SIT384 – Cyber Security Analytics

Task6.3HD Report

Name : Venujan Malaiyandi
Student ID : BSCP|CS|61|101

In this task 6.3HD, three classification algorithms were utilized on spambase dataset. They are decision trees, random forest, and finally, support vector machines. From the tests conducted, the accuracy results indicate that Random Forest outperformed the other models, achieving an accuracy of 95.55%. Decision Trees followed with an accuracy of 91.86%, while Support Vector Machines lagged behind with an accuracy of 66.23%.

In my opinion, the stark performance of Random Forest out of the three classification algorithms, can be traced to its ensemble nature. Random Forest algorithm combines multiple decision trees to form a more robust and accurate model. This prevents overfitting, by aggregating predictions from multiple trees. Additionally, Random Forest handles high-dimensional data well, making it suitable for the feature-rich Spambase dataset. On the other hand, Support Vector Machines struggled, possibly due to the dataset's non-linearity, as SVMs may not perform optimally when faced with complex decision boundaries.

The default hyperparameters were used for the initial implementation of each algorithm. I wanted not just an initial point of exploration, but also wanted to enquire about the baseline performance for all three models. This was especially useful in identifying and comparing the performance between the three classification models.

As a summary, the experimental results highlight the effectiveness of Random Forest in classifying the Spambase dataset. The ensemble approach of Random Forest, along with its ability to handle high-dimensional data, contributes to its superior performance. Decision Trees performed well but exhibited a slightly lower accuracy, while Support Vector Machines struggled, possibly due to the dataset's inherent complexity. Thus understanding the strengths and limitations of each algorithm is crucial in selecting the most suitable model for a given dataset and problem.