International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Index Copernicus Value (2015): 62.86 | Impact Factor (2015): 3.791

Document Classification Using Machine Learning Algorithms - A Review

P.V. Arivoli[1], T. Chakravarthy[2]

[1]Research Scholar, Department of Computer Science, A.V.V.M. S.P College, Poondi, TamilNadu, India

[2] Associate Professor, Department of Computer Science, A.V.V.M. S.P College College, Poondi, TamilNadu, India

Abstract: ***Automatic classification of text document plays a vital area of research in the field of Text Mining (TM) ever since the explosion of online text information. The sources like digital libraries, emails, blogs, etc., make the rapid evolving growth of text documents in the digital era. In general the categorization of text document includes several field of interest like Information Retrieval (IR), Machine Learning (ML) and Natural Language Processing (NLP). Hence, this paper focuses on various research attempts which employ both supervised and unsupervised ML techniques for classifying a set of text documents into pre-defined category labels.***

Keywords: Text Mining, Machine Learning, Natural Language Processing, Information Retrieval

1. Introduction

Text mining is the part of data mining which is used to discover the previously unknown as well as the interesting information from a huge amount of textual data [23]. It engages several fields like Information Retrieval (IR), Machine Learning (ML), Natural Language Processing (NLP) and Statistics [5]. Document classification is one among the emerging research area in TM. It is a well proven approach to organize the huge volume of textual data. It also widely used in knowledge extraction and knowledge representation from the text data sets. The common classification applications are email categorization, spam filtering, directory maintenance, mail routing, news monitoring and narrow casting, etc. The solutions to the most of the applications are solved by using machine learning algorithms. Ethem Alpaydin describes Machine Learning as: programming computers to optimise a performance criterion using example data or past experience [13]. In most cases, ML algorithms are applied in a situation where a programmer cannot explicitly tell the computer program what to do and what steps to take [25]. Generally, ML algorithms are classified namely by supervised, unsupervised and semi supervised. Firstly, in supervised learning technique the network is provided with a correct answer for every input pattern. Weights are determined to allow the network to produce answers as close as possible to the known correct answers. The examples of supervised learning techniques are k-Nearest Neighbor (k-NN), Maximum entropy Classifier, Support Vector Machine (SVM), Bayesian, etc. Secondly, unsupervised learning does not require a correct answer associated with each input pattern in the training data set. It explores the underlying structure in the data and organizes the patterns into specific category. Kohonen networks, Hopfield networks and self-organizing networks are some of the unsupervised learning techniques. Finally, Semi supervised learning is a combination of labeled and unlabeled data. Self training, co training (CO), Expectation Maximization (EM) and CO-EM are some of the semi supervised learning techniques.

In this paper, several research attempts in connection to the task of automatic text classification using machine learning algorithms have been studied. It mainly focuses on techniques like Maximum Entropy Classifier, Rocchio's Algorithm, Decision Trees, Vector Space Model (VSM), k-nearest neighbor (k-NN), Naive Bayes (NB), Artificial Neural Networks (ANN), Support Vector Machines (SVM), , Latent Semantic Analysis (LSA), Latent Semantic Index (LSI), Latent Dirichlet Allocation (LDA), Word2Vec, Doc2Vec, Deep Learning, Restricted Boltzmann Machine (RBM), Deep Boltzmann Machine (DBM), Deep Belief Network (DBN), Hybrid Deep Belief Network (HDBN) and Convolutional Neural Networks (CNN), etc.

The rest of the paper is organized as follows. Section 2 narrates an overview of text document representation which is necessary for automatic document classification. Section 3 specifies various machine learning algorithms used in the context of classification process and finally section 4 concludes with the summary of text classification in their concern application domain.

2. Document Representation

Figure 1 shows the overall process logic used in automatic document classification. It includes three major modules like feature extraction, feature selection and document classification. Among the three modules, document representation is the foremost segment which is characterized by feature extraction and feature selection. Feature extraction specifies various preprocessing activities which is used to reduce the document complexity and make easy the classification process. Normally, it includes process like stop words removal, stemming and tokenization. Meanwhile, feature selection handles the series of tokens from the feature extraction module to form a term-document vector matrix using the knowledge of Term Frequency (TF), Document Frequency and Inverse Document Frequency (IDF).

Firstly, read all the given documents and remove the stop words using the well defined procedure with respect to all documents. This preprocessing activity removes the words having low discrimination value. After the removal of the stop words, all the words in the documents are converted to lower case form to maintain the uniformity among the words in the entire document set. Next stemming process is initiated to find the root word for the commoner morphological ending words in English language using the most popular algorithm called Porter Stemmer [16]. These are the activities comprises in feature extraction module. Feature selection module emphasis on term frequency, inverse document frequency and term-document matrix. In particular term frequency specifies the number of times the word occurs in the respective document. Inverse document frequency specifies the relevance of the word respective to the entire document set. Finally the documents are normalized to unit length and represented by term-document matrix otherwise called as vector space model of the entire document set.

3. Machine Learning Algorithms

3.1 Maximum Entropy Classifier

Maximum entropy is a technique which works on the basis of estimating probability distributions from the given data. In the context of document classification, maximum entropy estimates the conditional distribution of the pre-defined label of a given document. Features are characterized for each document. Then the labeled training data is used to measure the expected value of these features [10]. Unlike Naive Bayes, Maximum Entropy Classifier makes no assumptions about the relationships between features and so might possibility do better when limited independence assumptions are not met.

3.2 Rocchio's Algorithm

A Rocchio's algorithm is normally used for document routing or filtering in Information Retrieval. It is a classical method for document classification which uses vector space model [36]. Rocchio's algorithm basically constructs a prototype vector for every label using a training set of documents by means of the similarity measures. A prototype vector is an average vector includes to class Ci, as in Eq.(1)

$$C_i = \alpha * Centroid_{C_i} - \beta * Centroid_{\overline{C_i}} \quad (1)$$

The algorithm is based on the assumption that most of the users have a general conception of which documents should be denoted as relevant or irrelevant. This algorithm is good in computability however it lacks in classification accuracy.

3.3 Decision Trees

Decision trees are hierarchical structure with the acyclic directed graphs; the root is starting node from the highest node these nodes are directly connected to the lower level nodes. End nodes (leafs) represent document categories, the tree leafs contain tests the classified documents must take in order to travel all nodes [22], [26], [37]. Branches connect nodes of neighboring levels, and then the testing process is performed on the selected attributes (features) of the document. Branches are related to the results of the test, leading to particular nodes of the lower level. Decision trees can be represented as influence diagrams, focusing on relationships between particular nodes. Their recursive construction uses a set of training examples and aims in separating examples belonging to separate categories.

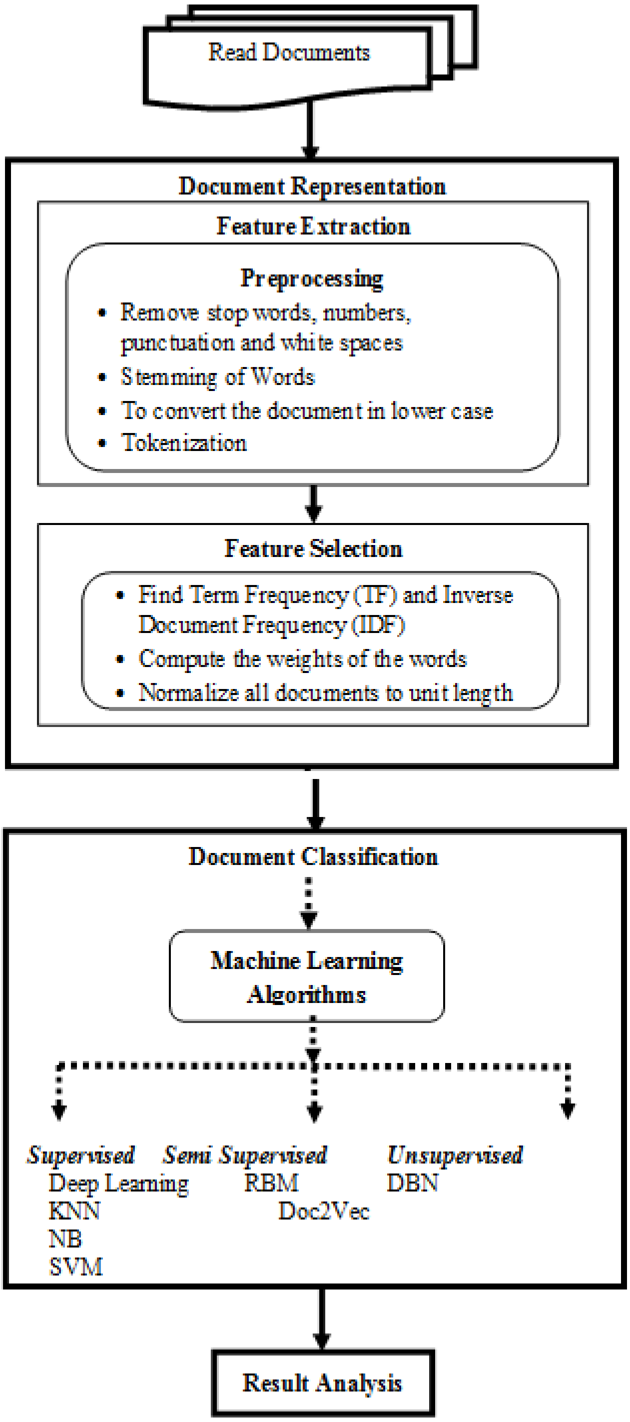

Figure 1: Text Document Classification Process

3.4 Vector Space Model (VSM)

Vector Space Model is basically a Boolean Based approach [15], [28]. It is mainly used for document retrieval tasks. Vector space model is used to convert the documents are vectors. Vector is an independent of other dimensions in each dimension, mutually independent of the vector space are assumed, when facing words which are related semantically not viable, like synonymy and polysomy.

3.5 k-Nearest Neighbor (k-NN)

The k-NN is supervised learning based algorithm and also a non parametric regression algorithm for text categorization [7], [11], [14], [42]. It is a first typical approach, classifies new cases based on a similarity measure, i.e. distance functions. By using some similarity measure such as Euclidean distance measure, etc., the distance is calculated which using the Euclidean formula, as in Eq. (2)

$$\text{Dist}(x,y) = \sqrt{\sum} (xi\text{-}yi)^2 \quad (2)$$

3.6 Naive Bayes (NB)

The Naive Bayes (NB) classifier is a supervised learning based probabilistic classifier and its derived from Bayes theorem [4], [20], [21], [34], [38], [39]. It is a classical approach and performs only on numeric and textual data. NB focuses on text classification process and many application areas like email spam detection, personal email sorting, document categorization, language detection and sentiment detection.

3.7 Artificial Neural Networks (ANN)

Artificial Neural Network includes the various approaches and has been used in document classification tasks. Some researchers are uses the single layer perceptron, it comprise only an input and an output layer [19]. Inputs are directly connects to the output layer. It is a simplest kind of feed forward network. The multilayer perceptron contains one input and output layer, but one or more hidden layers in between input and output layers [33]. The multilayer perceptron is a more sophisticated and widely used in classification tasks. In text categorization models used in back propagation and modified back propagation networks, because this model improves the performance accuracy and dimensionality reduction [8]. In [9], hybrid method combines the learning Phase Evolution Back Propagation Network (LPEBP) and Singular Value Decomposition (SVD). The improved version of the traditional Back Propagation Neural Network (BPNN) is LPEBP and it is greatly faster than BPNN. The Singular Value Decomposition increases the high dimensionality reduction and performance.

3.8 Support Vector Machine (SVM)

The Support Vector Machine is a method which works on the basis of statistical based method and also a supervised learning technique of ML. It is mainly used to solve the problems of regression and categorization [17], [40], [41], [43], [44], [48]. The SVM approach, using a sigmoid kernel function is alike to two-layer perceptron. It is used to discriminate positive and negative members of a given class of n-dimensional vectors, the training set supports positive and negative sets.. Computational learning theory that produces the structural risk minimation principle.

3.9 Latent Semantic Analysis (LSA)

The LSA method represents the contextual meaning of words by using the statistical computation performs on a corpus of documents, the basic idea of this method is complete information about the word context [2], [12], [32]. To fulfill the LSA's reflection of human erudition has been constituted in various paths, its create a LSA model, first building a m x n matrix A, where n is the number of documents in the corpus and m is the total number of terms that seem in all documents. Each column A represents the document d and each row represents term t. There are various methods on how to complete the elements at d of matrix A representing term frequencies. In this paper presents only one model of TF-IDF, and to get better result. The matrix A with LSA to apply SVD for the purpose of decomposition from latent semantic structure of the document inputs, then get n linearly independent vectors from a decomposition of documents, which represent the leading topics of the documents.

3.10 Latent Semantic Index (LSI)

LSI is a method of document indexing and retrieval method used to given set of documents, sometimes it also called LSA [30]. The basic ideas of the method, on the principle that words are used in the same context tend to have the same meaning. LSI uses a one of the mathematical technique called SVD to introduce patterns in the relation between the terms and concepts incorporated in the collection of text in unstructured format. The important key feature of LSI is to extract the meaning by establishing groups between terms of text.

3.11 Latent Dirichlet Allocation (LDA)

LDA is a elementary aspect of the model as fragment the collection of documents, probability distribution to represent the documents as a collection of topics and gives the probability distributions of every topics [32]. The LDA process is three major steps. Step1: The number of words used in a document is justified with sampling and Poisson distribution. Step2: Dirichlet distribution brings out to the distribution over the topics of a document. Step3: Topics are generated, then words for each topic generated based on the many measures for, like KL-divergence, Hellinger distance or Wassertein metric to name just a few. In [32], to choose the Hellinger distance along with the cosine similarity, and also possible statistical measure for calculate the similarity of two documents.

3.12 Word2Vec

Word2Vec does calculate the vector representation of words using the continuous bag-of-words and skin-gram architectures, along with context [32]. The skip-gram representation disseminated by Mikolov. Because generalized contexts generated and also more accuracy in other models. Word2Vec produces the output is word vocabulary, which emerge in the original document, additionally n-dimensional vector space represents their vector and gives the vector representation only for words. To get whole document, combine with some way and it creates a document vector, which can be compares cosine similarity by the another. Merge the word vector for phrases with n words, then calculate the similarity between the phrase-vector from both input document. In [32], to implements the Word2Vec method in Python.

3.13 Doc2Vec

Word2Vec project builds a lot of significance in the text mining society. Doc2Vec modified version from the Word2Vec model in an unsupervised learning passion of continuous representation of large block of text, like sentences, paragraph or whole documents [32].

3.14 Deep Learning (DL)

Deep learning permits the computational models that are combination of multilayer processing to learn representations of data with multiple levels of abstraction [29], [45], [47]. DL discovers complex structure in huge data sets by using the back propagation algorithm to indicate how a machine should change its internal parameters that are used to calculate the representation in every layer of the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shown light on sequential data such as text and speech.

3.15 Restricted Boltzmann Machine (RBM)

Basically Boltzmann Machines are stochastic recurrent neural network and log-linear Markvov random field [18]. Restricted Boltzmann machine (RBM) to creates an undirected generative models with the use of hidden layer variables through visible variables, then produce distributed model. Hidden layers are always binary and no intra layer connections within visible layer. Each RBM layer can acquisition of high correlations of hidden features between itself and the layer below.

3.16 Deep Boltzmann Machine (DBM)

Deep Boltzmann Machine is a network of symmetrically combination of stochastic binary units, and it is also composed of RBM [46]. It incorporates a set of visible and hidden units. In DBM model, all connections between layers are undirected. Many features are includes in DBM, it contains a catch layers presentation of the input with an efficient connected procedures; and also can trained unlabeled data. Parameters of all layers can be increased in combining function. DBM has a disadvantage that the training time high in respect of the machine's size and the more number of connection layers. In DBM model is not easy to make huge learning.

3.17 Deep Belief Network (DBN)

The DBN model is based on an unsupervised learning technique introduced by Hinton and Salakhutdinov [6], [27], [46]. The DBN can be viewed as a comprises of stacked Restricted Boltzmann Machines, that includes visible and hidden units. The document data reflects the visible unit and features learn reflects the hidden units from the visible units.

3.18 Hybrid Deep Belief Network (HDBN)

Hybrid Deep Belief Network (HDBN) combines the Deep Boltzmann Machine (DBM) and Deep Belief Network (DBN). The HDBN architecture divides the two layers, they are lower and upper layers [46]. The lower layer uses the Deep Boltzmann Machine and Deep Belief Network used in the upper layer. The important advantage of the HDBN is both used in directed and undirected graph models and it produces the best result and extracts the semantic information.

3.19 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are one type of neural networks that shares neuron weights on the same layer [24]. CNN to creates a major impact of computer vision and image understanding [1]. In [49], reports the experiments with convolutional neural networks trained on pre-trained words for sentence level classification tasks and it produces the excellent result performance in various benchmarks. In Kim [49], uses and compares the various types of CNN including random, static, non static and multichannel based word vectors are trained. The sentence as the concatenation of words and apply filters to each possible windows of words to produce a feature map. In [35], the CNN architecture to build four types of layers, they are convolutional layer (CONV), ReLu layer (RELU), pooling layer (POOL) and finally fully connected layer (FC). CONV receives the inputs from data set, and then computes the output of neurons that are connected to local regions in the input, each computing a dot product between their weights and regions to connect an input volume. Activation function executes on element wise in RELU layer. A down sampling operation with the spatial dimensions performs on POOL layer. Finally get the resulting in volume size and class size in FC layer.

Table 1: Summary of Machine Learning Algorithms for Document Classification

Sl. No	*Research Group*	*Year*	*Learning Algorithm Used*	*Training Environment*	*Data Set Used*
1	Sang-Bum Kim, Kyoung -Soo Han, Hae-Chang Rim, and Sung Hyon Myaeng	2006	Naive Bayes (NB)	Text Documents	Standard Reuters 21578 and 20 Newsgroup
2	Hugo Larochelle and Yoshua Bengio	2008	Restricted Boltzmann Machine (RBM)	Character Recognition and Text classification	MNIST dataset (Image)
3	Dennis Ramdass and Shreyes Seshasai	2009	Maximum Entropy Classification (MEC)	Text Classification	MIT Newspaper
4	Cheng Hua Li and Soon Choel Park	2009	Artificial Neural Networks (ANN) / Learning Phase Evolution Back Propagation Network (LPEBPN) and Singular	Document Classification	Standard Reuters 21578 and 20 Newsgroup

			Value Decomposition (SVD)		
5	Lawrence McAfee	2009	Deep Belief Network (DBN)	Document Classification	Wikipedia XML Corpus
6	Antonio Jimeno Yepes, Andrew MacKinlay, Justin Bedo, Rahil Garvani and Qiang Chen	2014		Bio Medical Domain	MTI ML site
7	Saurav Sahay	2011	Support Vector Machine (SVM)	Corporate Acquisition Documents	Standard Reuters 21578
8	Francis M. Kwale	2013	Vector Space Model (VSM)	Customer Relationship Management	Own example
9	Magnus La Fleur and Fredrik Renström	2015	Latent Semantic Index (LSI)	Customer Complaint Documents	Jeeves Support System Database
10	Michal Campr and Karel Jeˇ zek	2015	Latent Dirichlet Allocation (LDA), Word2Vec, Doc2Vec	Restaurant Reviews	www.fajnsmekr.cz
11	Yan Yan,Xu-Cheng Yin,Sujian Li,Mingyuan Yang, and Hong-Wei Hao	2015	Deep Boltzmann Machine (DBM)	News Articles	20 Newsgroup and BBC News Data
12	Yan Yan, Xu-Cheng Yin, Sujian Li, Mingyuan Yang, and Hong-Wei Hao	2015	Hybrid Deep Belief Network (HDBN)	News Article	20 News group data and BBC News Data
13	Andreas Holzinger, Johannes Schant, MiriamSchroettner, Christin Seifert, and Karin Verspoor	2014	Latent Semantic Analysis (LSA)	Medical Domain	GENIA and CRAFT
14	Edgar Altszyler, Mariano Sigman and Diego Fernández Slezak	2016		Document Classification	Dream Bank Report Corpus
15	Yann LeCun	2015	Deep Learning	Text Images Classification	Nature Database
16	Yan Yan, Xu-Cheng Yin, Bo-Wen Zhang, Chun Yang and Hong-Wei Hao	2016		Bio Medical Domain	Bio ASQ Data set and MEDLINE database
17	Lin-peng Jin1 and Jun Dong	2016		Bio Medical Data Classification	ECG Data set
18	L. Kang, J. Kumar, P. Ye, Y. Li, D. Doerman	2014	Convolutional Neural Networks (CNN)	Document Image Classification	Tobacco litigation data set and NIST tax-form data set
19	Yoon Kim	2014		Sentiment Analysis and Question Classification	Stanford Sentiment Tree bank, TREC data set ,Customer and Movie reviews
20	Andrej Karpathy, George Toderici and Sanketh Shetty	2014		Sports Video Classification	Sports IM data set
21	Ranti D. Sharma, Samarth Tripathi, Sunil K. Sahu, Sudhanshu Mittal, and Ashish Anand,	2016		Medical user review	Rate MDs's data set

4. Conclusion

Table 1 summarizes the various ML algorithms both supervised and unsupervised techniques used for document classification. In particular, Maximum Entropy Classifier, Rocchio's Algorithm, Decision Trees, Vector Space Model (VSM), K-nearest neighbor (KNN), Naive Bayes (NB), Artificial Neural Networks (ANN), Support Vector Machines (SVMs), , Latent Semantic Analysis (LSA), Latent Semantic Index (LSI), Latent Dirichlet Allocation (LDA), Word2Vec, Doc2Vec, Deep Learning, Restricted Boltzmann Machine (RBM), Deep Boltzmann Machine (DBM), Deep Belief Network (DBN), Hybrid Deep Belief Network (HDBN) and Convolutional Neural Networks (CNN) plays a very important role in classifying the text documents in concern to the respective application domain. Among the various research attempts, currently CNN model results in higher accuracy for document classification.

References

[1] Andrea Vedaldi Karel Lenc, "MatConvNet: Convolutional Neural Networks for MATLAB", MM '15 Proceedings of the 23rd ACM international conference on Multimedia, Pages 689-692,2015, doi.10.1145/2733373.28074122016
[2] Andreas Holzinger, Johannes Schant, MiriamSchroettner, Christin Seifert, and Karin Verspoor, "Biomedical Text Mining: State -of-the-Art,Open Problems and Future Challenges", Knowledge Discovery and Data Mining, LNCS 8401, Springer-Verlag Berlin, pp. 271–300, 2014.
[3] Andrej Karpathy, George Toderici and Sanketh Shetty, "Large-scale Video Classification with Convolutional Neural Networks", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1725-1732, 2014.

[4] Andrew McCallum and Kamal Nigam, “A Comparison of Event Models for Naïve Bayes Text Classification”, Journal of Machine Learning Research 3, pp. 1265-1287. 2003.
[5] Anne Kao and Steve Poteet, “Text Mining and Natural Language Processing: Introduction for the Special Issue”, ACM SIGKDD Explorations Newsletter - Natural language processing and text mining. Vol. 7, Issue 1, 2005. DOI: 10.1145/1089815.1089816
[6] Antonio Jimeno Yepes, Andrew MacKinlay, Justin Bedo, Rahil Garvani and Qiang Chen, "Deep Belief Networks and Biomedical Text Categorisation", In Proceedings of Australasian Language Technology Association Workshop, pages 123−12,2014.
[7] Bang, S. L., Yang, J. D., and Yang, H. J. , “ Hierarchical document categorization with k-NN and concept-based thesauri, Elsevier, Information Processing and Management”, pp. 397–406, 2006.
[8] Bo Yu, Zong-ben Xu and Cheng-hua Li ,“Latent semantic analysis for text categorization using neural network”, E lsevier, Knowledge-Based Systems Vol. 21, Issue. 8, pp. 900–904, 2008.
[9] Cheng Hua Li and Soon Choel Park, "An efficient document classification model using an improved back propagation neural network and singular value decomposition", Elsevier, Expert Systems with Applications, Vol. 36 ,pp- 3208–3215, 2009.
[10] Dennis Ramdass and Shreyes Seshasai, "Document Classification for Newspaper Articles", 6.863 Final Project, Spring 2009.
[11] Duoqian Miao , Qiguo Duan, Hongyun Zhang and Na Jiao, “Rough set based hybrid algorithm for text classification”, Elsevier, Expert Systems with Applications, Vol, 36, Issue 5, Pages 9168–9174, July 2009 .
[12] Edgar Altszyler, Mariano Sigman and Diego Fernández Slezak, "Comparative study of LSA vs Word2vec embeddings in small corpora: a case study in dreams database", Oct 2016. URL: https://arxiv.org/pdf/1610.01520v1.pdf
[13] Ethem Alpaydin, “Introduction to Machine Learning (Adaptive Computation and Machine Learning)”, The MIT Press, 2004.
[14] Eui-Hong (Sam) Han, George Karypis and Vipin Kumar, “Text Cat egorization Using Weighted Adjusted k-Nearest Neighbor Classification”, Department of Computer Science and Engineering, Army HPC Research Centre, University of Minnesota, Minneapolis, USA. 1999.
[15] Francis M. Kwale, “An Efficient Text Clustering Framework”, International Journal of Computer Applications (0975 – 8887) Volume 79, No.8, October 2013.
[16] Hao Lili and Hao Lizhu., “Automatic identification of stopwords in Chinese text classification”, In proceedings of the IEEE International Conference on Computer Science and Software Engineering, pp. 718 – 722, 2008.
[17] Heide Brücher, Gerhard Knolmayer and Marc-André Mittermayer, “Document Classification Methods for Organizing Explicit Knowledge”, Research Group Information Engineering, Institute of Information Systems, University of Bern, Engehaldenstrasse 8, CH - 3012 Bern, Switzerland. 2002.
[18] Hugo Larochelle and Yoshua Bengio,“ Classification using Discriminative Restricted Boltzmann Machines”, Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 2008.
[19] Hwee-Tou Ng, Wei-Boon Goh and Kok-Leong Low , “Feature Selection, Perceptron Learning, and a Usability Case Study for Text Categorization, In Proceedings of the 20th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, pp.67-73. 1997.
[20] Irina Rish, “An Empirical Study of the Naïve Bayes Classifier”, In Pro ceedings of the IJCAI-01 Workshop on Empirical Methods in Artificial Intelligence. 2001.
[21] Irina Rish, Joseph Hellerstein and Jayram Thathachar, “An Analysis of Data Characteristics that affect Naïve Bayes Performance”, IBM T.J. Watson Research Center 30 Saw Mill River Road, Hawthorne, NY 10532, USA. 2001.
[22] Joachims, T, “Text Categorization With Support Vector Machines: Learning with Many Relevant Features”, In: European Conference on Machine Learning, Chemnitz, Germany 1998, pp.137-142 , 1998.
[23] Joachims. T, “ Transductive Inference for Text Classification Using Support Vector Machines,” Proc. 16th Int’l Conf. Machine Learning (ICML ’99), pp. 200-209, 1999.
[24] Kang. L, Kumar. J and Ye, Li. Y, Doerman. D, "Convolutional neural networks for document image classification", ICPR, pp. 3168-3172, 2014.
[25] Katariina Nyberg, “Document Classification Using Machine Learning and Ontologies” - Master’s Thesis, AALTO UNIVERSITY, Espoo, January 31, 2011.
[26] Kim. J, Lee. B, Shaw. M, Chang. H and Nelson. W, “Application of Decision -Tree Induction Techniques to Personalized Advertisements on Internet Storefronts”, International Journal of Electronic Commerce 5(3) pp.45-62, 2001.
[27] Lawrence McAfee, “Document Classification using Deep Belief Nets”, CS224n, Sprint 2008.
[28] Lee Sangno, Song Jaeki and Kim Yongjin, “An Empirical Comparison Of Four Text Mining Methods”, The Journal of Computer Information Systems,Vol. 51,No.1, 2010.
[29] Lin-peng Jin and Jun Dong, "Ensemble Deep Learning for Biomedical Time Series Classification", Hindawi Publishing Corporation Computational Intelligence and Neuroscience, Vol. 2016, Article ID 6212684,2016. URL: http://dx.doi.org/10.1155/2016/6212684
[30] Magnus La Fleur and Fredrik Renström, “Conceptual Indexing using Latent Semantic Indexing A Case Study”, Department of Infor mation Technology, UPPSALA University,2015.
[31] Max Jaderberg , Karen Simonyan , Andrea Vedaldi and Andrew Zisserman, "Reading Text in the Wild with Convolutional Neural Networks", Springer, International Journal of Computer Vision, 2014.
[32] Michal Campr and Karel Jeˇzek, “Comparing Semantic Models for Evaluating Automatic Document Summarization”, Springer Lecture Notes in Computer Science : Text, Speech, and Dialogue,Vol.9302, pp. 252-260, 2015. DOI: 10.1007/978-3-319-24033-6_29
[33] Miguel E. Ruiz and Padmini Srinivasan, “Automatic Text Categorization Using Neural Network”, In

Proceedings of the 8th ASIS SIG/CR Workshop on Classification Research, pp. 59-72. 1998.
[34] Pedro Domingos and Michael Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss, Machine Learning", Vol. 29, No. 2 -3, pp.103-130. 1997.
[35] Ranti D. Sharma, Samarth Tripathi, Sunil K. Sahu, Sudhanshu Mittal and Ashish Anand, "Predicting Online Doctor Ratings from User Reviews Using Convolutional Neural Networks", International Journal of Machine Learning and Computing, Vol. 6, No. 2, April 2016.
[36] Rocchio. J, "Relevance Feedback in Information Retrieval", In G. Salton (ed.). Computer Science and Applied Cognitive Science, The SMART System: pp.67-88, 1971.
[37] Russell Greiner and Jonathan Schaffer, "AIxploratorium – Decision Trees", Department of Computing Science, University of Alberta, Edmonton, ABT6G2H1, Canada.2001. URL :http://www.cs.ualberta.ca/ ~aixplore/ learning/ DecisionTrees
[38] Sang-Bum Kim, Hue-Chang Rim, Dong-Suk Yook and Huei-Seok Lim, "Effective Methods for Improving Naïve Bayes Text Classification", 7th Pacific Rim International Conference on Artificial Intelligence, Vol. 2417. 2002.
[39] Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim, and Sung Hyon Myaeng, "Some Effective Techniques for Naive Bayes Text Classification", IEEE Transactions On Knowledge And Data Engineering, Vol. 18, No. 11, November 2006.
[40] Saurav Sahay, "Support Vector Machines and Document Classification",URL:http://www.static.cc.gatech.edu/~ssahay/sauravsahay7001-2.pdf
[41] Soumen Chakrabarti, Shourya Roy and Mahesh V. Soundalgekar, "Fast and Accurate Text Classification via Multiple Linear Discriminant Projection" , The International Journal on Very Large Data Bases (VLDB), pp.170-185, 2003.
[42] Tam. V, Santoso. A, and Setiono. R, "A comparative study of centroid-based, neighborhood-based and statistical approaches for effective document categorization", Proceedings of the 16th International Conference on Pattern Recognition, pp.235–238, 2002.
[43] Thorsten Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features" ECML -98, 10th European Conference on Machine Learning, pp. 137-142, 1998.
[44] Vladimir N. Vapnik, "The Nature of Statistical Learning Theory" , Springer, NewYork. 1995.
[45] Yan Yan, Xu-Cheng Yin, Bo-Wen Zhang, Chun Yang and Hong-Wei Hao,"Semantic indexing with deep learning: a case study", Big Data Analytics, Vol. 1, No. 7, 2016. DOI:10.1186/s41044-016-0007-z.
[46] Yan Yan,Xu-Cheng Yin,Sujian Li,Mingyuan Yang, and Hong-Wei Hao, "Learning Document Semantic Representation with Hybrid Deep Belief Network" , Hindawi Publishing Corporation Computational Intelligence and Neuroscience, Vol. 2015, http://dx.doi.org/10.1155/2015/650527
[47] Yann LeCun, Yoshua Bengio and Geoffrey Hinton, "Deep learning - A Review", Nature, Vol. 521, No. 28, May 2015.
[48] Yi Lin, "Support V ector Machines and the Bayes Rule in Classification", Technical Report No.1014, Department of Statistics, University of Wiscousin, Madison. 1999.
[49] Yoon Kim, "Convolutional Neural Networks for Sentence Classification", 2014, URL: https://arXiv preprintarXiv:1408.5882.