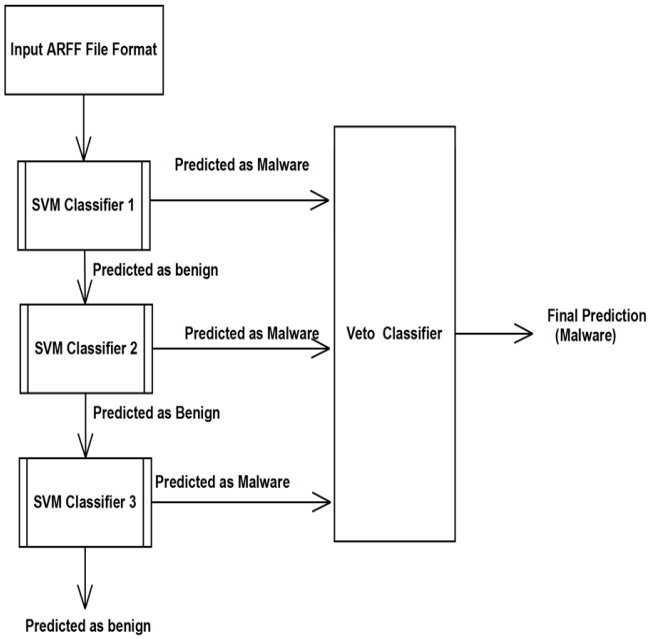
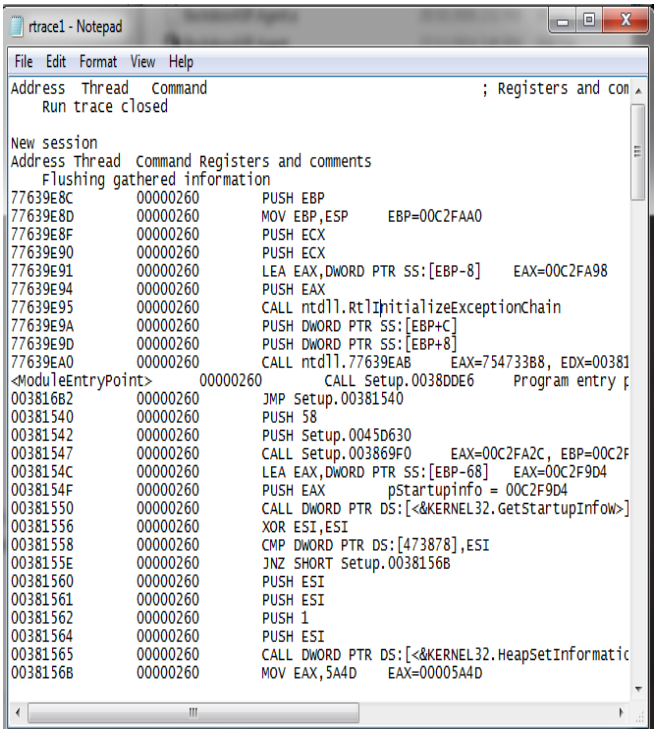


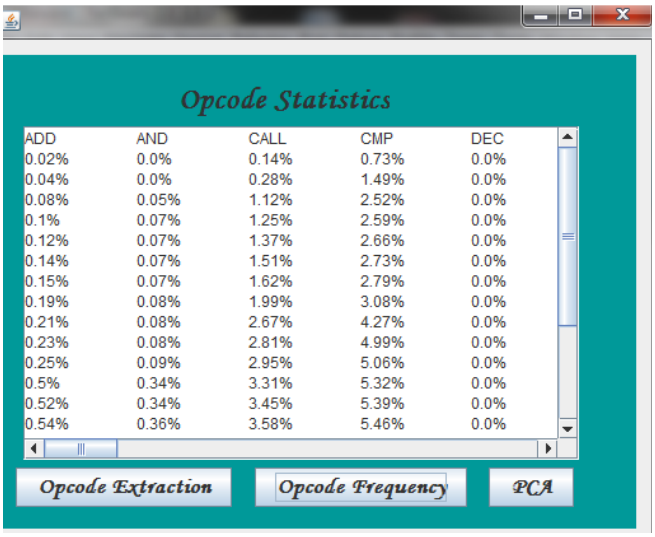
Figure 2:.Ensemble classification approach using veto Voting

4. Results

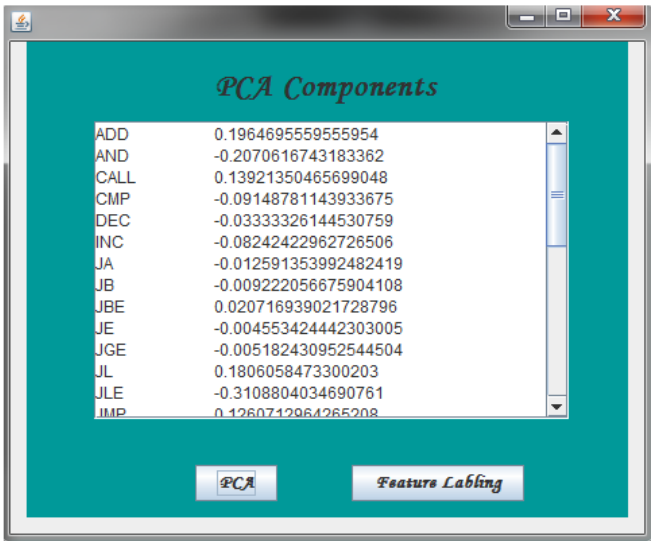
4.1 Dataset Creation using OllyDbg Disassembler



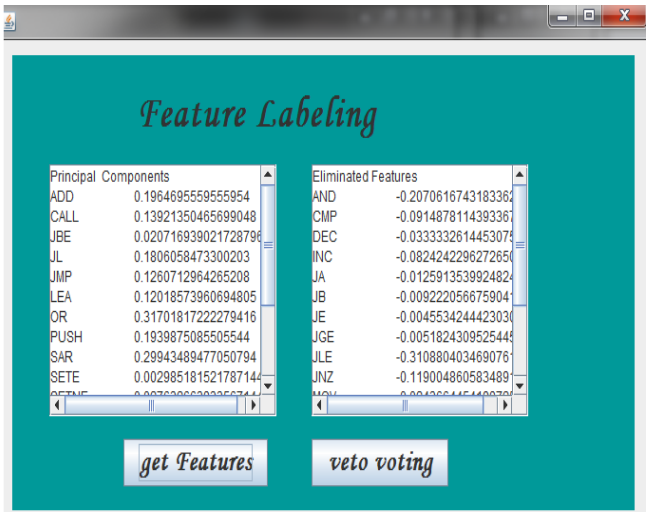
4.2 Opcode Extraction with Occurences for each *.txt file in the dataset



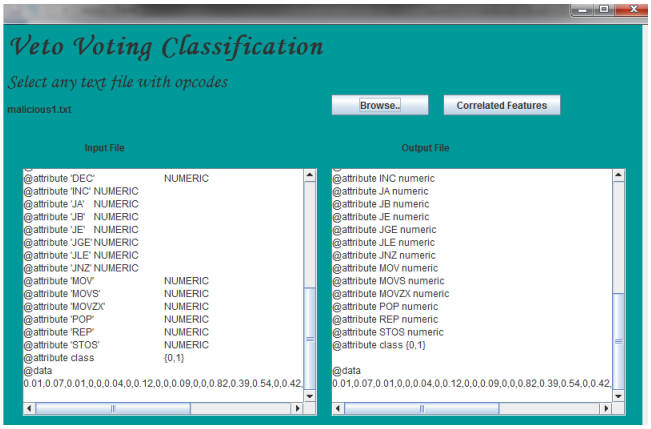
4.3 Principle Component Analysis

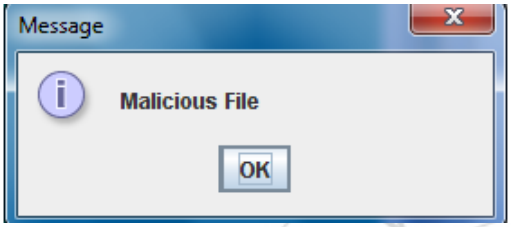


4.4 Feature Labelling By using PCA method



4.5 Veto Voting Ensemble Classification





5. Conclusions

1. The less frequent opcodes having an importance rating for classifying benign and malicious software. While mov has a negative impact on the classification and identification of software.
2. Opcode mov is a poor indicator of benign and malicious software. mov opcodes inhibits the ability to correctly classify software when used with other opcodes. Using the eigenvector prefilter, irrelevant features can safely remove.
3. Ensemble SVM classifier provides good accuracy to detectmalware as compared to other methods.

Acknowledgement

I avail this opportunity to express my deep sense of gratitude and whole hearted thanks to my guide Prof. D. M. Padulkar for giving his valuable guidance, inspiration and encouragement to embark this paper. Without his Coordination, guidance and reviewing, this task could not be completed alone.

References

[1] P. O'Kane, S. Sezer, K. McLaughlin, and E. Im, "SVM Training Phase Reduction Using Dataset Feature Filtering for Malware Detection", IEEE Transaction on information Forensic And Security, VOL. 8, NO. 3, March 2013.
[2] A. Shabtai, R. Moskovitch, C. Feher, S. Dolev, and Y. Elovici, "Detecting unknown malicious code by applying classification techniques on opcode patterns, Security Informatics, vol. 1, pp. 1-22, 2012.
[3] D. Bilar, "Opcodes as predictor for malware, Int. J. Electron. Security Digital Forensics, vol. 1, no. 2, pp. 156 - 168, 2007.
[4] D. Bilar, "Callgraph properties of executables and generative mechanisms, AI Commun., Special Issue on Network Anal. in Natural Sci.and Eng., vol. 20, no. 4, pp. 231-243, 2007.
[5] I. Santos, Y. K. Penya, J. Devesa, and P. G. Garcia, "N-grams-based file signatures for malware detection," S3Lab, Deusto Technological Found., 2009.
[6] R. Sekar, M. Bendre, D. Bollineni, and Bollineni, R. Needham and M.Abadi, Eds., "A fast automaton-based method for detecting anomalous program behaviors," in Proc. 2001 IEEE Symp. Security and Privacy,IEEE Comput. Soc., Los Alamitos, CA, USA, 2001, pp. 144-155."
[7] Wei-Jen Li, W. L. K. Wang, S. Stolfo, and B. Herzog, "Fileprints: Identifying file types by n-gram analysis, in Proc. 6th IEEE Inform. Assurance Workshop,June 2005, pp. : 64-71.
[8] I. Santos, F. Brezo, B. Sanz, C. Laorden, and Y. P. G. Bringas, "Using opcode sequences in single-class learning to detect unknown malware," IET Inform. Security, vol. 5, no. 4, pp. 220227, 2011.

Author Profile

Ms. Shital Kuber, Received the Bachelors degree(B. E.) computer Engineering in 2013 from VPCOE, Baramati. she is now pursuing M. E. degree, from VPCOE, Baramati.

Prof. Digambar Padulkar is working as Assistant professor, Department Of Computer Engineering, Vidya Pratishthan's College Of Engineering, Baramati, Pune, Maharashtra.. His research areas include Uncertain Data Mining, Applications of Data Mining in Business and Intelligent predictions.