

## 1 Introduction

In the process of modeling the cybersecurity data, I explored multiple methodologies. These included logistic regression, a neural network model with Keras, a decision tree, and finally, a Gradient Boosting Machine (GBM).

## 2 Logistic Regression

Initially, I considered logistic regression due to its simplicity and interpretability. However, due to the complex and non-linear nature of the features in the dataset, this method was soon ruled out. Logistic regression can struggle with complex and high dimensional datasets as it assumes a linear relationship between the independent and dependent variables, which did not hold true for this dataset.

## 3 Neural Network with Keras

Next, a neural network using Keras was implemented. This approach was adopted due to the ability of neural networks to learn complex patterns and their proven success in various fields. Despite its potential, the performance on the dataset was sub-optimal. This could be due to the fact that neural networks require large amounts of data, and it can be challenging to determine the appropriate architecture for the specific problem.

## 4 Decision Tree

I then tried a decision tree model, which offers interpretability and can capture non-linear relationships. However, a decision tree on its own can easily overfit and perform poorly on unseen data. This was the case for this dataset, with the model showing high variance.

## 5 Gradient Boosting Machine (GBM)

Given the shortcomings of the previous models, I decided to use Gradient Boosting Machine (GBM). GBM is a powerful ensemble method that constructs new predictors that aim to correct the residuals errors of the prior predictor, reducing both bias and variance. The GBM model provided significantly improved results compared to the other models, handling the complexities and non-linearity in the data effectively.

## 6 Conclusion

The modeling process showed the importance of selecting a method suitable for the data's characteristics. Despite the potential of logistic regression, neu-

ral networks, and decision trees, these models were unsuitable for this specific dataset. In contrast, GBM provided a robust solution that effectively handled the complexities and idiosyncrasies of the cybersecurity data.