Supervised Learning

Implication

- Goal: predict **Claim Status** (Approved or Denied)
  - **Operational Efficiency:** Reduce manual error, saving time and costs.
  - **Customer Experience:** Faster claim approvals enhance satisfaction.
  - **Risk Management**: Insights help refine premiums and target high-risk customers

- Chose **Classification Tree Model**
  - Better performance and metrics compared to logistic regression model

Model issue and evaluations

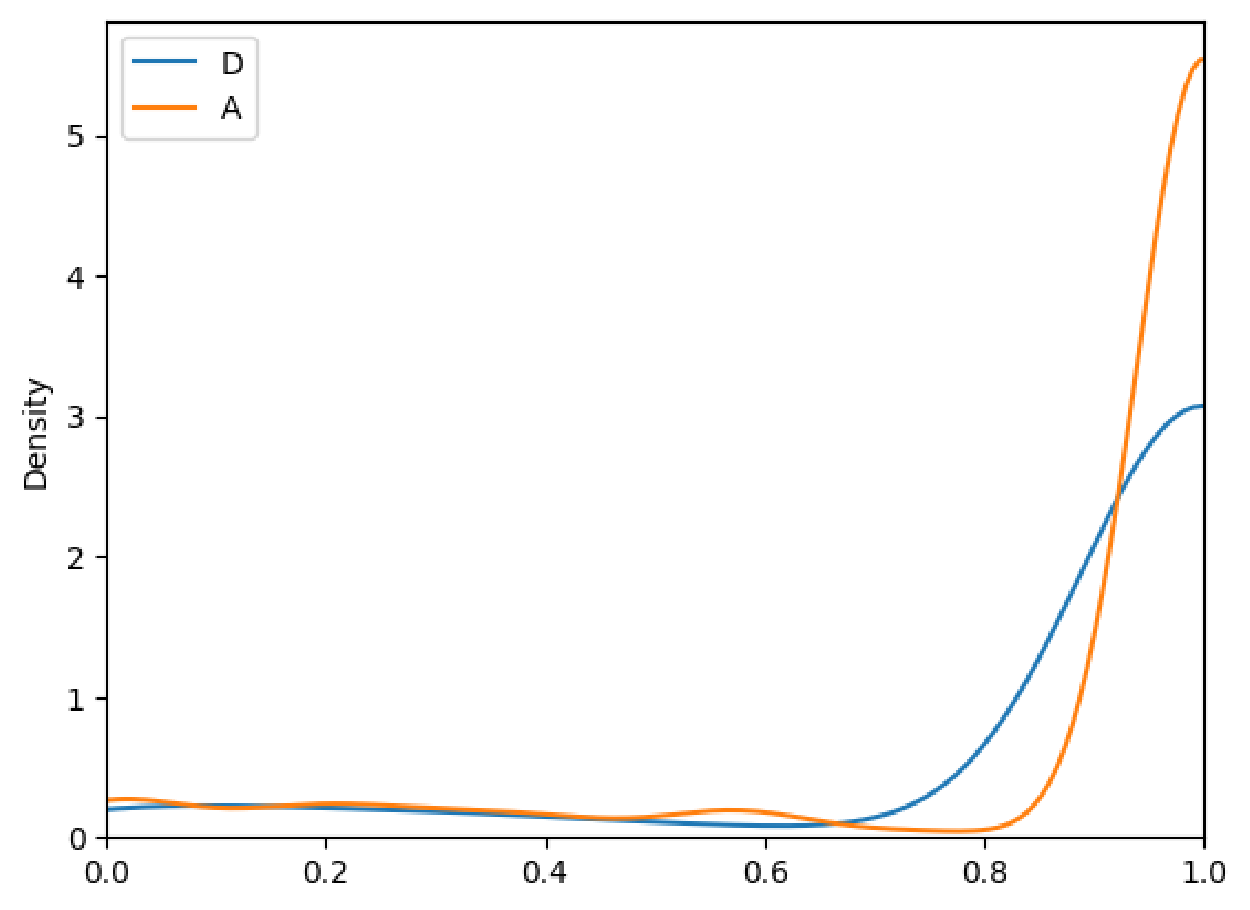
- Issue
  - Imbalance between approved and denied
- Solution
  - Overfit training data to resolve imbalance

	count
CLAIM_STATUS	
1	9497
0	503

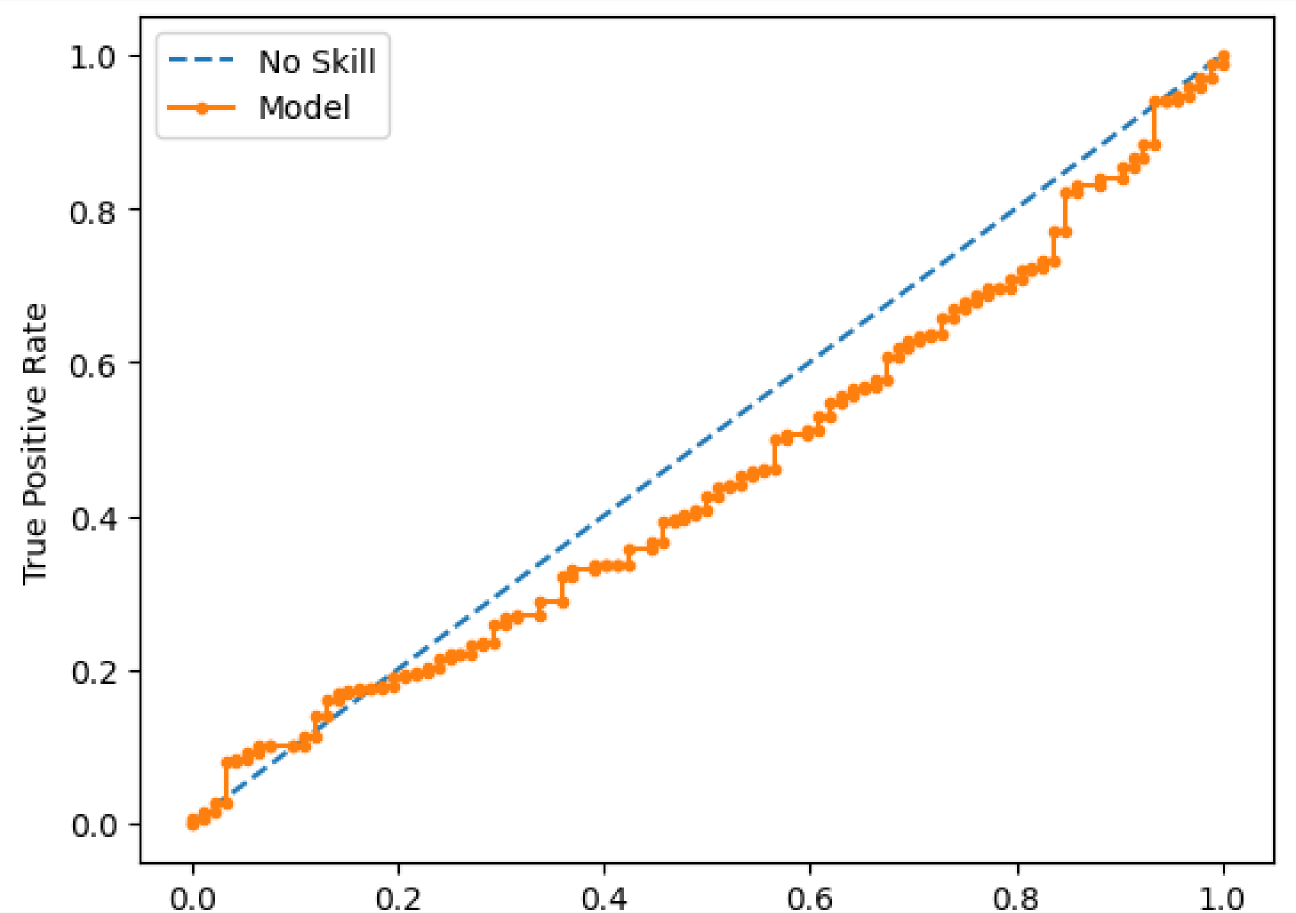
- Training Score: 0.95
- Testing Score: 0.85
- Implication:
  - Model can predict claim outcome accurately with in model and outside with new data

Continued

- High confidence in predicting Approval
- Moderate confidence in predicting Denied
- Slight overlapping 0.4-0.8 indicate ambiguous prediction



ROC Curve



- Decision Tree: ROC AUC=0.452
- Model struggles to differentiate between approved and denied claim

Confusion Matrix

- Recall: 88.05%
  - High recall indicates that the model correctly identifies a large proportion of approved claims (true positives).
  - This is excellent for use cases where missing approved claims would be costly.

- Precision: 95.4%
  - High precision means that when the model predicts an approved claim, it is almost always correct.
  - This reduces the risk of false positives, which is important for maintaining operational efficiency.

Comparison

Decision Tree

- Training Score: 0.95
- Testing Score: 0.85
- ROC AUC=0.452
- Recall: 88.05%
- Precision: 95.4%

Logistic Regression

- Training Score: 0.56
- Testing Score: 0.51
- ROC AUC=0.452
- Recall: 51.36%
- Precision: 94.59%