Milestone II Report

Comparison table of models

model	Recall Score	Training Time	Prediction Time
Decision Tree	0.2	3.22(s)	0.02(s)
Linear Regression	0	0.14(s)	0.02(s)
Random Forest	0.011	10.54(s)	0.2(s)
Lasso Regression	0	0.11(s)	0.003(s)
KNN	0.0046	1.38(s)	30.96(s)

Table 1: Models Performance

Metric Selection

Since our labels mostly consist of 0, we aim for our model to predict a substantial number of ones, making Recall a reasonable metric to use.

Models Evaluation

Decision Tree: Among the five models we use, this particular model exhibits the highest recall score. We anticipate that it will likely outperform regression models, which it indeed does. After hyperparameter tuning, this model achieves its peak performance at a maximum depth of 28, with a recall score of 0.2. One of the reasons this model may not perform better is likely due to the nature of the dataset, as mentioned previously; the majority of our labels are 0.

Linear Regression: We didn't have high expectations for its performance since our problem leans more towards classification, but we wanted to assess how this regression model would fare on our dataset. One challenge we encountered with regression models was converting their continuous predictions into discrete 0s and 1s for calculating the recall score. To achieve this, we had to set a threshold: predictions higher than the threshold were mapped to 1, and those lower were mapped to 0. We initially used the average of prediction labels on the training dataset as the threshold, but both the recall score and accuracy were poor. After experimenting with different thresholds, including 0.5, we couldn't find an optimal value. Ultimately, we settled on 0.5, but the recall score remained at 0, indicating poor performance. In addition to the reasons mentioned for the decision tree model's performance, another factor was our inability to determine the optimal threshold, as discussed earlier.

Random Forest: After observing the performance of the Decision Tree (DT) model, we anticipated that this new model's performance would be comparable. However, to our surprise, its performance was notably worse. The underlying reasons for this disappointing outcome appear to align closely with those observed in the DT model.

Lasso Regression: We sought to evaluate the predictive capabilities of an additional regression model on our dataset. Despite conducting tuning efforts, the performance closely resembled that of linear regression (LR). This finding suggests a consistency in model behavior across regression approaches.

KNN: We initially anticipated that this particular potentially reaching a similar level of performance to the Decision Tree (DT) model. However, to our surprise, this was not the case. One contributing factor to this discrepancy, observed in both the K-Nearest Neighbors (KNN) and Random Forest (RF) models, is the inability to adequately assess the impact of hyperparameter tuning due to longer-than-anticipated runtimes. We believe that after incorporating the results of our hyperparameter tuning efforts for these models, their performance should converge towards that of the DT model.