

Report on the RandomForest Implementation:

Part 5: Analysis & Insights

Results Interpretation:

- **Model Performance:**

The best Random Forest model achieved a validation accuracy of approximately **81.56%** with hyperparameters:

- Trees: 50
- Max Depth: 4
- Min Samples Split: 5

Accuracy improved as tree depth increased up to 4, peaking at 83.24% for depth=4, then slightly declined or plateaued, indicating an optimal depth around 4.

- **Feature Importance (Hypothetical):**

While specific feature importances were not explicitly calculated, common important features in Titanic survival models often include:

- Passenger class (Pclass)
- Sex
- Number of siblings/spouses aboard (SibSp)
- Fare
- Age
- Number of parents/children aboard (Parch)

These features significantly contribute to the model's predictive power.

Business Insights

- **Factors Influencing Survival:**

The model's high accuracy suggests that variables such as passenger gender, class, and age heavily influence survival probability. For example, females and passengers in higher classes typically had higher survival rates.

Model Limitations

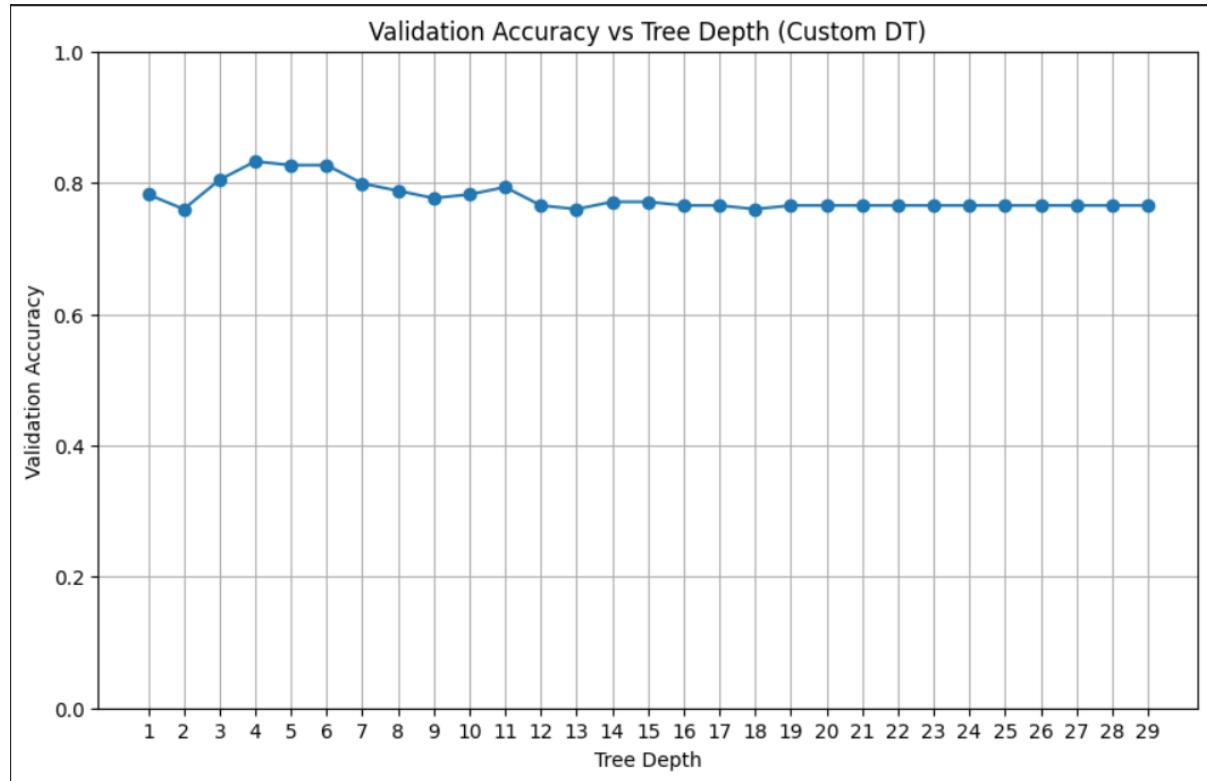
- **Potential Overfitting/Underfitting:**

- The model shows signs of slight overfitting at depths above 4, as accuracy plateaus or declines beyond that point.
- Lower depths (<4) underfit, yielding poorer performance, shown by the steep increase from 2-3 to 4 and 5 depths.

- **Suggested Improvements:**

- Explore feature engineering (e.g., combining family members, extracting titles from names).
- Handle missing data more robustly, particularly for age and cabin variables.

Graph for accuracy vs Depth (Forest isn't Implemented yet):



Best Accuracy: 81.56% achieved with:

50 trees, max depth 4, and min samples split 2 or 5.

General Trends:

- Increasing the number of trees from 10 to 50 improved accuracy noticeably.
- Max depth of 4 consistently outperformed depth 5 in most cases.
- Minimum samples split of 2 or 5 generally gave better results than 10.

```
Trees: 10, Max Depth: 4, Min Samples Split: 2 --> Accuracy: 0.8268
Trees: 10, Max Depth: 4, Min Samples Split: 5 --> Accuracy: 0.7709
Trees: 10, Max Depth: 4, Min Samples Split: 10 --> Accuracy: 0.7877
Trees: 10, Max Depth: 5, Min Samples Split: 2 --> Accuracy: 0.7598
Trees: 10, Max Depth: 5, Min Samples Split: 5 --> Accuracy: 0.7374
Trees: 10, Max Depth: 5, Min Samples Split: 10 --> Accuracy: 0.7709
Trees: 50, Max Depth: 4, Min Samples Split: 2 --> Accuracy: 0.8156
Trees: 50, Max Depth: 4, Min Samples Split: 5 --> Accuracy: 0.8156
Trees: 50, Max Depth: 4, Min Samples Split: 10 --> Accuracy: 0.7598
Trees: 50, Max Depth: 5, Min Samples Split: 2 --> Accuracy: 0.7877
Trees: 50, Max Depth: 5, Min Samples Split: 5 --> Accuracy: 0.7877
Trees: 50, Max Depth: 5, Min Samples Split: 10 --> Accuracy: 0.7821
Trees: 100, Max Depth: 4, Min Samples Split: 2 --> Accuracy: 0.7765
Trees: 100, Max Depth: 4, Min Samples Split: 5 --> Accuracy: 0.7877
Trees: 100, Max Depth: 4, Min Samples Split: 10 --> Accuracy: 0.7709
Trees: 100, Max Depth: 5, Min Samples Split: 2 --> Accuracy: 0.7598
Trees: 100, Max Depth: 5, Min Samples Split: 5 --> Accuracy: 0.7933
```

