A table with numbers and letters

Description automatically generatedPart 1: Model exploration table

A chart with different colored boxes

Description automatically generatedPart 2: Box plot

The Decision Tree model's train and test box plots indicate minimal overfitting; however, the Out of Time (OOT) range from 0.47 to 0.58 is quite broad.

The Random Forest model's train and test plots show consistency, suggesting no significant overfitting. Nevertheless, the OOT performance is relatively low, failing to exceed 0.5, which could indicate underfitting or poor generalization on unseen data.

The LightGBM model's train and test box plots demonstrate stable and moderate fitting. The OOT range is narrower, from 0.48 to 0.55, tighter compared to the Decision Tree model.

The Neural Network model's train and test box plots reveal overfitting. Despite numerous tuning attempts, this remains the best result I could achieve.

Overall, I would choose **LightGBM** as the best model due to its robust performance on both train and test data and its smaller variability in OOT performance.

Part 3: Four “complexity plots”

1. Single Decision Tree:

model = DecisionTreeClassifier(max\_depth=i)

The graph displays how a DecisionTreeClassifier behaves as its depth increases from 1 to 15. As the tree depth goes up, the model fits the training data better, shown by the rising blue line. However, its ability to perform well on new data (orange line for testing and green line for OOT) starts well but then stops improving and even gets worse, which means the model is too complex and not generalizing well.

This indicates that the best tree depth is probably around where the testing accuracy stops getting better.

1. Random Forest



model = RandomForestClassifier(max\_depth=i, random\_state=42)

The graph illustrates the performance of a RandomForestClassifier as its depth increases from 1 to 15. As the tree depth increases, the model fits the training data increasingly well, as shown by the ascending blue line. Meanwhile, its ability to perform well on new data (orange line for testing) also improves, but at a slower pace, hinting at the beginning of overfitting as the model complexity continues to rise. The OOT performance, represented by the green line, remains relatively stable yet low, indicating that the model's generalization to completely new datasets is not improving in tandem with the training and testing improvements. This suggests that an optimal tree depth might be at a point before the testing accuracy gains begin to level off.

1. LightGBM

A graph of a number of different colored lines

Description automatically generated with medium confidence

model = lgb.LGBMClassifier(num\_leaves=10\*i, random\_state=42)

The graph illustrates the performance of a LightGBMClassifier as the number of leaves increases, controlled by 10\*i. Training accuracy (blue line) rapidly improves and then stabilizes, indicating a good fit to the training data. Testing accuracy (orange line) also improves but levels off, suggesting an optimal complexity point where further increases do not enhance model performance on unseen data. The OOT performance (green line), however, remains low and stable, highlighting a consistent failure to generalize to entirely new datasets. This suggests that while initial increases in complexity benefit the model, there's a threshold beyond which more complexity adds little value and could hinder generalization. Optimizing model settings or simplifying the model could improve its applicability and robustness in real-world scenarios. Regular evaluation with new data is crucial to maintain effectiveness.

1. Neural Network



model = MLPClassifier(hidden\_layer\_sizes=(10\*i,),random\_state=42)

The graph shows the performance of an MLPClassifier as its complexity increases, indicated by the number of neurons which scale with 10\*i. Training accuracy (blue line) consistently improves, demonstrating that more complex models fit training data better. However, testing accuracy (orange line) improves initially but begins to plateau, suggesting the beginning of overfitting. OOT performance (green line) remains flat and low, indicating poor generalization to completely new data. This implies that the optimal complexity, where the model best balances fit and generalization, may occur just before the testing accuracy levels off.