

MARSHA ARIF MUQRI

22F-BSAI-07

AI (A1)

MACHINE LEARNING ASSIGNMENT REPORT

Titanic Survival Prediction

INTRODUCTION:

The following project applies implementing multiple machine learning models to predict whether a passenger survived the Titanic tragedy and uses ensembling techniques to improve accuracy. The prediction is made using passenger details such as age, gender, and other available attributes.

DATA PREPROCESSING:

Before training the models, the Titanic dataset was prepared to ensure it was suitable for machine learning algorithms. Therefore several preprocessing steps were carried out, including:

- Handling Missing Values:
Some passengers did not have age information available. These missing age values were replaced with the average age of the dataset. Similarly, missing embarkation ports were filled using the most frequently occurring port.
- Removing Irrelevant Features:
Certain columns, such as city names, were either repetitive or had too many missing values. These features were removed to reduce noise and improve model performance.
- Encoding Categorical Variables:
Since machine learning models require numerical input, categorical data such as gender was converted into numeric form (for example, Male = 1 and Female = 0).
- Feature Scaling:
To prevent features with larger numerical ranges from dominating the learning process, all numerical values were scaled to a standard range.

MODEL EVALUATION AND SELECTION:

In this project, we used four different classification models and tested them to determine which one provided the best performance. Their results and the reason for using the model is summarized below:

Model	Accuracy	Reason for Use
Decision Tree	78.77%	We used this model as it uses a simple rule-based, tree-like structure,
Naive Bayes	77.65%	We used this model as it is a probability-based and computationally efficient model.
Support Vector Machine (SVM)	78.77%	This model is effective at separating data into distinct classes.
Random Forest	82.12%	This model was our best performer with the accuracy of 82%. This is because it combines multiple decision trees to improve accuracy.

CONCLUSION:

In summary, We successfully developed a model capable of predicting Titanic survival with an accuracy of approximately 82% and that was the Random Forest model. The findings indicate that survival outcomes were influenced by identifiable patterns rather than chance. Future improvements could include fine-tuning the Random Forest model's parameters to further enhance prediction accuracy.

\