# Support Vector Machine and Back Propagation Neural Network Approach for Text Classification*

Yaqeen Saad

Department of Computer Science,
University of Anbar
Anbar, Iraq
Yaqeen.cs91@gmail.com

Khaled Shaker

Department of Computer Science,
University of Anbar
Anbar, Iraq
khalidalhity@gmail.com

*Abstract*— **Text classification is the process of inserting text into one or additional categories. Text categorization has many of significant application, Mostly in the field of organization, and for browsing within great groups of document. It is sometimes completed by means of "machine learning.". Since the system is built based on a wide range of document features."Feature selection." is an important approach within this process, since there are typically several thousand possible features terms. Within text categorization, The target goal of features selection is to improve the efficiency of procedures and reliability of classification by deleting features that have no relevance and non-essential terms. While keeping terms which hold enough data that facilitate with the classification task. The target goal of this work is to increase the efficient text categorization models. Within the "text mining" algorithms, a document is appearing as "vector" whose dimension is that the range of special keywords in it, which can be very large. Classic document categorization may be computationally costly. Therefore, feature extraction through the singular valued decomposition is employed for decrease the dimensionality of the documents, we are applying classification algorithms based on "Back propagation" and "Support Vector Machine." methodology. before the classification we applied "Principle Component Analysis." technique in order to improve the result accuracy . We then compared the performance of these two algorithms via computing standard precision and recall for the documentscollection.**

*Keywords- Text Classification, classification algorithm, SVM, Back propagation.*

## I.  INTRODUCTION

In recent years ago, content based document arrangement tasks (collectively indicate to as information retrieval IR) have gained an obvious standing inside the information systems field, due to the increase availability of text records in digital convenience to arrive them in flexible ways [1]. For the quick growth of the net and on-line knowledge, automatic text classification has attracted many researchers and companies. Numbers of models for (ML) have been used to task of document categorization, for example, decision trees, k-nearest-neighbor and Naive-Bayes [2]. Document categorization is the procedure of arranging (attaching) text into one class or more from predefined categories or classes, in other words to assign category labels to documents. Spontaneous text categorization can suppose a demanding section in a wide variety of more-flexible, dynamic, land personalized information management task as well: real-time sorting of email or files into folder hierarchies. The variance within outcome of this compartmentalization grow from an attribute group elected for foundation the unity of particular, texts with a given class the summon of text assortment distinguish that the separation of text documents into groups of like documents minimize the overhead desired for speedy recall of such documents and supply minimal domains in which the users may scout about similar documents [3]. Text classification is "An active research area of text mining, where the documents are classified into predefined classes," (wajeed;2004). There are numbers of approaches have been applied to solve the problem of text classification. In the few years that have elapsed "Support Vector Machines classification" Become a priority. SVM's deep theoretical partition, precision, computational scalability. Also Back Propagation offered by Neural Network is effective algorithms for the task of text classification. It can be dealing with large dimensional data space [3] [6] [19].

## II.  BASIC DEFINITION FOR TEXT CATEGORIZATION

Text Categorization is an active research area of text mining where the text is prepared with supervised, unsupervised or semi-supervised, Knowledge [4]. Text assortment indicates to fix the trouble to classify documents based on their content into a specific set of predefined labels. The major aim of Text Categorization is to assign the class to the novel document. TC plays very significant role in vast set of areas such as information recovery, web page classification, and many more [5].

## III.  KIND OF DOCUMENT SORTING

### A.  *Flat and Hierarchical classification*

Mostly, text arrangement may be split into two kinds, flat and hierarchal classification. In the flat classification, the sub-headings are not seen as shown in Figure 1.
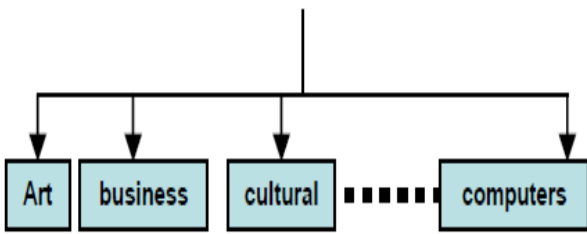


Figure 1: Flat classification

When it becomes a large number of documents in category, The search through this kind of classification is complicated and the compartmentalization reliability of the text classification system that applied this data will to be affected This problem led to the use of hierarchical classification as shown in Figure2 [6] [7].
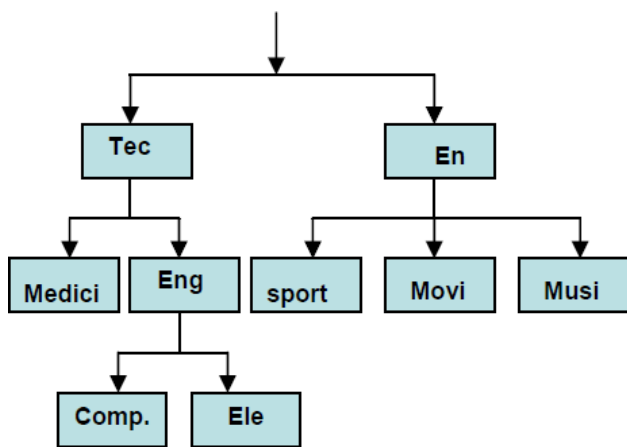


Figure 2: Hierarchical classification.

### B.  *Supervised vs Unsupervised Methods*

At supervised method, the information as offered to ML is totally classified. That means: whole examples are offered with compartmentalization that the machine is intended to reproduce .With the unsupervised method the systemstaretnottsupplied anyttraining models on whole and manner of clustering. This is incision of datum byword for numerous sets. Information driven is The score of clustering algorithms, hence further 'natural' and best compatible to the tacit frame of the data. [12][13].

### C.  *One-label vs Multi-label Text Assortment*

In text classification the Single-label refer to the procedure of allocating document ( dj ) from dataset to just one from pre-defined group , in this case there are number of categories but only one can be assigned to the document, whereas in multi-label job, the document (dj) may be able to allocated to zero, one or more than one category [8].

## IV.  THE GENERAL STEP OF TC

The steps of document assortment are discussing as following points.
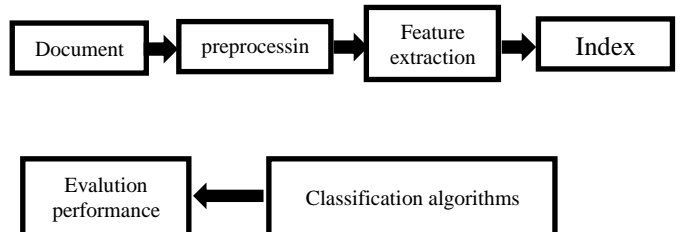


Figure 3: Document Classification Process

### A.  *Preprocessing*

At the first phase of preprocessing which is utilized to return texts for obvious term formation. The records of text intended to the following stage in categorization of the text are performed by great amount of attributes. Commonly the steps taken are:

- Tokenization: Tokenization takes the text (i.e. a string) and find the sentences and `tokens'. This may seem like a `no-brainer' (just divide the string at every space), but this is not the case. Abbreviations like `etc.', `i.e.' and `e.g.' are typically written with punctuations, so punctuations do not always end sentences. Also, hyphens, digits, cases and more need to be handled. Combinations of rules and predefined

lists can be used to decide what to do with ambiguities [10].

- Stop words Removal: Stop words, additionally called function words or popular words, are terms that are so popular in a text collection that they do not insert any information to the categorization process. terms like " prepositions" and paper that show extremely and do not assist in recognition among the categories are named : `stop words.', These terms are as a rule extracted using a stop words list. [11].

- Stemming word: Whilst putting the stemming algorithm that procedure is changed sundry term form for comparable canonical form. This step resbonsible for restoration tokens to their origin design for example : writing to write [11].
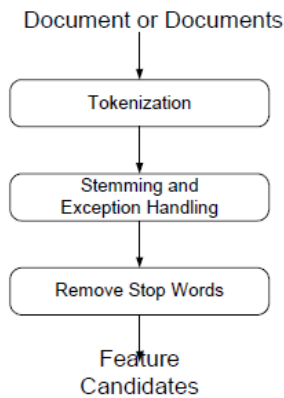


Figure 4:  Text preprocessing

### B.  Feature Extraction

Feature extraction additionally referred to as "Feature Transformation" there are a lot of methods for FE. These procedures are working on reducing input data size. Principle Component Analysis is a fully famed method with feature transformation. PCA This algorithm is designed to define the matrix peculiar with a view to lowering  first attribute size to lower dimensional feature space because reduce the complication of ranking job without any trade-off within precision [26].

- Latent Semantic Indexing Technique: Latent Semantic Indexing (LSI) is  mechanism which comes the initial  input with great dimensional of document into space with ''latent linguistics dimensions. LSA required Singular Value Decomposition of this matrix to put up semantic vector space that can be wont to used-to perform conceptual term-document associations.
- Singular value decomposition: SVD is well-developed methodology for the task of electing fundamental attributes from great dataset and for minimizing the dimensions of the data. At first all terms within the text documents can be built. Once a

term by document matrix is made, SVD is employed to decompose the term by document matrix to construct a linguistics vector space which can be used totrepresenttconceptualttermdocumenttassociation [25].

- Principle component Analysis: PCA it is a successful and important technique as it has the ability to reduce the size of data via converted the main attribute space into smaller spaces. In another meaning, the target aims of PCA are to derive new variables that are combinations of the original variables and are interrelated [27]. Steps of PCA "in order to reduce dimensionality from d to m":

1. Post the information

2. Compute the covariance matrix dxd: $C = A^T A / N$

3. Find "eigenvectors" of the covariance form (perpendicular).

4. Select "m eigenvectors" which match to the heights "m eigenvalues" to be the modern area dimensions. Variance in every new dimension is given by the eigenvalues.

### C.  Text representation

The document indexing  is one of the most important mechanism for preprocessing which is applied to minimize the difficulty of the records and And create themteasyttothandle, the document must be converted from thetwhole documenttversion totdocumenttvector. The Perhaps should commonly used document ntationt is called vector space model.vector space model, represented via `vectors`to words.Usually one has group of documents that is performed by "word by word." document Matrix "Matrix BoW/VSM" impersonation planner has its own limitations. Several of them are: towering dimensional from the performance, lack of linage with relative words, and the damage of 'semantic' connection which live between expression within the document. in order to avoid that troubles, feature vector space are applied for give suitable weights for word as shown

$$
\begin{pmatrix}
& T_1 & T_2 & \ldots & T_{at} & c_i \\
D_1 & w_{11} & w21 & \ldots & w_{t1} & c_1 \\
D_2 & w_{12} & w22 & \ldots & w_{t2} & c_2 \\
\vdots & \vdots & \vdots & & \vdots \\
Dn & w_{1n} & w2n & \ldots & w_{tn} & C_n
\end{pmatrix}
$$

in above matrix [11]. In this matrix every input carry out appearance such term with in the paper, wherever wtn is the weight of term i in document n.Because each word does not normally become visible

within each document .Thereare different methode to defining the weight w11. Such as "Boolean weighting", word frequencyweighting, tf-idf, entropy etc. however main disadvantage for this system is that itoutcome in hugesparse matrix, that elevate problem of large dimensionality [14] .

## D. Feature Selection FS

Feature selection has great impact within data mining in general and text mining in particular, where FS is an active and effective area of research since the 1970s. FS is An important and fundamental issue within categorization and recognition of pattern as most standing learning algorithms are not created to compact with large size of attributes space. Within document classification problem, FS aims to enhance the classification reliability and computational efficiency of pattern recognition and classification techniques by deleting redundant terms and irrelevant from the corpus, It is as well applied to elicit attributes that hold enough information about the text dataset". Feature selection has two common methods: "wrapper." and "filter," . Within the wrapper approach, A partial set from the main attributes is elected depending on the accuracy of the classifiers while in the filter approach, asubset of features is selected or filtered using feature scoring Metric [15][16][17].

_ For assistance within constructing accurate and fast pattern recognition , classification samples via reject unnecessary and rowdy terms

_ Take off undesirable attributes leadsfor decreasing work-time and minimize wanted, storage media.

_ Boost the generalization ability of pattern recognition and classification models.

## E. Model Classification by Support Vector Machine

Support Vector Machine and "their Variables and stretching", often known as kernel-based ways (or merely kernel methods), is set of supervised machine learning, SVM has been worked extensively and used to difference function approximation trouble and pattern classification. model sorting applied for assort several topic into one between the specified classes re-owned as groups. The classification of Pattern applied for assort several topic into one between the specified classes renowned as groups.in picked pattern classification problem classifier, which is computer software, is progressing thus, objects are distributing properly moderately smart accuracy. The features are the input to the classifier , and for this reason they are specified because they act every object good or so that data pertinence for other types are completely unattached, in the input space [20]. Within task of SVM the input data was divided into a coaching true instance part and false instance part. It as well make "hyper planes" like a line between true and false examples that be the upper. Those hyper planes do optimal solution

based on notion, of constitutional change minimization. The conceptual framework, of SVM is shown in Figure 5. SVM calculate hyper planes that break away true part from false part in hyper space. We define a space between the amidst true-portion, hyper plane closer the false case and false-part hyper plane nearest the true examples the margin [21].
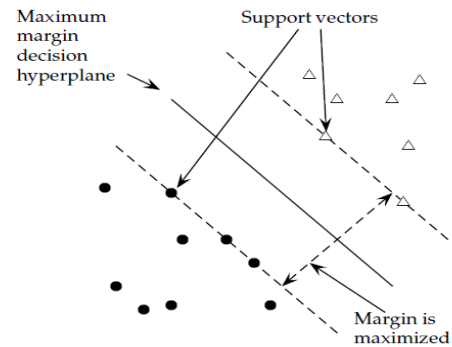


Figure 5: the hyperplane, support vectors laying on supporting hyper planes.

- *Support vector machine for Binary Classification:* SVM are a mostly applicable method for machine learning. It can be apply on binary ranking as linear separable state: assume we are specified with coaching set 'xi', plus with the symbol rates $y_i \in \{-1,1\}$. SVM seeking into dividing hyper plane, which disconnect affirmative plus passive instances for all another togather with greater side, on other meaning, the range of the ruling superficial with relative example is maximal [22] (see Figure 6). The decision function can bewritten in the following form:

$$f(x) = \text{sign}(\sum_{i=1}^{N} \alpha_i \cdot y_i \cdot K(\mathbf{x}_i, \mathbf{x}) + b)$$
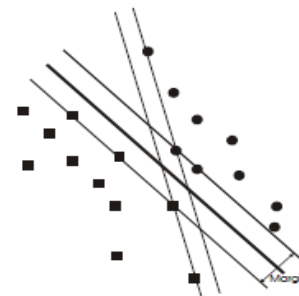


Figure 6: The best ruling roof with greater margin vs. a non-optimal ruling surface.

- *SVMs for Multi class Classification*: Oftentimes, datum are obtained more than two categories in the case of more than two group state , x∈ r may be taken as random r-vector selection for sorting purposes and Y ∈ {1, 2, . . .,N} is a class label, wherever N that the categories numbers . So as that SVM classifiers are designed for under 2 categories, we want to understand How expanded methodology of SVM to acknowledge inter alia N >2 classes [20].

## F. Model Classification by Neural Network

The neural networks area unit made from a large variety of components with Associate in nursing the entries size greater than within computational pieces of imitative architectures. these ones elements, chiefly 'artificial neuron' are interrelated, for set using numerical system to the datum handling based on linking approach to computation, figure 7. Appeared "Neural network". The NN, create their neuron sensitive to store item. Various styles from neural network methods are enforced to the task of document compartmentalization. Many researchers have applied the "single layer." perceptron, that consists of only from `Input.' and `Output.' Without any number of hidden layer. Inputs area unit fed in the products via groups of weights. during the approach it will be thought-about the best reasonably `feed-forward.' network. The `multi-layer.' perceptron that is additional refined, that consists of associate input layer, one or more hidden layers, and an output layer in its structure, also widely implemented for sorting functions [24].
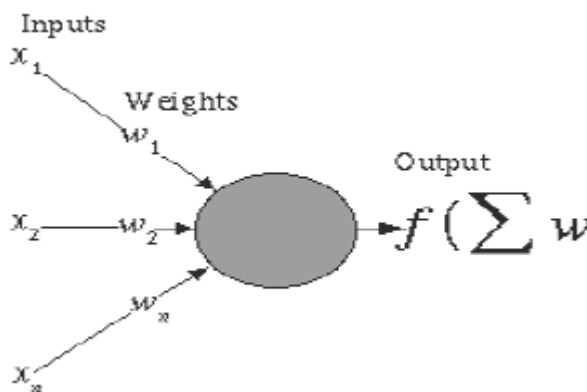


Figure 7:  Artificial Neural Network

The back-propagation (BP) algorithm impart the classification sample by employment a multilayer "feed-forward,". NN. The general public design of the neural network for BP is shown within the following diagrams, with some hidden layers ,one input layer, and one output layer. every layer embody some units or perceptron. every unit can be connected to others by weighted connections.

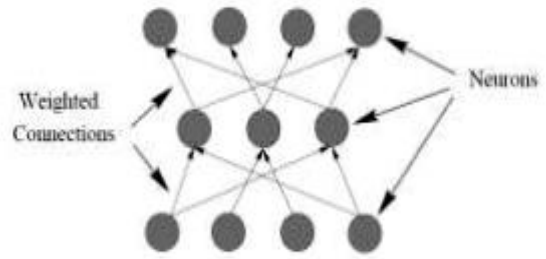Before the coaching, The values of the weights area unit initialized [84].



Figure 8: Backpropagation Network

## V. METHODOLOGY

### A. Data-Set

For the purpose of achievement of these assortment techniques, input data summation was obtained from tReuters-21578, distribution 1.0 test collectiont. "There are a group of 21,578 of newswire stories of Reuters classified in to several categories by personnel from Reuters Ltd,"(Lewis;2004). And Carnegie Group, Inc in 1987 and were further formated by David D. Lewis and Peter Shoemaker in 1991 [26]. There are a total of 674 categories in Reuters – 21578 collections .They are totally divided in to five fields. Each field has several categories of document collection. in our work,we work on 5000 documents deviding between the trainig and testing set. Table 1 shows topic category with texts numbers for each class.

TABLE I. The numbers of document for each classes

| Category | Document Numbers |
|---|---|
| earn | 500 |
| Acqu | 500 |
| Money-supply | 500 |
| trade | 500 |
| Crude | 500 |
| coffee | 500 |
| grain | 500 |
| interest | 500 |
| Ship | 500 |
| Corn | 500 |

### B. Expermantal Results

After preprocessing for document the second process the feature extraction using Singular Valued Decomposition (SVD) by applying this technique the total number of features (high dimension input space) will be reduced in order to achieve best result. The next significant step within the categorization of text document is applying of feature selection FS is important operation in the field

of text mining in this paper we applied mutual information (MI), this method compute the scoring for each term after that the term with high scoring will be chosen. As additional step we applied another technique to be worked on the selected features from previous step ,this technique is principle component analysis (PCA) for dimension reduction (DR) in the last step we are making the classification using two approach, Back-Propagation Neural Network and Support Vector Machine

TABLE II. The result of support vector machine

| Class name | Recall | F1 Measure | Precision | Accuracy |
|---|---|---|---|---|
| Earn | 8.29% | 8.84% | 9.48% | 8.84% |
| Acq | 8.56% | 9.02% | 9.52% | 9.02% |
| Trade | 9.31% | 9.37% | 9.43% | 9.37% |
| Money-fx | 9.52% | 9.50% | 9.49% | 9.50% |
| Interest | 9.83% | 9.68% | 9.53% | 9.68% |
| Ship | 9.85% | 9.65% | 9.45% | 9.65% |
| Sugar | 9.85% | 9.64% | 9.43% | 9.64% |
| Coffee | 9.96% | 9.72% | 9.49% | 9.72% |
| Gold | 9.98% | 9.70% | 9.43% | 9.70% |
| Crude | 9.99% | 9.74% | 9.51% | 9.74% |
| Average | 9.57% | 9.49% | 9.48% | 9.49% |

As shown in Table II the result of document classification using( SVM) algorithm after applying Feature selection (FS) using Mutual Information technique (MI) and dimension reduction using (PCA).Within the feature selection method sub-group of attributes will be chosen depending on their Scoring metrics ( Attribute with high value) Since the total number of features is (400) terms while after applying of the FS , (200) terms have been chosen depending on their scoring metrics. So this group also reduced to (172) terms by the Principle component Analysis technique that reduced this set of features. The average result for SVM algorithm with Dimension Reduction techniques is : precision : 94.8% . Recall : 95.7% and Accuracy: 94.9%.

Table 3 shows the result of document classification using Back Propagation Neural Network (BPNN) algorithm after applying Feature selection (FS) using Mutual Information technique(MI). Also there are numbers of features will be selected as a partial set from the main group Depending on their importance. After that

this partial set of attributes will be decreased to another smaller group. The average result for SVM algorithm with dimension reduction techniques is: precision : 99.4% . Recall : 99.4% and Accuracy: 99.4%.

TABLE III. The result of BPNN

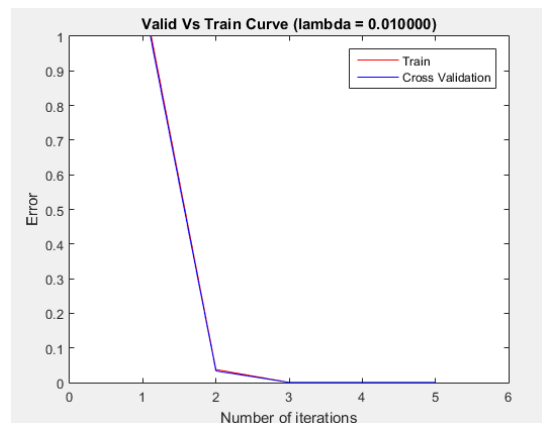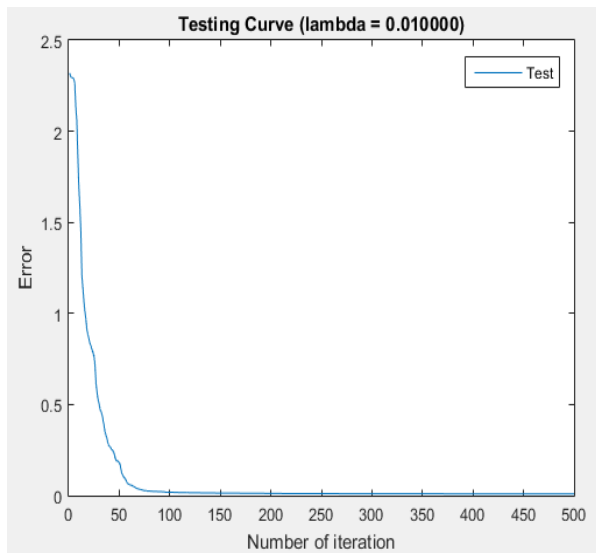| Class number | Accuracy | F1 Measure | precision |
|---|---|---|---|
| Earn | 98.6% | 98.6% | 99.2% |
| Acq | 99.0% | 99.0% | 99.6% |
| Trade | 99.3% | 99.3% | 99.9% |
| Money-fx | 99.6% | 99.6% | 99.2% |
| Interest | 99.7% | 99.7% | 99.7% |
| Ship | 99.6% | 99.6% | 99.6% |
| Sugar | 99.6% | 99.6% | 99.0% |
| Coffee | 99.7% | 99.7% | 99.8% |
| Gold | 99.5% | 99.5% | 99.1% |
| Crude | 99.7% | 99.7% | 99.6% |
| Average | 99.4% | 99.4% | 99.4% |



Figure 9: The curve of BPNN for (Training vs. Valid).

Figure 10: Testing curve for BPNN

Durin the total comparison between BPNN and SVM for Reuters dataset, the outcomes through whole status for both precision and recall indicate significantly variation in the implementation of back propagation neural network over support vector machine algorithm and of the decrease attributes group over the larger feature set. It is possible to realize that the SVM algorithm is less computationally computational than the BPNN algorithm and require less time for execution over Back prppagation Neural Network, but the result show on this type of dataset (example) Back propagation Neyral Network perfoem best than Support vector machine, moreover the neural network hold to executed massive matrix computions on matrices as within much lines as features. likewise we found that decreasing the vector volume via half does not own passive impact on showing ,really get better production, and is hence too more suitable.

## VI. CONCLUSIONS

Text categorization refers to the procedure of allocating one or more category to the text document. In this work we studied and applied two types of important algorithms required in the process of classification of texts which are support vector machine and back Propagation neural network. Both algorithms work with high dimension input space, in which the dimension of attributes space is very large. So this dimensionality can be reduced via applying the techniques of dimension reduction. Within this work we are applying Singular value decomposition and Principle Component Analysis for Feature extraction and Mutual information for Feature selection .The data sets of this experiments gathering from Reuters data/set. mat file. The experimental results demonstrated that BPNN overcome the result of SVM at the degree of reliability with recall. But Neural Network take more time for execution than Support vector machine

.In future work we want to improve the performance and get the best results accuracy by using Hybrid classifier.

## REFERENCES

[1] Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., and Spyropoulos, C. D. "An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages". In Proceedings of SIGIR-00, 23rd ACM International Conference on Research and Development in Information Retrieval (Athens, GR, 2000), pp. 160–167.

[2] Janez Brank, Marko Grobelnik, Nataˇsa Miliˊc-Frayling, Dunja Mladeniˊc" Interaction of feature selection methods and linear classification model"s. Proc. ICML-2002 Workshop on Text Learning, pp. 12–17, 2002. Longer version available as Microsoft Research Technical Report MSR-TR-2002-63, 12 June 2002.34

[3] Kwok, J.T-K. " Automated Text Categorization Using Support Vector Machine". *Proceedings of the International Conference on Neural Information Processing (ICONIP)1998.*

[4] Ashis Kumar Mandal1and Rikta Sen "Supervised Learning MethodsFor Bangla Web Document Categorization" International Journal ofArtificial Intelligence & Applications (IJAIA), Vol. 5, No. 5, September2014 DOI.

[5] Erlin,Unang Rio "Text Message Categorization of Collaborative Learning Skills in Online Discussion Using Support Vector Machine".International Conference on Computer, Control, Informatics and Its Applications. 2013

[6] M. A. Wajeed and T. Adilakshmi, "Text Classification Using Machine Learning," Journal of Theoretical and Applied Information Technology, vol. 7, no. 2,pp. 119_123, 2009.

[7] Fabrizio Sebastiani. *"automatic Text classification",*. In Keith Brown (ed.), The Encyclopedia of Language and Linguistics, 2ndEdition, Vol. 14, Elsevier Science, Amsterdam, NL, 2004.

[8] Addis, A. "Study and Development of Novel Techniques for Hierarchical Text Categorization*".* PhD Thesis, Electrical and Electronic Engineering Dept., University of Cagliari, Italy.2010

[9] Dr. Ronen Feldman and James Sanger,*"THE TEXT MINING HANDBOOK"*, Advanced Approaches in Analyzing Unstructured Data, Published in the United States of America by Cambridge University Press, New York,2007.

[10] Ronen Feldman and James Sanger. "The Text Mining Handbook:Advanced Approaches in Analyzing Unstructured Data". Cambridge University Press, New York, NY, USA, 2007.

[11] G. Forman, *"An extensive empirical study of feature selection metrics for text classification,"* The Journal of Machine Learning Research, vol. 3, pp. 1289{ 1305, 2003.

[12] A. Dasgupta, *"Feature selection methods for text classification.",*In Proceedings of the 13th ACMSIGKDD international conference on Knowledge discovery and data mining,pp. 230 -239, 2007

[13] S. M. Kamruzzaman1, Farhana Haider and Ahmed Ryadh Hasan," Text Classification Using Data Mining",2005.

[14] J. Ramos, "Using tf-idf to determine word relevance in document queries," inProceedings of the First Instructional Conference on Machine Learning, 2003.

[15] Anirban Dasgupta "Feature Selection Methods for Text Classification "KDD'07, August 12–15,2007.

[16] Y. Saeys, I. Inza, and P. Larra~naga, "A review of feature selection techniquesin bioinformatics," Bioinformatics, vol. 23, no. 19, pp. 2507{2517, 2007.

[17] M. Aljlayl and O. Frieder, *"On arabic search: improving the retrieval effectiveness via a light stemming approach,"* in

Proceedings of the eleventh inter- national conference on Information and knowledge management, pp. 340{347, ACM, 2002.

[18] Kiritchenko, S. "Hierarchical Text Categorization and Its Application to Bioinformatics". PhD Thesis, School of Information Technology and Engineering, Faculty of Engineering, University of Ottawa, Ottawa, Canada.2005.

[19] Z. Markov and D. T. Larose, "Data mining the Web: uncovering patterns inWeb content, structure, and usage." Wiley-Interscience, 2007.

[20] Shigeo Abe, "support vector machine for pattern recognition" second edition ,springer, new york ,2010.

[21] Gunn, S., "Support Vector Machines for Classification and Regression". http://homepages.cae.wisc.edu/~ece539/software/svmtoolbox/svm. pdf, (Date:05/ 02/06).

[22] Thorsten Joachims" Text categorization with support vector machines" learning with many relevant features, Proc. of ECML-98, 10th EuropeanConference on Machine Learning, Springer Verlag, Heidelberg, DE, ,pp. 137-142.1998

[23] Christopher J. C. Burges" A tutorial on support vector machines for patternRecognition". Data Mining and Knowledge Discovery, 2(2), pp. 955-974.1998

[24] Miguel E. Ruiz, Padmini Srinivasan; "Automatic Text Categorization Using Neural Network", In Proceedings of the 8th ASIS SIG/CR Workshop on Classification Research, pp. 59-72. 1998.

[25] Mikhail Krivenko, Vitaly Vasilyev " Sequential Latent Semantic Indexing" ,proceedings of the 2nd Workshop on Data Mining using Matrices and Tensors ,Article no3 , 2009.

[26] David D. Lewis, 'Reuters 21578, Distribution 1.0 Test collection'(n.d.)www.daviddlewis.com/resources/testcollections/reuters21578/