













various problems associated with spam and different methods and techniques attempting to deal with it. From the study we identified that, many of the filtering techniques are based on text categorization methods and there is no technique can claim to provide an ideal solution with 0% false positive.

## REFERENCES

- [1] Han, J., Kamber, M., & Pei, J. (2006). Data mining: concepts and techniques. Morgan kaufmann.
- [2] Jiawei, H., & Kamber, M. (2001). Data mining: concepts and techniques. San Francisco, CA, itd: Morgan Kaufmann, Data, C. H. D. (2010). Data Mining: Concepts and Techniques.
- [3] Androutsopoulos, I., Koutsias, J., Chandrinou, K. V., & Spyropoulos, C. D. (2000, July). An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages. In Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval (pp. 160-167). ACM.
- [4] Androutsopoulos, I., Paliouras, G., Karkaletsis, V., Sakkis, C., Spyropoulos, C. D., & Stamatopoulos, P. (2000). Learning to filter spam e-mail: A comparison of a naive bayesian and a memory-based approach. arXiv preprint cs/0009009.
- [5] Basavaraju, M., & Prabhakar, R. (2010). A novel method of spam mail detection using text based clustering approach. International Journal of Computer Applications, 5(4).
- [6] Hovold, J. (2005, July). Naive bayes spam filtering using word position-based attributes. In Proceedings of the 2nd Conference on Email and Anti-Spam (CEAS 2005).
- [7] Jin, X., Wang, L., Lu, Y., & Shi, L. (2003). MC tree: A dynamic index structure for partially clustered multi-dimensional database. Tsinghua Science and Technology, 8(2), 174-180.
- [8] Liu, P. Y., Zhang, L. Y., & Zhu, Z. F. (2009). Research on e-mail filtering based on improved Bayesian. Journal of Computers, 8(3), 271-275.
- [9] Rajput, A., & Tewari, D. Adaptive Spam Filtering based on Bayesian Algorithm.
- [10] Rennie, J. (2000, August). ifile: An application of machine learning to e-mail filtering. In Proc. KDD 2000 Workshop on Text Mining, Boston, MA.
- [11] Sahami, S., Dumais, S., Heckerman, D., & Horvitz, E. (1998, July). A Bayesian approach to filtering junk e-mail. In Learning for Text Categorization: Papers from the 1998 workshop (Vol. 62, pp. 98-105).
- [12] Song, Y., Kolcz, A., & Giles, C. L. (2009). Better Naive Bayes classification for high-precision spam detection. Software: Practice and Experience, 39(11), 1003-1024.
- [13] TIAN Jinlan, ZHANG Suqin, ZHU Lin, LIU Lu. (2005). Improvement and Parallelism of k-Means Clustering Algorithm. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China, 10(3), 277-281.