Comparison of Intrusion Detection Systems for Use in Low-Powered Devices

Pratyay Amrit (140953430)

Under the guidance of

Ms. Ipsita Upasana,
Assistant Professor
Department of I & CT
MIT, Manipal

Contents

Introduction

- Security in low powered devices
- Rise of IoT and WSN
- Intrusion Detection System (IDS)

Problem Definition

- Why IDS for low powered devices?
- Signature based IDS
- Anomaly based IDS
- Problems with low powered devices

Objectives

- To determine useful features in classifying network data.
- To find appropriate measures to score the different models.
- To study the causes of different models performing differently
- To open scope for future studies to improve the given score.

Scope

- Cyber Security service providers
- People working in IoT/WSN
- Further research

Methodology: Dataset

- UNSW-NB15 (over 2.5 million records)
- 9 types of attacks:
  Fuzzers
  Analysis
  Backdoor
  DoS
  Exploit
  Generic
  Reconnaissance
  Shell Code
  Worm

Methodology: Dataset

- 49 features divided into 6 categories:
- Flow Features(5), Basic Features(13), Content Features(8), Time Features(10), Additional Generated Features(11), Labeled Features(2)

Methodology: Pre-processing

- Import data (python pandas)
- Fill empty values
- Transform nominal to numeric
- Add feature names to Pandas DataFrame

Methodology: Feature Selection

- Use ExtraTreesClassifier
- Feature Importance Scores (gini impurity)

$$I_g(p) = 1 - \sum_{i=1}^{J} \left(p_i^2\right)$$

Methodology: K-Nearest Neighbors

- Popular classification algorithm.
- An object is classified as the class of the majority vote from the nearest k vectors in the feature space.

Methodology: Naive Bayes

- Popular classification algorithm.
- Assumption that features have no correlation.

$$p(C_k|X) = p(C_k)\prod_{i=1}^{n}(p(x_i|C_k))$$

Methodology: Decision Trees

- Generates tree-like model of decisions and their consequences.
- Information gain, gini impurity etc. are used to decide which feature to split at every level in the tree.

Methodology: Random Forests

- Ensemble Method
- Many small trees formed from random samples of the dataset.
- Decision on the basis of majority vote among individual trees.

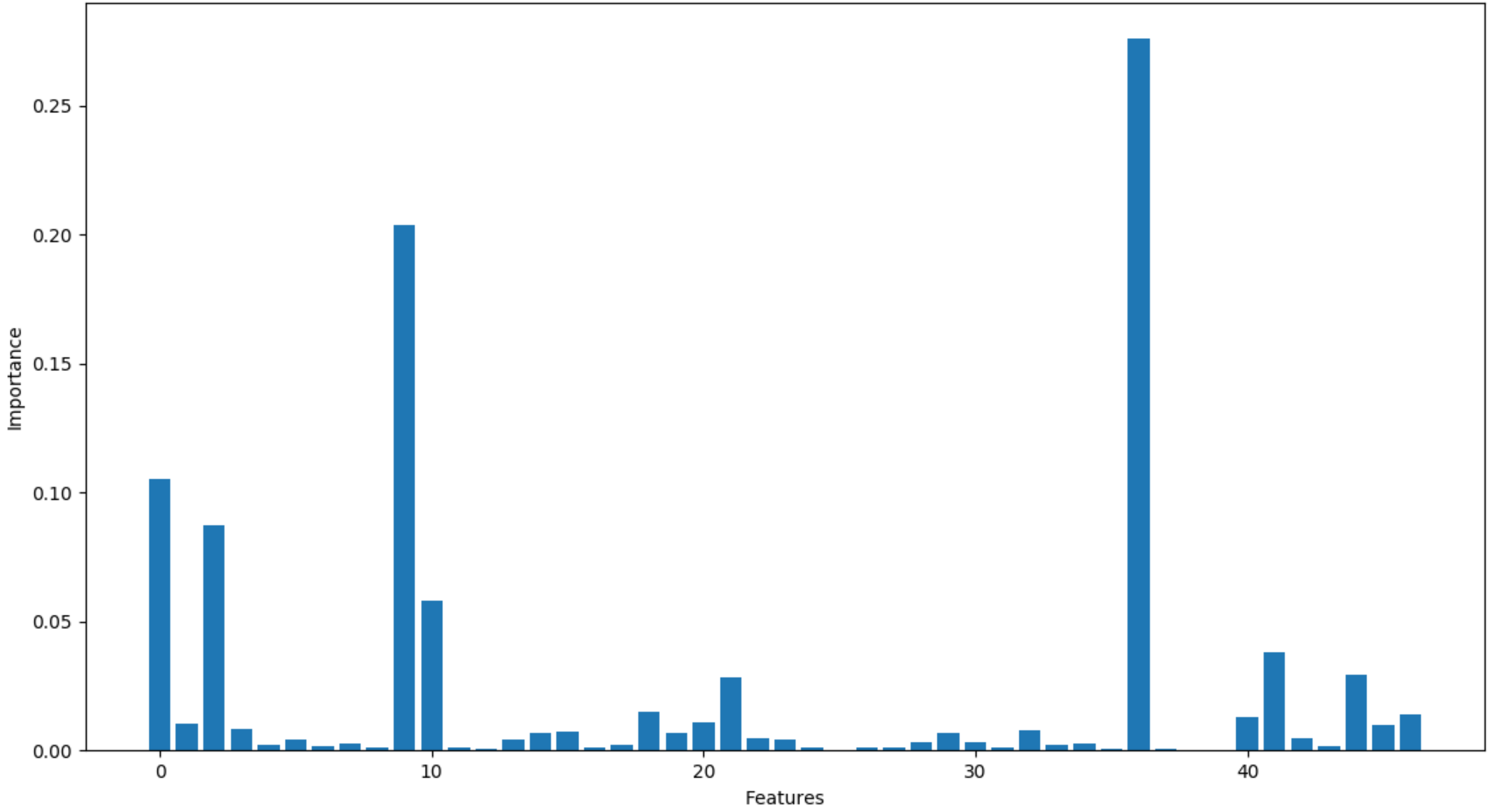
Methodology: Extra Trees

- Ensemble Method
- Similar to Random Forests.
- Split is decided using the entire dataset, instead of a random subset.

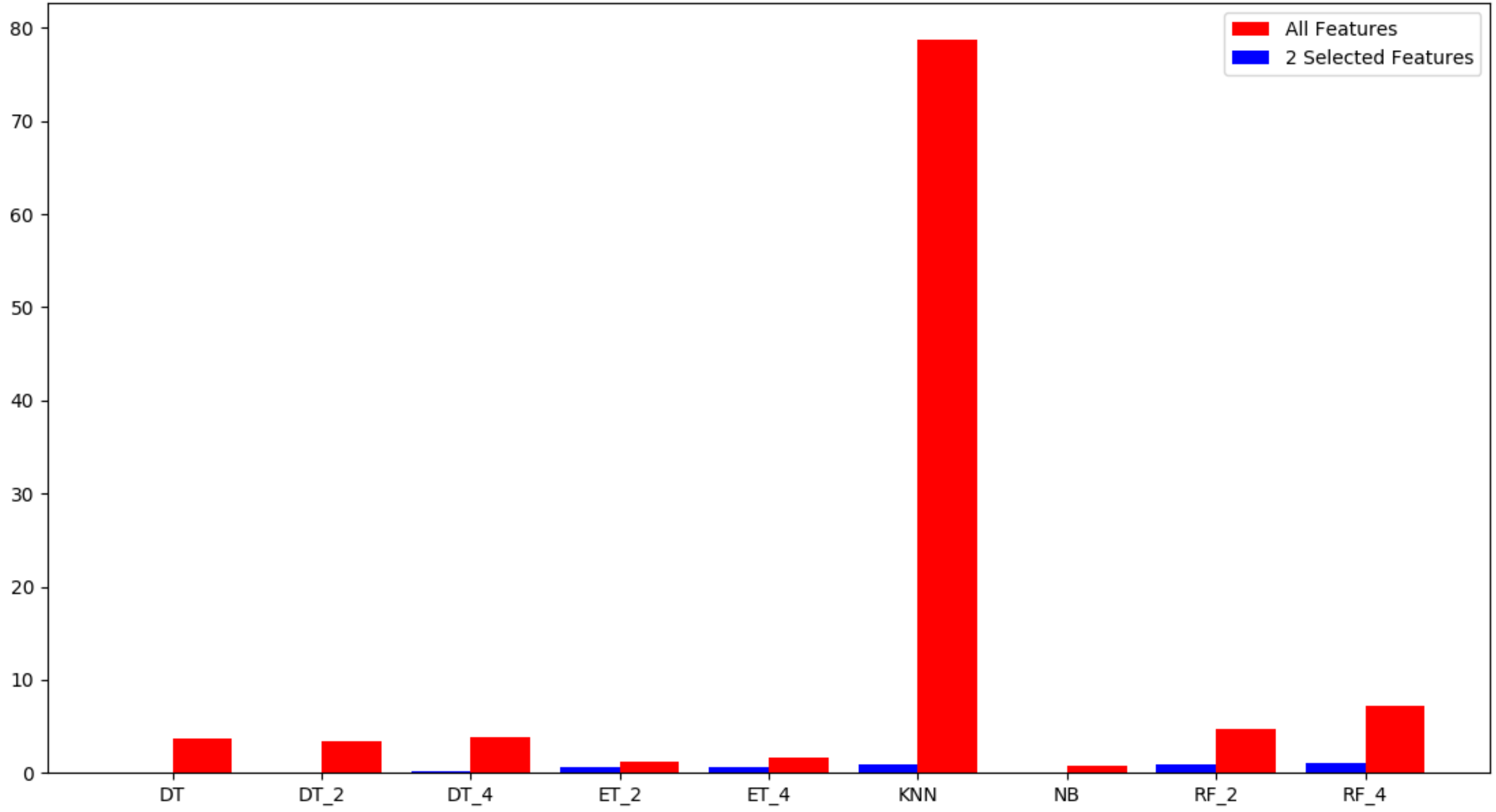
Methodology: Performance Parameters

- True Positive (TP): correctly classified attacks.
- True Negative (TN): correctly classified normal.
- False Positive (FP): incorrectly classified attacks.
- False Negative (FN): incorrectly classified normal.
- $accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
- $TS = \frac{accuracy}{ttf}$ $\quad$ $PS = \frac{accuracy}{ttp}$

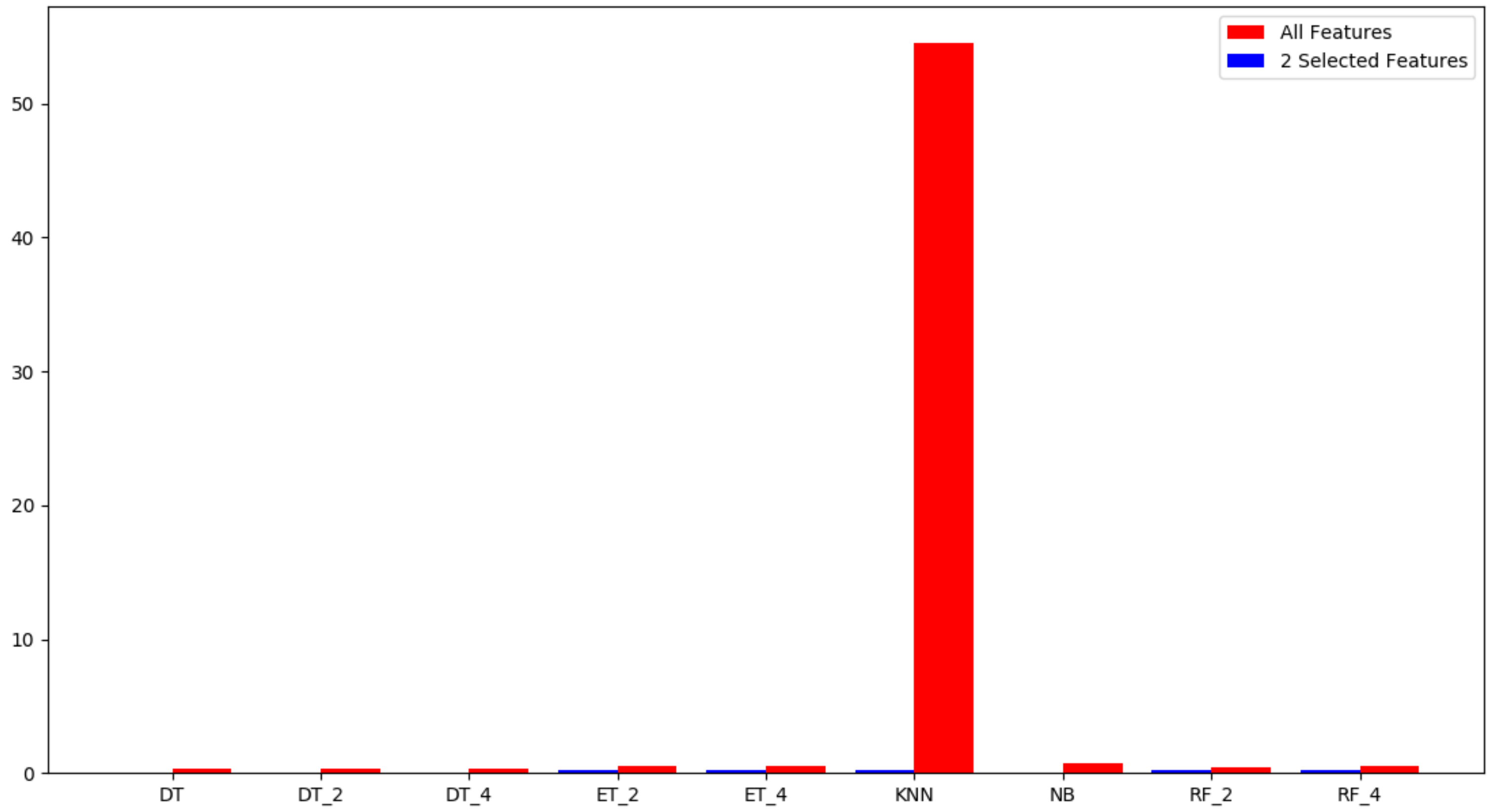
Results: Feature Selection



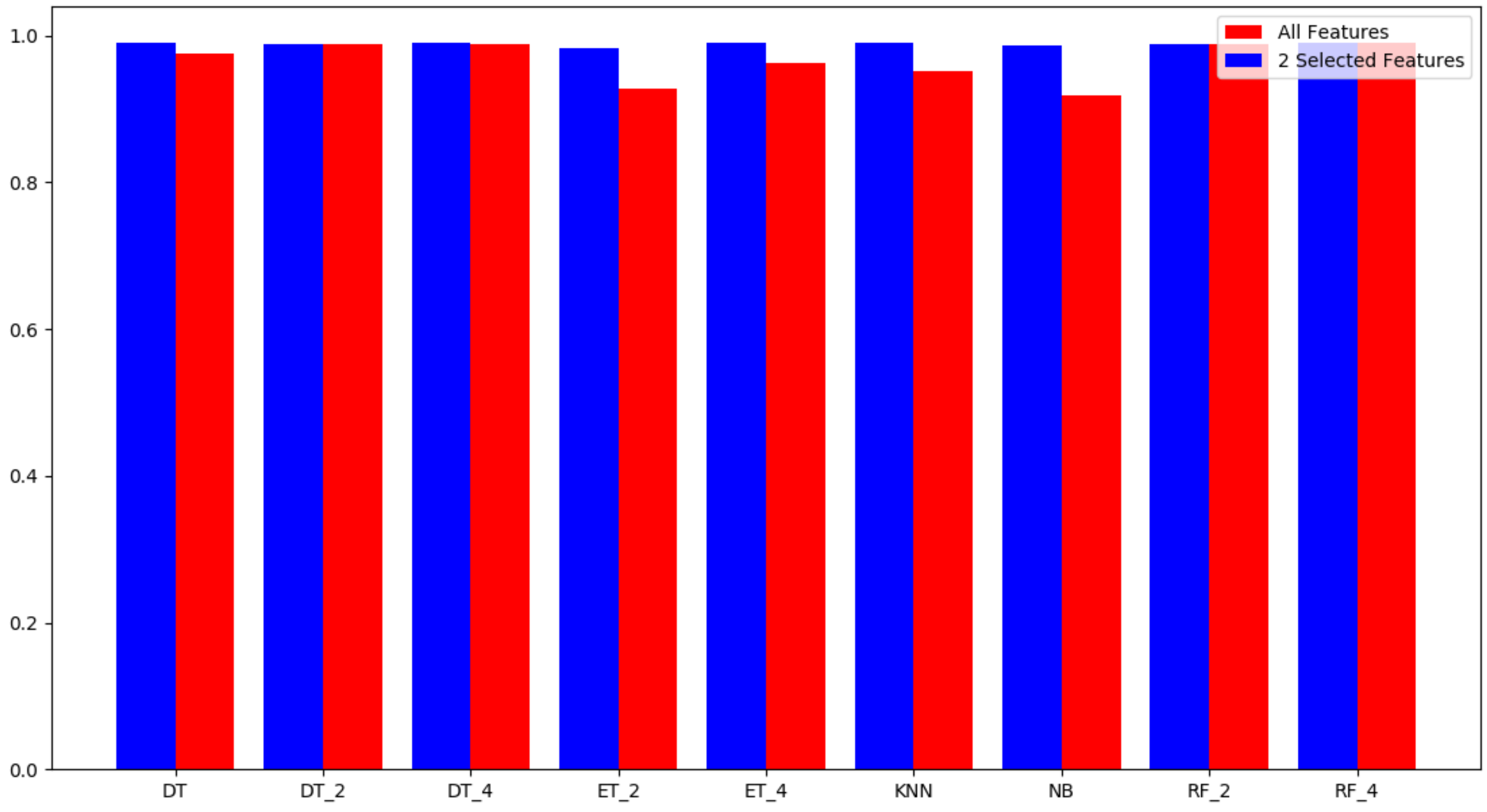
Results: Change in TTF (Table 4.1, 4.2)



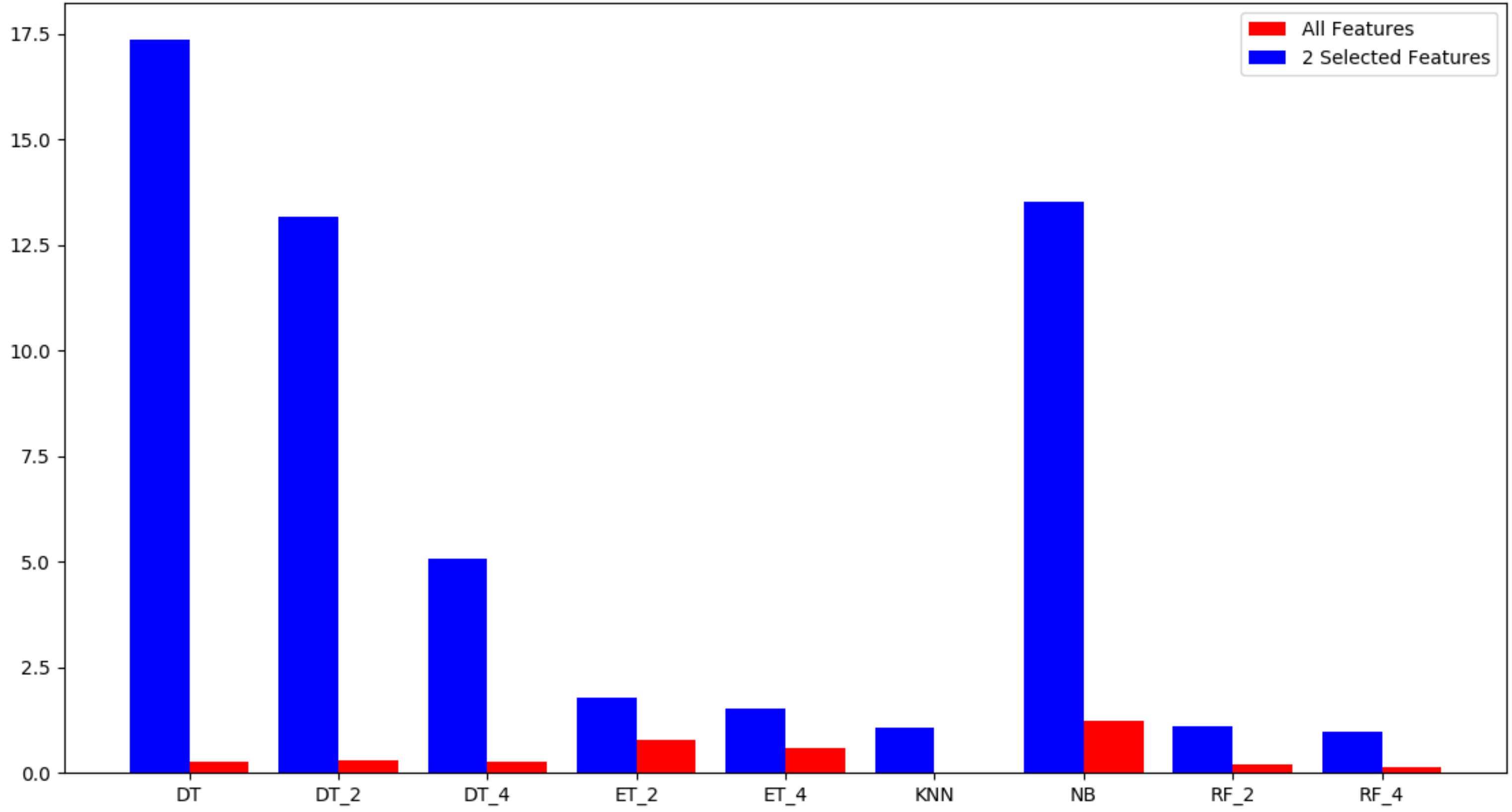
Results: Change in TTP (Table 4.1, 4.2)



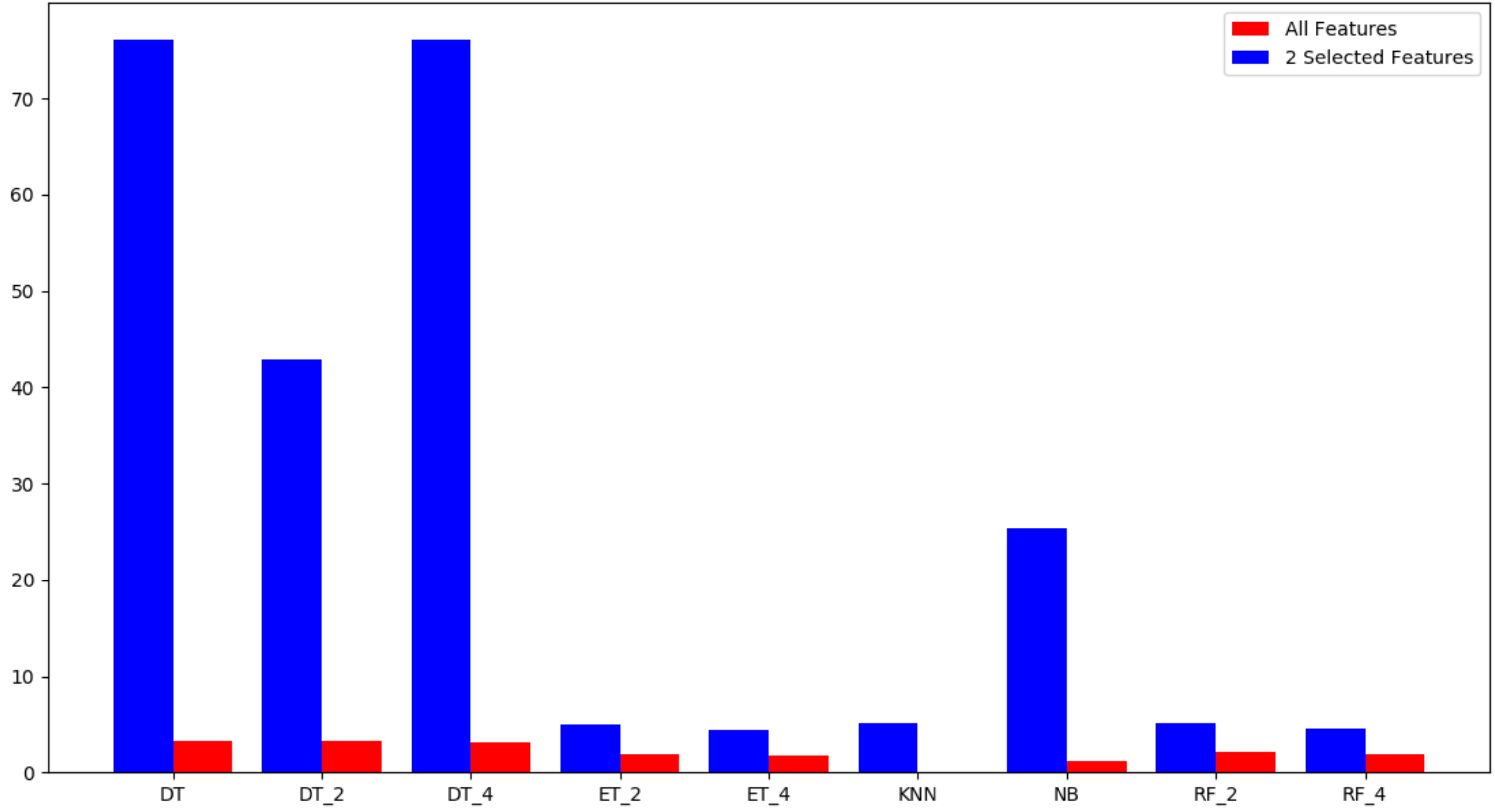
Results: Change in Accuracy (Table 4.1, 4.2)



Results: Change in TS (Table 4.3)



Results: Change in PS (Table 4.3)



Conclusion

- Importance of sttl, dttl and state for IDS.
- Decision Trees

References

- H. Debar, M. Dacier, and A. Wespi, “A revised taxonomy for intrusion-detection systems,” Annales Des T´el ´ecommunications, vol. 55, no. 7, pp. 361–378, Jul 2000. [Online]. Available: https://doi.org/10.1007/BF02994844
- P. V. S. Alpao, J. R. I. Pedrasa, and R. Atienza, “Multilayer perceptron with binary weights and activations for intrusion detection of cyberphysical systems,” in TENCON 2017 - 2017 IEEE Region 10 Conference, Nov 2017, pp. 2825–2829.
- M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the kdd cup 99 data set,” in 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, July 2009, pp. 1–6
- N. Moustafa and J. Slay, “Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set),” in 2015 Military Communications and Information Systems Conference (MilCIS), Nov 2015, pp. 1–6.
- N. Moustafa and J. Slay, “The significant features of the unsw-nb15 and the kdd99 data sets for network intrusion detection systems,” in 2015 4th International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS), Nov 2015, pp. 25–31.

References

- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikitlearn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
- H. Zhang, “The optimality of naive bayes,” 01 2004.
- T. K. Ho, “Random decision forests,” in Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, Aug 1995, pp.278–282 vol.1.
- P. Geurts, D. Ernst, and L. Wehenkel, “Extremely randomized trees,” Machine Learning, vol. 63, no. 1, pp. 3–42, Apr 2006. [Online]. Available:https://doi.org/10.1007/s10994-006-6226-1