

highly correlated and does not considering the word count in that case it show very poor result.

In case of k-NN if the pre-processing is proper then this method shows the good results. It can suitable for n number of documents. But difficult to find the value of k in this method it takes to long time. Jaccard coefficient shows good result for categorizing the message into two categories. These methods first identify the terms and then only check for those terms in two documents with simple formula which gives result in minimum time, efficiency is high.

6. Conclusion

Some more research is required for performance improvement and more accuracy of the classification process. Classification and clustering of semi-structured documents is challenging. By reducing the training and testing time and improves the accuracy of the classification precision and recall. By studying all the techniques learn that every technique has some drawbacks and advantages with some specified condition. Among all of the some shows the good results. While choosing the technique we have to consider the advantage as well as limitation of each technique.

References

- [1] Chauhan Shrihari R, Amish Desai, "A Review on Knowledge Discovery using Text Classification Techniques in Text Mining", International Journal of Computer Applications (0975 – 8887) Volume 111 – No 6, February 2015
- [2] Vandana Korde,"Text classification and classifiers:" International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.2, March 2012".
- [3] Zakaria Elberrichi, Karima Abidi, "Arabic text categorization: a comparative study of different representation modes." Int. Arab J. Inf. Technol. 9(5): 465-470 (2012)
- [4] S. Subbaiah "Extracting Knowledge using Probabilistic Classifier for Text Mining" Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22, IEEE-2013
- [5] M. Janaki Meena , K. R. Chandran "Naive Bayes Text Classification with Positive Features Selected by Statistical Method" ©2009 IEEE vaishali Bhujade, N.J.Janwe "knowledge discovery in text mining techniques using association rule extraction" International Conference on Computational Intelligence and Communication Systems, IEEE-2011
- [6] Jiawei Han and Micheline Kamber "Data Mining Concepts And Techniques" ,Morgan kaufman publishers, San Francisco, Elsevier, 2011, pp. 285-351
- [7] Aurangzeb Khan, Baharum Baharudin, Lam Hong Lee," A Review of Machine Learning Algorithms for Text-Documents Classification", Journal Of Advances In Information Technology, Vol. 1, No. 1, February 2010
- [8] Fabrizio Sebastiani, "Machine Learning in Automated Text Categorization", ACM Computing Surveys, Vol. 34, No. 1, March 2010.

- [9] Lena Tenenboim, Bracha Shapira, Peretz Shoval "Ontology- Based Classification Of News In An Electronic Newspaper" International Conference "Intelligent Information and Engineering Systems" INFOS 2008, Varna, Bulgaria,June-July2008.
- [10] Anna Huang, "Similarity Measures for Text Document Clustering", NZCSRSC 2008, April 2008, Christchurch, New Zealand.
- [11] Dasgupta, "Feature selection methods for text classification.",In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining,pp.230-239,2007.

Author Profile



Ankita R. Jadhao she is pursuing M.tech in Dept. of Computer Science and Engineering, Ramdeobaba College of Engineering and Management , Nagpur, India. She passed B.E from Sipna College of Engineering Technology, Amravati, India. She had actively participated in various international conference and workshop. Her field of interest is Natural Language Processing.



Avinash J. Agrawal received Bachelor of Engineering Degree in Computer Technology from Nagpur University, India and Master of Technology degree in Computer Technology from National Institute of Technology, Raipur, India in 1998 and 2005 respectively. He has Ph.D. in Computer Science and Engineering from Visvesvaraya National Institute of Technology, Nagpur in 2013. His research area is Natural Language Processing and Artificial Intelligence. He is having 18 years of teaching experience. Presently he is Associate Professor in Shri Ramdeobaba College of Engineering and Management, Nagpur. He is the author of more than 50 research papers in International Journal and Conferences.