****

**VIRGINIA COMMONWEALTH UNIVERSITY**

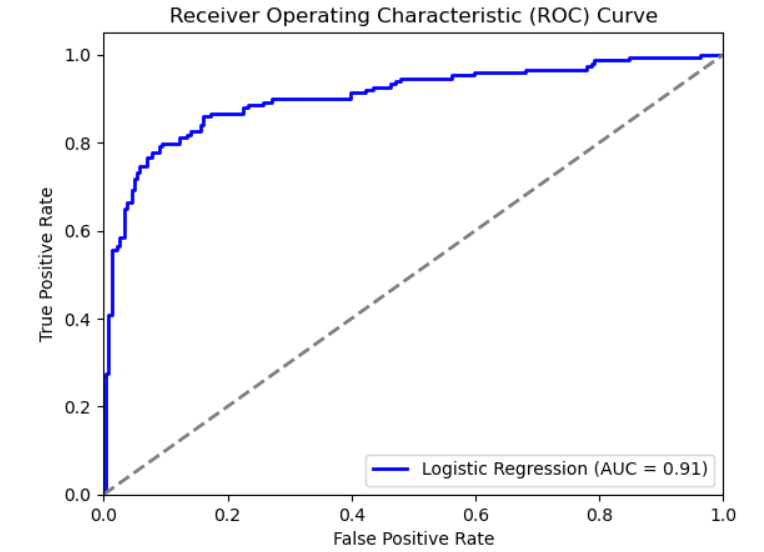
**Statistical analysis and modelling (SCMA 632)**

**A3 Part A**

**Rohan Arpit Dungdung**

**V01108247**

**Date of Submission: 15-07-2024**



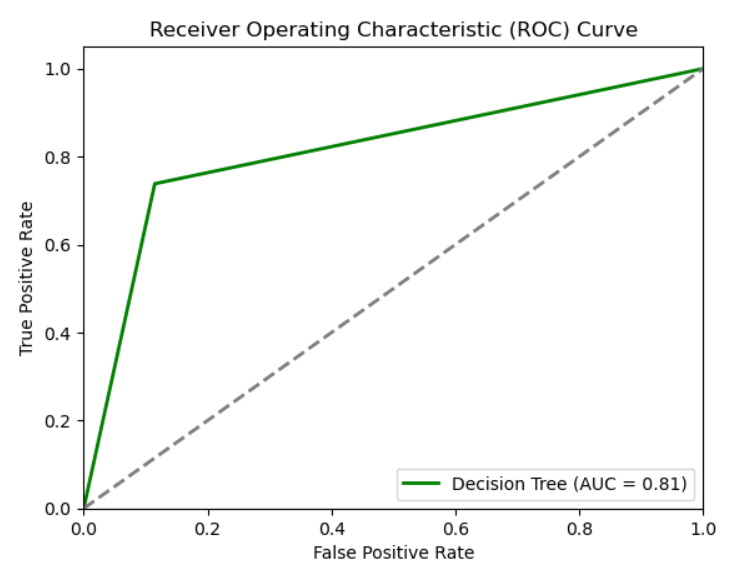
Interpretation

The ROC curve in this image is positioned significantly above the diagonal line, indicating that the logistic regression model has good discriminative power. The curve's shape shows a steep rise towards the top left corner, which is indicative of high sensitivity and specificity.

The Area Under the Curve (AUC) is a key metric in evaluating the performance of a classification model. In this case, the AUC is 0.91, which is quite high. An AUC of 0.91 means that the model has a 91% chance of correctly distinguishing between positive and negative classes (e.g., predicting whether a passenger survived or not).

A higher AUC value (closer to 1) indicates better performance, with 1 being a perfect classifier and 0.5 representing a random guess.

The ROC curve shows that the logistic regression model achieves high sensitivity and specificity, which means it can accurately identify true positives for example correctly predicting survival) while minimizing false positives for example predicting survival when the passenger did not survive).



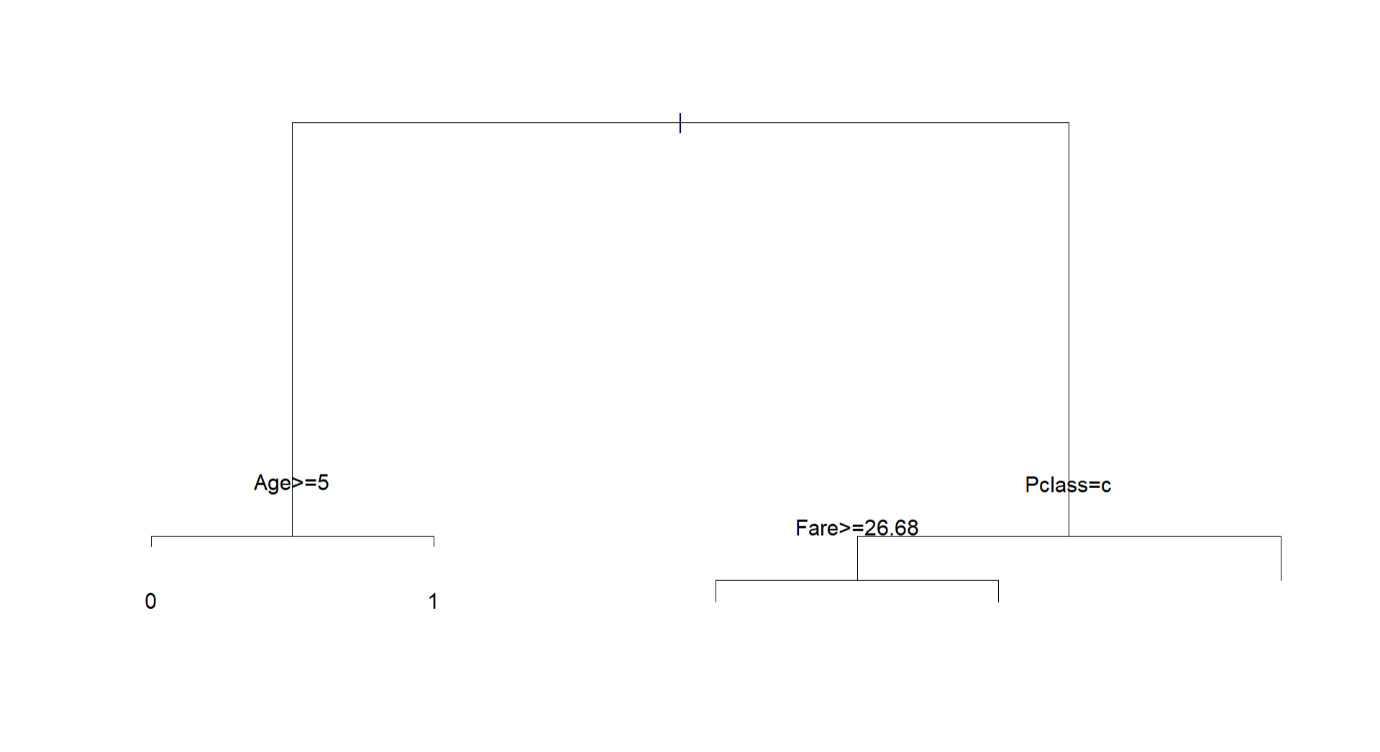
Interpretation

The ROC curve for the decision tree model rises above the diagonal line, indicating that the model performs better than random guessing. However, compared to an ideal model, the curve does not reach as close to the top-left corner, suggesting there are some limitations in distinguishing between positive and negative classes.

The Area Under the Curve (AUC) for the decision tree model is 0.81. This value indicates that the model has a good but not outstanding ability to discriminate between the positive and negative classes. An AUC of 0.81 means that there is an 81% chance that the model will correctly classify a randomly chosen positive instance higher than a randomly chosen negative instance.

An AUC closer to 1 indicates a better-performing model, while an AUC of 0.5 suggests no discriminative power (random guessing).

The ROC curve demonstrates the trade-off between sensitivity and specificity. The decision tree model appears to achieve a good balance, but there is room for improvement, particularly in minimizing false positives while maintaining high sensitivity.



Interpretation

The decision tree highlights some key factors associated with survival on the Titanic, such as class, age, and fare. For example, being in a higher class (class C) and paying a higher fare seems to correlate with a higher likelihood of survival. The model is simple and interpretable, providing clear decision rules, but it may also be limited in capturing more complex relationships in the data.