a probabilistic model of news categories is presented. Model is first learned and train data is then classified. Probability matrix of each news headline across all classes is generated and then, stop words are removed on basis of frequency.
- Learning Initial Probabilistic Model from Headlines Learning probabilistic model from headlines consists of following steps:
- 1) News headlines sample x is taken from X and later on, unique words from \overline{X} are identified. Each sample is checked and X_{unique} matrix is generated.
- 2) After identifying unique words from \vec{X} , frequency matrix $Freq_{i,j}$ for each word *i* in class *j* is generated. This is achieved by initializing the freq matrix by zeros. Select a headline from training data and tokenize words from it. Let *i* be the index of the selected word in the unique words list and *j* be the index of the class to which a selected headline belongs, the freq matrix for each word w_i in class c_j is generated as:

$$Freq_{i,j} = Freq_{i,j} + 1 \quad \forall w_i \in X_{unique}$$
(3)

where i = 1, 2, ... #words, j = 1, 2, ... #classes, and X_{unique} is the unique words list.

Using frequency matrix generated in step 2, compute the probabilities of sample headlines x in all c_j classes.

$$P(c_j \mid w_i) = Freq(c_j \mid w_i) / \sum_{j=1}^{N} Freq(c_j \mid w_i) \ \forall i, j$$

$$(4)$$

where $P(c_j | w_i)$ is the probability of occurrence of word w_i in class c_j , $Freq(c_j | w_i)$ is frequency of

word w_i in class c_j , and $\sum_{j=1}^{N} Freq$ $(c_j | w_i)$ is the sum of all w_i word frequencies in all c_j classes.

- 4) Initialize the probability matrix $prob=\{1,2,...,c_n\}$ where c_n is the number of class from 1,..., n.
- 5) Take square root of the cumulative probabilities word w_i at extracted index. Update the probabilities across c_j classes as:

$$prab_{c_j} = \sqrt{prab_{c_j} + p(c_j|w_i)} \quad \forall j$$
 (5)

where $prob_{c_j}$ is is the probability of class c_j and $prob_{c_j}$ is probability of each word w_i at extracted index of word w_i .

- 6) Repeat the steps 4 and 5 for all w_i words in headlines x.
- 7) Identify the maximum probability for *class j*.

$$class_{j}^{x^{(0)}} = \arg \max_{\substack{\forall c \mid a \neq j}} prob_{c_{j}} \forall i, j \tag{6}$$

where $prob_{c_j}$ is the probability of *class*_j given $x^{(t)}$ and $arg \max_{vclass_j}$ returns index of *class*_j having maximum probability.

- Find Domain Specific Less Informative Words For removing less inforative words, an approach named as word frequency inverse class frequency (*W_f*-*Ic_f*) is presented where stop words are removed on basis of fixed threshold. Process of for removing stop words using *W_f*-*Ic_f* is presented below.
- 1) Let W_f be word wi frequency in class c_j and Ic_f be the inverse class frequency of word w_i , word frequency inverse class frequency W_f - Ic_f is computed as:

$$W_{f} - Ic_{f_{(i)}}^{(i)} = Freq_{i,j} \times \log_{10}(\#words \div Freq_{i,j})$$
(7)

where $Freq_{i,j}$ is frequency of word w_i in class c_j and #words are total number of words in headlines.

2) Generate list of word frequencies across all classes i.e. $\sum_{i=1}^{N} Freq_{i,j}$

- *3)* Sort sum of word frequencies in descending order i.e. words with highest frequencies must be at the top of list.
- Sort W_f-Ic_f results against each word in ascending order i.e. the word most frequent will have lowest W_f-Ic_f value.
- 5) Filter the word from word list if it satisfies the following condition.

$$X_{f} = \left\{ W_{f} - Ic_{f_{j}}^{i} < \alpha \right\}$$

$$\tag{8}$$

where X_f is a filtered stop words list and \propto is the threshold. The value of \propto is set as 0.000115 through empirical evaluation.

- 6) Merge this generated stop words list with existing generic stop words list
- 7) Stop words list S_L is now updated as:

$$\overline{S}_{L} = \left\{ S_{L} \cup X_{f} \right\}$$
(9)

where S_L is a generic stop words list, X_f frequency based generated stop words list.

- *C) News Headlines Classification* The methodology for news headlines classification comprises of following steps.
- 1) Select a sample headline *tx* from sample test data *TX*.
- 2) Pre-process the selected sample headline *tx* where stop words list $\overline{S_L}$ is updated earlier in section 3.3.2.3 (step 5).
- 3) Extract words w_i from pre-processed headline.
- Synonyms sets are made where synonyms are taken from WorldNet. These synonyms sets comprise of words and their possible synonyms. Update word list w by including synonyms.
- 5) Identify the index of extracted word \widetilde{W}_{t} in X_{unique}
- 6) Compute sum of probabilities P at extracted index of word w_i . Update the probabilities across c_j classes as:

$$prob_{c_j} = \sqrt{prob_{c_j} + p(c_j | \overline{w_i})} \quad \forall j$$
(10)

where $prob_{c_j}$ is the probability of class c_j and $p(c_j | \overline{w_i})$ is the probability of each word $\overline{w_i}$ at extracted index.

- 7) Repeat the steps 5 and 6 for all w_i words in headlines tx.
- 8) Identify the maximum probability for *class j*.

$$class_{j}^{tx^{(l)}} = \arg \max_{\forall class_{j}} prob_{c_{j}} \forall j$$
(11)

where $prab_{c_j}$ is the probability of $class_j$ given $tx^{(t)}$ and $arg \max_{vclass_j}$ returns index of class $class_j$ having maximum probability.

4. Experimental Studeis. In this section, discussion will revolve around experimental datasets along with 4 experiments.

A. Experimantal Datasets To evaluate the performance of proposed novel probabilistic approach, two different datasets have been used including DS1 and DS2. A Brief overview of these datasets is given below in TABLE I.

Data set	Dataset Description	# of instances	# of Classes
DS1	News headlines data set where headlines are 10,000(Train Data) gathered from RSS feeds of 2,200(Test Data) various popular news websites.		6
DS2	Data set consisting of news headlines taken from Reuters data set. Reuters have been adopted in carrying various text categorization tasks.	8,553(Train Data) 2,928(Test Data)	10

TABLE I: A BRIEF OVERVIEW OF EXPERIMENTAL DATASETS

B. Experiment 1: Evaluation of Proposed Approach using Generic Stop Words List The aim of this experiment is to illustrate the classification accuracy and effectiveness of proposed approach by using existing generic stop words list. The experiment has been carried out on DS1 dataset, which has been used most of the time for news headlines classification. Experiment has been carried out on total of six classes comprising of news headlines collected from various websites.

The generic stop words list is publicly available and consists of about 545 stop words. Experiment is conducted on DS1 dataset and experimental results of proposed approach are presented. These results are obtained against existing stop words SMART list. Proposed probabilistic classifier is evaluated, and it is observed from Table II that the proposed probabilistic classifier yields the overall accuracy of 83%. Moreover, it can also be observed that by probabilistic learning, it is determined that how much each word is effective and probable for any class. Also, it is observed that this SMART list consists of many relevant words, which are considered stop words. Therefore, this generic list is not always reliable for carrying out news headline classification tasks effectively.

C. Experiment 2: Evaluation of Proposed Approach using Updated Stop Words List The aim of performing this experiment is to illustrate the classification accuracy of the proposed probabilistic approach by using updated stop words list. The experiment has been carried out on the same DS1 dataset, which has been used for categorizing news headlines. Here, existing stop words list is updated by adding newly identified frequency based stop words from the corpus. The word frequencies computed by equation (3) are used for identifying domain-specific fewer informative words. The stop words identification is performed by computing inverse of word frequencies first, which is calculated by the total number of words in the corpus divided by their distinctive class frequencies as mentioned in equation (7). Logarithmic value of inverse word frequencies is taken, which is later multiplied by individual word frequencies across all classes have been used including DS1 and DS2. The word occurring the most in any class gives very less frequency weight as compared to those words occurring very rare. Afterwards, stop words are removed on the basis of fixed frequency threshold as mentioned in equation (8) and identified stop words list is next generated

and merged with existing generic SMART stop words list using equation (9). General stopwords list consists of total 545 stop words but after adding 200 frequencies based stop words, the list is updated to 745 words. Proposed probabilistic classifier is evaluated on this updated stop words list and it is observed that proposed probabilistic classifier yields the overall accuracy of 84.63%. It can also be observed that by adding frequency based stop words in existing generic stop words list, slight improvement in classification accuracy is achieved as shown in Table II below.

	Proposed Classifier/ Classification Method	Headlines Classification using Generic Stop Words List	Headlines Classification using Updated Stop Words List	Accur acy (%)
	1	✓		83
	2		\checkmark	84.63
ſ	3	✓	\checkmark	86.66

TABLE II: Contribution of various components in proposed framework using DS1

D. Experiment 3: Comparing Proposed Approach with Competitor Using Dataset DS2 Here we present the comparison of proposed approach with competitors using DS2 dataset. The classification accuracy results are compared with those of DS2 where competitor has used Naïve Bayes classifier for news headlines classification. We have compared our Naïve Bayes results using dataset DS2. All of these experimental results are reported in Table III. By experimental evaluation, we get the better accuracy of Naïve Bayes classifier on our experimental setup on DS2 dataset i.e. 75% whereas competitor's accuracy results are 53% using Naïve Bayes classifier. By experimental evaluation, it is determined that our results have outperformed competitor's results on Naïve Bayes classifier and have given more improved accuracy.

 TABLE III: Comparing proposed approach with competitor using data set DS2

Results/Classifier	Naive Bayes Accuracy (%)	
Competitor Reuslts with Naive Bayes for News Headlines Classification	53	
Competitor Reuslts with Naive Bayes for News Headlines Classification	79.37	

E. Experiment 4: Comparing Proposed Approach with Competitor Using Dataset DS1 This experimentcompares our accuracy results with competitor using DS1 dataset and analyses Naïve Bayes results. Experiment is carried out by using Naïve Bayes parameters and calculating probabilities of words within all classes. Using Naïve Bayes, it is observed that more frequent classes are possibly considered correct classified classes as compared to infrequent classes. The sum of log and posterior probabilities signifies that how much probable is each word for being in any class. The words with maximum probabilities are considered useful for any class. Logarithms of probabilities are taken, and classification accuracies are reported as shown in Table IV.

TABLE IV: Comparing	proposed approach	with competitor	using data set DS	1
	p p			

Dataset	Naive Bayes Accuracy (%)	Proposed Classifier Accuracy (%)
DS1	84.09	86.66

By empirical evaluation, it is observed that Naïve Bayes results with our approach give improved accuracy i.e. 86.66% as compared to the competitor on similar datasets. Furthermore, it is also determined that our Naïve Bayes results outperform competitor's results keeping the experimental setup equivalent.

Conclusion. In this paper, a novel probabilistic approach is discussed for classifying news, based on their headlines. Each of the modules is separately discussed. All the three modules i.e. pre-processing, model learning, and news headlines classification are applied in this approach. In addition, stop words filtering is done using frequency based stop words removal approach, where word frequency inverse class frequency

method is applied. To show the performance and effectiveness of our proposed method on each individual classifier, we conducted four experiments on our data sets for measuring classification accuracy. Competitor analysis is also performed using both experimental datasets. Experimental studies show that our proposed approach gives superior results as compared to the competitor. It also gives best retrieval accuracies for generic as well as updated stop words list. Proposed system classifies news headlines in a much better fashion as compared to previous classifications that have already been conducted for full documents news.

REFERENCES

- Aggarwal, C.C., Zhai, C.: A survey of text clustering algorithms. In: Mining Text Data, pp. 77–128. Springer
- [2] Ali, A. R., & Ijaz, M. (2009). Urdu text classification. Paper presented at the Proceedings of the 7th International Conference on Frontiers of Information Technology.
- [3] Deshmukh, R., & Kirange, M. D. Classifying News Headlines for Providing User Centered E-Newspaper Using SVM
- [4] Dilrukshi, I., De Zoysa, K., & Caldera, A. (2013). Twitter news classification using SVM. Paper presented at the Computer Science & Education (ICCSE), 2013 8th International Conference on IEEE (2013) 287-291
- [5] Drury, B., Torgo, L., & Almeida, J. (2011). Classifying news stories to estimate the direction of a stock market index. Paper presented at the Information Systems and Technologies (CISTI), 2011 6th Iberian Conference
- [6] Frakes, W. (1992). Introduction to information storage and retrieval systems. Space, 14, 10. Frakes, W. B. (1992). Stemming Algorithms.
- [7] Goyal, R. D. (2007). Knowledge based neural network for text classification. Paper presented at the Granular Computing, 2007. GRC 2007. IEEE International Conference
- [8] Heß, A., Dopichaj, P., & Maaß, C. (2008). Multi-Value Classification of Very Short Texts KI 2008: Advances in Artificial Intelligence (pp. 70-77): Springer.
- [9] Hotho, A., Nürnberger, A., & Paaß, G. (2005). A Brief Survey of Text Mining. Paper presented at the Ldv Forum.
- [10] Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text classification using machine learning techniques. WSEAS Transactions on Computers, 4(8), 966-974.
- [11] Jia, Y., Chen, Z., & Yu, S. (2009). Reader emotion classification of news headlines. Paper presented at the Natural Language Processing and Knowledge Engineering, 2009. NLP-KE 2009. International Conference.
- [12] Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features: Springer.
- [13] Khamar, K. (2013). Short Text Classification Using kNN Based on Distance Function. IJARCCE International Journal of Advanced Research in Computer and Communication Engineering. Government Engineering College, Modasa.(ISSN Print: 2319-5940 ISSN Online: 2278-1021).
- [14] Kim, S.-B., Han, K.-S., Rim, H.-C., & Myaeng, S. H. (2006). Some effective techniques for naïve bayes text classification. Knowledge and Data Engineering, IEEE Transactions on, 18(11), 1457-1466.
- [15] Kirange, D. (2013). Emotion Classification of News Headlines Using SVM. Asian Journal of Computer Science & Information Technology, 2(5).
- [16] Krishnakumar, A. (2006). TEXT CATEGORIZATION Building a kNN classifier for the Reuters-21578 collection: Citeseer.
- [17] Liu, M., & Yang, J. (2012). An improvement of TFIDF weighting in text categorization. International Proceedings of Computer Science and Information Technology, 44-47.

- [18] Liu, X., Rujia, G., & Liufu, S. (2012). Internet news headlines classification method based on the n-gram language model. Paper presented at the Computer Science and Information Processing (CSIP), 2012 International Conference.
- [19] Lo, R. T.-W., He, B., & Ounis, I. (2005). Automatically building a stopword list for an information retrieval system. Paper presented at the Journal on Digital Information Management: Special Issue on the 5th Dutch-Belgian Information Retrieval Workshop (DIR).
- [20] Mahmood, A. M., Satuluri, N., & Kuppa, M. R. (2011). An Overview of Recent and Traditional Decision Tree Classifiers in Machine Learning. International Journal of Research & Reviews in Ad hoc Networks, 1(1).
- [21] Podgorelec, V., Kokol, P., Stiglic, B., & Rozman, I. (2002). Decision trees: an overview and their use in medicine. Journal of medical systems, 26(5), 445-463.
- [22] Pope, M. W. (2007). Automatic classification of online news headlines. A Master's paper, University of North Carolina at Chapel Hill.
- [23] Ramasubramanian, C., & Ramya, R. (2013). Effective Pre-Processing Activities in Text Mining using Improved Porter's Stemming Algorithm. International Journal of Advanced Research in Computer and Communication Engineering, 2(12).
- [24] Ramos, J. (2003). Using tf-idf to determine word relevance in document queries. Paper presented at the Proceedings of the first instructional conference on machine learning.
- [25] Ting, S., Ip, W., & Tsang, A. H. (2011). Is Naive Bayes a good classifier for document classification? International Journal of Software Engineering and Its Applications, 5(3), 37-46.