Text Categorization using Semantic and Discriminative Selected Features

Madhuri B.Bhusare
Department of Computer Engineering
G.E.Ss.R.H.Sapat C.O.E.M.S.and R, Nashik.
Savitribai Phule Pune University, Pune
madhuribhusare128@gmail.com
Chandrakant R. Barde
Department of Computer Engineering
G.E.Ss.R.H.Sapat C.O.E.M.S.and R, Nashik.
Savitribai Phule Pune University, Pune
erchandubarde@gmail.com

***Abstract*—Data mining is also known as knowledge discovery is a task of analysing data from several outlook and summarize it into useful information. It has huge application in the area of classification such as pattern identification. Feature selection has been a condemnatory exploration subject in pattern mining, in light of the factor that the actual datasets frequently have huge dimensional elements, ex. The bio-computing and content data mining applications. Lot of a ongoingchannelfeatureselectionmethod-rankapproachbyadvancingcertainelementrankingorder,such that interconnected elements normally have comparable rankings. These connected components are tedious and don't give considerable shared data which help data mining. So, when we select a set number of features, we plan to select the top non-repetitive elements such that the expensive common data can be boost. Feature Selection model to minimize the repetition between accordingly selected features. In any case, this structure uses the impatient hunt, through these lines the global component wasn't reviewed and the outcome are not perfect. Text categorization is a set of categories and collection of document, the activity of searching topic for each document. Predefine categories and ladled document classify newer document and done standard classification problem. Single label categorization, each document belongs to exactly one category. For improving the performance text categorization here use Kullback-Leibler (KL) Divergence and Jefferys Divergence as binomial representation that matter type I and type II errors of Bayesian classifier. Then introduce a new divergence measure called Jefferys Multi Hypothesis (JMH) Divergence for multiclass classification. Develop two efficient method of feature selection first is maximum discrimination and second is maximum discriminationX^2 . Further to improve the bios capacity feature selection algorithm by weighting each distinct feature. The promising results of extensive experiment demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Feature selection, Feature Ranking, Redundancy Minimization, Kullback-Leibler (KL), Jefferys Multi Hypothesis (JMH) Divergence.**

I. INTRODUCTION

Automated text classification is a particularly challenging task in modern data analysis, both from an empirical and From a theoretical perspective[2]. This problem is of central interest in many internet applications, and Consequently. It has received attention from researchers in such diverse areas as information retrieval. Late fast up grades in investigation and progressing advancements in data transformation empower us together tremendous measure so information. A lot of information mining and machine learning procedure have been develop to breakdown and comprehend the exploratory information for exclusive applications from that feature selection is one of more authoritative methods, and can upgrade other information mining assignments[6][27][30]. The fundamental approach that a maximal grouping of contiguous word characters in the content stream constitutes a word. Common content preparing applications standardize the upper forcing so as to case of every word each character is into lowercase. Let this change controlled by the character function to Lower Case(char). The survey is not important for some Unicode letters. Furthermore, if underscores or digit share absorb among the word characters, then their lowercase mapping is the identity function. As the objective of features selection is to choose a reduced subset of features to communicate with information, expect the chose features can give best shared data target fluctuating. There are three category of feature selection methods: filter method, wrapper method, and embedded method.[3]

The filter methods have low performance expense, yet they choose features frequently can't accomplish great grouping execution. The features chose by wrapper of ten than not have very large execution. Be that as it may, the wrapper methods use grouping results to choose feature[6], consequently their computational expenses high and are not suitable for expansive scale applications. The embedded methods fu se feature pursuit and order modelling to as solitary streamlining issue, and normally is fast earthen the wrapper methods and slower than the filter methods. Inspire of the fact that there are a wide range of sorts of feature selection approaches, their component is the same, i.e. every one of them use distinctive approaches to rank features, for example, score capacity, grouping results, weights from the model parameter grid. Subsequently, numerous top positioned features are frequently associated to one another. From measurement perspectives, these couple features are excess and more uninterested features may not acquaint additional expensive data with help information mining. So, when select a breaking point number of features, would like to accept the top features such not refuse that the helpful common data can be augmented. Feature repetition can be measured by the cosine likeness between features. Here utilize world wide repetition to speak to the aggregate of excess of a feature repetition with every single other feature. In past inspection, Picked this exceptive issue and proposed the base demanding most relevance Feature Selection model to reduce the overhead between important selected features[6]. To mark this issue suggest one more feature selection method to among many features reduce feature duplication with extending the place of scores. The feature location accordingly of any feature selection process can be used as the information of our new structure. The proposed model minimizes the global repetition and purifies the location scores, such that the feature rankings are advance on ward. Our new structure can be attached to both unsupervised and managed feature selections procedure. Experimental result on standard information sets indicate that the proposed redundancy reducing system reliably boost the feature selection results compare with the first methods.

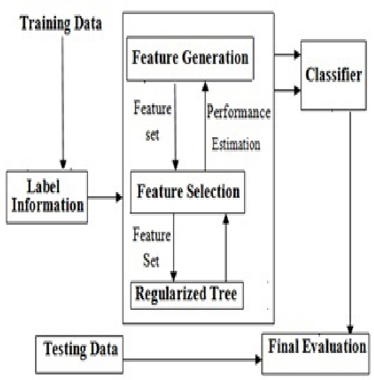

Fig. 1. Document classification on the basis Feature.

II. Literature Survey

This paper introduces support vector machine for text categorization. It provides both theoretical and empirical evidence that SVMS are very well suitable for text categorization the theoretical analysis concludes that SVMS acknowledge the particular properties of text. i.e. a. High dimensional features spaces. b. Few irrelevant features c. Sparse instance vector [1].
Our system develops a new approach towards the automatic text categorization. Model learn and then it is used for classifying the feature document. This categorization approach is derived from a machine learning paradigms which known as dynamic learning. In which we retrieve advance document is known as retrieval feedback [2][27].

In most of resent text categorization research focuses on addressing specific issue in text categorization [Ex. Feature Selection, Dimensional Reduction] linear measure for feature selection are very few new approaches to be advise [21][30].

In recent days the problem of ranking has gained much attention in machine learning ranking methods may filter feature to reduce dimensionality of the feature space this is very much effective for classification method that do not have any inherent feature selection built in Ex. Nearest neighbour method, some type of neural network[17].

The system have demon striated that hierarchical neural language model can actually out perform its non-hierarchical counter parts and achieve state of art performance the main motto is to making a hierarchical model perform well is using to carefully constructed hierarchy over words[7][9].

This paper describe paragraph vector, an unsupervised learning algorithms That learn vector represent for variable length pieces of text such as sentences and documents the vector representation are learned to predict the surrounding word in context. Sampled from the paragraph [25].

For successful information retrieval, the Naive Bayes model has been used, producing some of the best result. In the recent comparisons of learning method for text categorisation have been somewhat less favourable nave bayes model ,while still showing them achieve respectable effectiveness[24].
The recent increasing of system that hierarchically organize massive amounts of text based documents calls algorithms that hierarchically categorization new document as they come in. The paper describes an approach which utilizes the existing reach hierarchical structure in order to facilitate this process [11][12].
This paper present an extensive comparative study for feature selection matrix for high dimensional domain of text classification. While focusing on support vector machine and to class problems typically with high class skew, it has revelled the surprising performance of a new feature selection matrix by normal separation [15].

By achieving the equivalent Feature Selection within the con-text of Well-founded bastion, the losso approach provides the great scope for Domain knowledge. Combining with learning from training data. Simplest care would be to give rare world prioress with higher variance, reflecting that they have higher contain than more common words words and distribution clustering and learning logic.[4].

The paper has compared the theory and practice of to different first order probabilistic classifier, both of which make the Naive Bayes assumption. The multinomial model is found to be almost uniformly better than multivariate born Olly model in impaired result. On five real word corpora. We find that the multi nominal model reduces error by an average of 27% and sometimes by more than 50% [10].

In this paper we have discuss and evaluated experimentally in span filtering Context 5 different versions of naive Bayes classifier this investigation included two version of naive Bayes that have not been used widely in spam filtering literature namely flexible Bayes and multinomial nave Bayes with bullion attribute for classification of text document[13][18].

Our system introduces a new filtering measure for feature selection in text categorization. Which have simple expression in terms of appearance of the word in the different documents and text categorization study for feature selection It also shows that these measures have interesting property.[14]

The system proposes a distributed Naive Bayes text classification model with wait enhancing method. These module assumes a documents is generated by multivariate distribution module so the system suggest per document term frequency normalization to estimate the distribution parameters [5][13].

Covetous process for reducing feature to take out the unwanted features: if one element fs is chosen, different features extremely according with fs are refused and these declined features may have low globally repetition on the whole, so the not identical element selection results. For instance, the reare four features with positioning scores:9.9,9.7,9.3,2 (higher score indicates more discriminative), the location score of an element is fundamentally controlled by the connection between the component and there sponser variable.Logistic regressions for text classification[16]
The model proposed in our work is without parameter. More vital, there location ship framework utilized as apart of may not be sure semi-unmistakable, which prompt a non curved issue, and the worldwide ideal cannot be acquired. Conversely, the positive semi-definiteness of the relationship grid is ensured in our work[6][7]
Most research in speeding up text mining involves algorithmic improvement so induction algorithms, and yet form any large scale applications, such as classifying or indexing large document repositories, the time spent extracting word features from text scan itself greatly exceed the initial training time[5][14]. Naive Bayes classifier has been widely used for text categorization due to its simplicity and efficiency It is a model-based classification method and offers competitive classification performance for text categorization compared with other data-driven classification methods, such as neu- ral network, support vector machine (SVM), logistic regression, and k-nearest neighbors. The naive Bayes applies the Bayes theorem with the naive assump- tion that any pair of features are independent for a given class[18][19][29]
A fast method for text feature extraction that folds together Unicode conversion, forced lowercasing ,word boundary detection, and string hash computation. It show empirically that our integer hash features result in classifiers with equivalent statistical performance to those built using string word features, but require fearless computation and less memory. The obviously Feature selection is to choose significant[26][28] and instructive elements from the high-dimensional space, and assumes critical part in numerous investigative and down to earth applications, on the grounds that it can accelerate the learning procedure; enhance the mode speculation ability, and abatement the calculation running time in the genuine applications.

III. Proposed System

A. Problem Statement

- Feature selection method can ranks and also assign weight to the original features, so maximize the discriminative performance for text categorization, when naive Bayes classifiers are used as learning algorithms.
- Efficient approach to rank the order of features to approximately produce the maximum JMH divergence. The theoretical analysis shows that the JMH divergence is monotonically increasing when more features are selected.

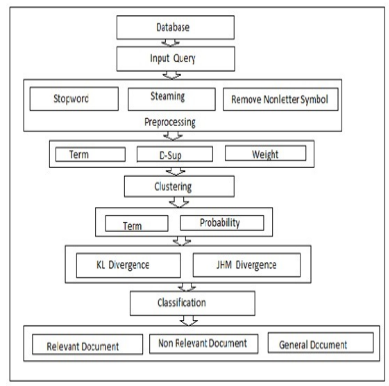

Fig. 2. Architecture of system

Regardless of the way that there are an extensive variety of sorts of feature selection approaches, their instrument is the same, i.e. each one of them use different ways to deal with rank features, for instance, score limit, classification results, weights from the model parameter structure[33]. Associated features regularly tend to get similar rankings, in light of the fact that they are viewed as pretty much as imperative for classification[2][14]. In this way, various top situated features are consistently compared to each other. From estimations viewpoint, these associated features are redundant and more redundant features may not familiarize extra significant information with help data mining. Henceforth, when select a cut off number of features; here need to pick the top non- redundant features such that the supportive regular information can be increased.

Feature Selection

Feature selection is one of the most necessary data preprocessing steps in data mining and knowledge engineering. Let us say we are interested in a task which is finding employees prone to attritions[2][6]. Each employee is represented by various attributes/features like their age, designation, marital status, average working hours, average number of leaves taken, take-home salary, last ratings, last increments, number of awards received, number of hours spent in training time from the last promotion, and so forth[20][22]. In feature selection, the idea is to select best few features from the above, so as to we perform equivalently in performing the task, in terms of some evaluation measure,

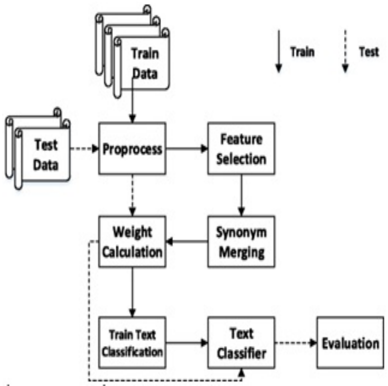

Fig. 3. Feature selection on the basis of weight

So for a classification task, a standard evaluation measure like classification accuracy and Score, and so forth, and for clustering it can be internal measures like silhouette width or an external measure like purity.

Feature selection offers the following three advantages:(i)better model understand ability and visualization: it might not be possible to reduce to a two-dimensional or a three-dimensional feature set, but even if we want to visualize with a combination of two or three features, the combinations will be much lesser in the reduced feature space;(ii)generalization of the model and reduction over fitting: as a result better learning accuracy is achieved;(iii) efficiency in terms of time and space complexity: for both training and execution time.

Information Gain

Measure the information if one knows the presence or absence of a term in a document, which is defined as

$$IG(t) = -\sum_{i=1}^{m} p(c_i)\log p(c_i) + p(t)\sum_{i=1}^{m} p(c_i|t)\log p(c_i|t) + p(\bar{t})\sum_{i=1}^{m} p(c_i|\bar{t})\log p(c_i|\bar{t}) \tag{1}$$

Feature Reduction:
We can reduce data by eliminating feature to obtain accuracy in data mining and remove noise[19][21]. It also compress the data for efficient retrieval and storage purpose. Data are pre-process for good pattern mining and machine learning. Diversion Measures for Binary Hypothesis Testing: In probability theory and information theory the Kullback-Leibler divergence also called discrimination information[26].

$$\begin{aligned} KL(P_1, P_2) &= \int_x p(x|H_1)\log\frac{p(x|H_1)}{p(x|H_2)}dx \\ &= E_{p1}\log\frac{p(x|H_1)}{p(x|H_2)} \end{aligned} \tag{2}$$

Description
Jeffreys Multi Hypothesis Divergence: [29][32]first generation to next multi hypothesis divergence uses scheme that one verse all.It can also calculate N binary hypothesis testing detectors. Jensen Shannon used in calculating the similarity between two properties. Divergence is the one that can be used to measure multi-distribution divergence, In that divergences of each individual distribution with a reference distribution are computed and sum together[32][33].
Let $P = P_1, P_2, , P_N$ be the set of N distributions. The Jeffreys-Multi-Hypothesis (JMH) divergence, denoted by JMH$P_1, P_2, , P_N$, is defined to be

$$JMH(P_1, P_2, ..., P_N) = \sum_{i=1}^{N} KL(P_1, P_2) \tag{3}$$

Rank Feature Index:
Search engine can calculate the relevance rank. Ranking model has collection of ranking features to calculate the rank score that document. It can take data from the search index from that document.

- Weighting to feature in that each feature in a document is assign a feature weight based on a weighting scheme
- Text Categorization Naive Bayes Classification: Naive Bayes is a simple and efficient technique for classification. It assign class labels to data instances, represented as feature values, where the class labels are drawn from some finite set of document. All naive Bayes classifiers.
  Term Frequency(TF)

$$\begin{gathered} TF = \frac{tf}{doc-length} \qquad TF = \frac{tf}{maxtf_d} \\ TF = \frac{tf}{tf + 0.5 + 1.5 * \frac{doc-length}{avg-doc-length}} \end{gathered} \tag{4}$$

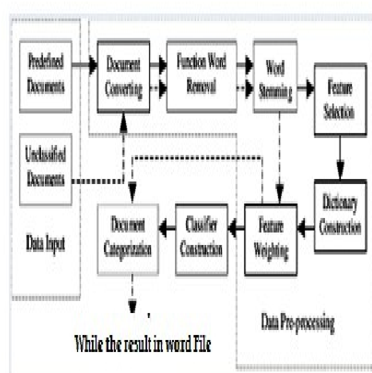

Fig. 4. System Working

Text classification system have been adopted by a growing number of organizations to effectively manage the ever growing inflow of unstructured information. The goal of text classification Systems is to increase discoverability of information and make all the knowledge discovered available or actionable to support strategic decision making. Essentially there are really three text classification algorithm[8][10].

1) Mannual approach : The bag of keywords , it requires compiling a list of key terms that qualifies the type of content in question to particular topic[13].
2) Statistical approach : This depends on manual tagging and identification of a training set of documents that covers the same topic.
3) Rule Based approach : This classification depends on linguistic rules that capture all of the attribute and the feature of a document to assign it to a category. Rules can be written manually or generated with an automatic analysis and validated manually[2].

Algorithm Discriminated Feature Selection

Input : A training dataset D_{m*n} with m samples and n features in space F and the target C,P Predefine parameters;
Output : Selected feature subset S;
Step 1: Initialize Parameters : $S = \phi$ // Selected feature subset
Step 2: Group training sample D by class ; // Pre-processing the training data D
Step 3: While $(|S| \leq P)$
Step 4: For each Feature f in F
$S \leftarrow S \cup (f)$
Constructing the $|S| - FSGin|S|$- simentional feature space and calculate $Q_{f_i \cup s}$
// Analyszing the relevant independency between f and selected subset S//
End For
Step 5: Selcted the feature F with maximum value of $Q_{f_i \cup s}$
Set $S \leftarrow S \cup f$.
$F \leftarrow F/(f)$; // Selecting the Feature with max RI
Step 6: End While

B. Mathematical Model

S : D, KL, JHM, RF, AX, NB, CI

Where, S = is a System,
D = Input dataset,
KL = Calculate K L Divergence, JHM = JHM Divergence,
RF = Rank Feature Index, AX = Apply X^2,
NB = Nave Bayes
CL =Classification

Y : KL, JHM, RF, AX, NB, CI

D: d1, d2,., dn Set of Web Data.
F: f1, f2,.., fn Functions for Feature Selection.
Y is a set of techniques use in our Application.
Start State(S)
At client side: text classification from collection of document.
At server side: Perform Calculation.
End State (E)
At client side: class label get and do verification.
At Server side: According to user requirement send classification results.
Input (I): Data set send to the system which are further get classified.
Output (O): According to user requirement data set get classified on system.
Function (F): Probability calculation, Computing KL Divergence and JMH Divergence, Rank Feature, Weighting Feature
F1: Calculate KL Divergence.

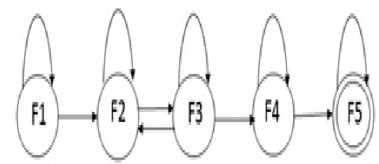

Fig. 5. State Transition Diagram.

F2: Calculate JHM Divergence.
F3: Rank Feature Index.
F4: Apply X^2.
F5: Naive Bayes Classification.

IV. EXPERIMENTAL SETUP AND RESULTS

Table 1 and Table 2 shows the comparison of existing systems with proposed system. There are various existing system with their classifier is compare with proposed system .Proposed System is distinguished with the high value of precision from all existing system. Proposed System also represent the higher average of weighting feature selection algorithm with better accuracy and precision etc. hence table shows the proposed system with the high average values of all classifier.

The proposed system has introduce new feature selection on the basis of information measures for Naive Bayes classifiers, aiming to select the features that offer the

TABLE I
CLASSIFICATION TABLE FOR PRECISION.

Classification	20-News Gruop	TDT2	Reuters
KNN	7.3	3.6	3.2
K-means	4.6	7.4	5.4
Multinomial NB	7.6	3.3	4.7
SVM	5.5	3.7	8.7
Wfeature	9.7	8.6	8.9

TABLE II
CLASSIFICATION TABLE FOR ACCURACY

Classification	20-News Gruop	TDT2	Reuters
KNN	87.56%	82.80%	82.31%
K-means	82.87%	87.28%	83.23%
Multinomial NB	85.45%	80.92%	85.56%
SVM	87.23%	82.39%	92.12%
Wfeature	93.50%	96.24%	97.20%

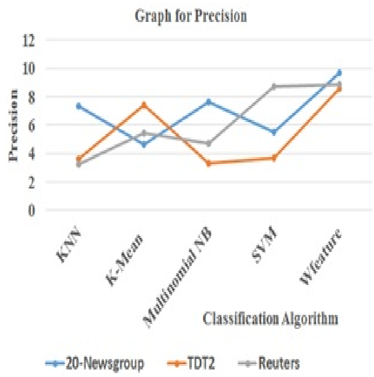

Fig. 6. Classification graph for precision.

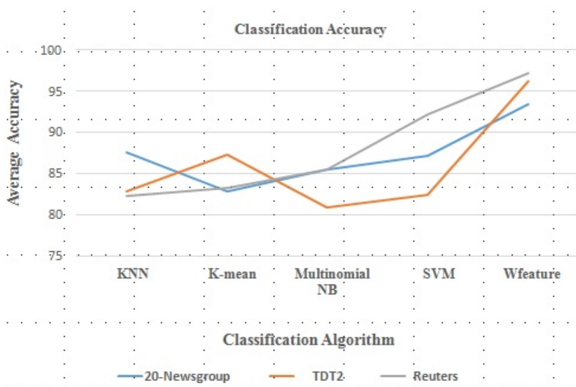

Fig. 7. Classification graph for accuracy.

maximum discriminative capacity for text classification, Thus we can achieve the optimal feature selection in Naive Bayes for text categorization. We had carry out experiments on the dataset mentioned in paper (Reuters) when naive Bayes and SVM are used as classifiers. To compare the performance of these feature selection methods, we evaluate the classification-measure accuracy and precision metric of these classifiers with different number of features ranging from 10 to 2000. With different database like 20-newsgroup, TDT2 and Reuters and we obtain better result from weighted feature method.

V. OUTCOMES SUCCESS DEFINITION OF WORK

Introduced new feature selection approaches based on the information measures for naive Bayes classifiers, aiming to select the features that offer the maximum discriminative capacity for text classification. It also develop feature selection algorithms by weighting each individual features, aiming to maximize the discriminative capacity.

VI. CONCLUSION

The proposed system has introduce new feature selection on the basis of information measures for Naive Bayes classifiers, aiming to select the features that offer the maximum discriminative capacity for text classification, Thus we can achieve the optimal feature selection in Naive Bayes for text categorization. Proposed System Also work better at precision and accuracy based classifier as shown in graph.

ACKNOWLEDGMENT

I would like to express gratefulness to P.G Dept. of Computer Engineering, Gokhale. Education. Society, R.H. Sapat C.O.E. and Research Centre, Nasik.

REFERENCES

[1] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features” , in Proc. Eur. Conf. Mach. Learn., 1998, pp. 137142.
[2] W. Lam, M. Ruiz, and P. Srinivas an,“Automatic text categorization and its application to text retrieval”, IEEE Trans. Knowl. Data Eng., vol. 11, no. 6, pp. 865879, Nov./Dec. 1999.
[3] I. Witten and E. Frank, “Data Mining: Practical Machine Learning Tools and Techniques”, Second Edition, ser. The Morgan KaufmannSeries in Data Management Systems. Elsevier Science, 2005
[4] H. Al-Mubaid and S. Umair, “A new text categorization technique using distributional clustering and learning logic”, IEEE Trans. Knowl. Data Eng., vol. 18, no. 9, pp. 11561165, Sep. 2006.
[5] S. Haykin, “Neural Networks: A Comprehensive Foundation”, 2nd ed. Prentice Hall, 1998.
[6] H. Liu and L. Yu, “Toward integrating feature selection algorithms for classification and clustering”, IEEE Trans. Knowl. Data Eng., vol. 17, no. 4, pp. 491502, Apr. 2005.
[7] V. N. Vapnik,“The Nature of Statistical Learning Theory”. New York, NY, USA: Springer-Verlag New York, Inc., 1995.
[8] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations, in Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability”, 1967, pp. 281-297.
[9] A. Mnih and G. E. Hinton,“A scalable hierarchical distributed language model”, in Proc. Adv. Neural Inform. Process. Syst., 2009, pp. 1081-1088.
[10] D. D. Lewis,“Naive (Bayes) at forty: The independence assump-tion in Inform. retrieval”, in Proc. Eur. Conf. Mach. Learn., 1998, pp. 415
[11] D. Koller and M. Sahami, “Hierarchically classifying documents using very few words”, in Proc. Int. Conf. Mach. Learn., 1997, pp. 170-178.
[12] Q. V. Le and T. Mikolov, “Distributed representations of sentences and documents”, arXiv preprint arXiv:1405.4053, 2014.
[13] Y. H. Li and A. K. Jain, “Classification of text documents”, Com-put. J., vol. 41, no. 8, pp. 537-546, 1998.
[14] G. Forman, “An extensive empirical study of feature selection metrics for text classification, J. Mach. Learn. Res., vol. 3, 1289-1305, 2003.
[15] Y. Yang and J. O. Pedersen, “A comparative study on feature selection in text categorization”, in Proc. Int. Conf. Mach. Learn., vol. 97, 1997, pp. 412-420.
[16] A. Genkin, D. D. Lewis, and D. Madigan, “Large-scale Bayesian logistic regression for text categorization”, Technimetrics, vol. 49, no. 3, pp. 291-304, 2007.
[17] B. Tang and H. He, “ENN: Extended nearest neighbour method for pattern recognition [research frontier]”, IEEE Compute. Intel. Mag., vol. 10, no. 3, pp. 5260, 2015.
[18] A. McCallum and K. Nigam, “A comparison of event models for naive Bayes text classification”, in Proc. AAAI Workshop Learning Text Categorization, 1998, vol. 752, pp. 4148.
[19] S. Eyheramendy, D. D. Lewis, and D. Madigan, “On the naive Bayes model for text categorization”, in Proc. 9th Int. Workshop Artif. Intel. Statist., 2003, pp. 332-339.
[20] V. Metsis, I. Androutsopoulos, and G. Paliouras, “Spam filtering with naive Bayes-which naive Bayes?”, in Proc. Conf. Email Anti-Spam, 2006, pp. 2728.
[21] E. F. Combarro, E. Montanes, I. Diaz, J. Ranilla, and R. Mones, “Introducing a family of linear measures for feature selection intext categorization”, IEEE Trans. Knowl. Data Eng., vol. 17, no. 9, pp
[22] S. Kay, Q. Ding, B. Tang, and H. He, “Probability density function estimation using the EEF with application to subset/feature selection”, IEEE Trans. Signal Process., vol. 64, no. 3, pp. 641651, Feb. 2016.
[23] B. Tang, H. He, P. M. Baggenstoss, and S. Kay,“A Bayesian classification approach using class-specific features for text categorization”, IEEE Trans. Knowl. Data Eng., vol. 28, no. 6, pp. 1602 1606, Jun. 2016.
[24] D. Mladenic and M. Grobelnik, “Feature selection for unbalanced class distribution and naive Bayes”, in Proc. Int. Conf. Mach. Learn., 1999, vol. 99, pp. 258-267.

[25] G. Salton, A. Wong, and C.-S. Yang, “A vector space model for automatic indexing”, Common ACM, vol. 18, no. 11, pp. 613620, 1975.

[26] J. Lin, Divergence measures based on the Shannon entropy”, IEEE Trans. Inf. Theory, vol. 37, no. 1, pp. 145151, Jan. 1991.

[27] M. F. Caropreso, S. Matwin, and F. Sebastiani, “A learner-independent evaluation of the usefulness of statistical phrases for automated text categorization, Text Databases and Document Management: Theory and Practice”. Hershey, PA, USA: IGI Publish-ing, pp. 78102, 2001.

[28] K. Sawyer, “A multiple divergence criterion for testing between separate hypotheses, Statist. Probability Lett”, vol. 1, no. 1, pp. 26 30, 1982.

[29] R. Rifkin and A. Klautau, “In defence of one-vs-all classification”, J. Mach. Learn. Res., vol. 5, pp. 101141, 2004.

[30] Y. Yang and J. O. Pedersen, A comparative study on feature selection in text categorization, in ICML, vol. 97, 1997, pp. 412420.

[31] Bo Tang, Student Member, IEEE, Steven Kay, Fellow, IEEE, and Haibo He, “Toward Optimal Feature Selection in Naive Bayes for Text Categorization”,IEEETrans.Knowl. Data Eng., vol. 28, no. 9, Sep. 2006.

[32] Madhuri B.Bhusare,Chandrakant R. Barde, “ Weighting Individual Optimal Feature Selection In Nave Bayes For Text Classification”, International Journal of Science Technology Management and Research (IJSTMER) volume 1 ,Issue 8, November 2016.

[33] Madhuri B.Bhusare,Chandrakant R. Barde, “ Weighting Individual Optimal Feature Selection In Nave Bayes For Text Classification”, International Journal of Advance Research in Science and Engineering (IJARSE) volume 6,Issue 10,February 2017.

Ms. Madhuri B. Bhusare is pursuing the Master Degree in Computer Engineering from GESs R. H. Sapat College of Engineering Management Studies Research,Nasik under Pune University. She has pursued her Bachelors Degree in Computer Technology from Matoshri college of engineering, Aklahare, Nasik under Pune University in 2014.

Prof. Chandrakant R. Barde is currently working as an Assistant Professor in the Computer Engineering Department of GESs R. H. Sapat College of Engineering Management Studies and Research, Nashik (INDIA).He received his Masters Degree in Computer Engineering from Dr.B.A.T.U., Lonere, Raigad in 2012 and Bachelors Degree in Computer Science and Engineering from Amruthvahini College of Engineering in 2005. He has an academic experience of 11 years (since 2005).