

NAME: Ayesha Tariq

Roll No: 22-F-BSAI-41

Project Report: Predicting Survival on the Titanic

1. Introduction

This project uses Machine Learning to predict whether a passenger would have survived the Titanic disaster based on information like their age, gender.

2. Data Preprocessing Decisions

Before feeding the data to our models, we had to "clean" it to make sure the computer could understand it.

the key steps are following:

- **Handling Missing Information:**

Not all passengers had their age recorded. We filled in missing ages with the "average age" and missing ports with the most common one.

- **Dropping Unnecessary Details:**

Some information like the city name, was either mostly missing or repetitive. We removed these to keep the model focused on the most important details.

- **Converting Words to Numbers:**

Computers work best with numbers. We converted categories like "Male/Female" into numerical codes (e.g., 0 and 1).

- **Standardizing the Scale:**

Since some numbers are very large and others like ("Number of Siblings") are small, we scaled all data so that no single feature would unfairly dominate the model's decision-making.

3. Model Comparison and Justification

We tested four different types of models to see which performed best. Here is how they compared:

Model	Accuracy	Why we used it
Decision Tree	78.77%	It follows a flowchart-like logic.
Naive Bayes	77.65%	A simple model based on probability.
SVM (Support Vector Machine)	78.77%	Good at finding a clear boundary between groups.
Random Forest (Ensemble)	82.12%	The Winner. It combines many Decision Trees together to get a more reliable "vote".

Conclusion

In simple words, we successfully built a tool that can predict Titanic survival with 82% accuracy. The data shows that survival wasn't just random; it followed specific patterns based on a passenger's background. For future work, we could improve this further by "tuning" the settings of the Random Forest to squeeze out even more accuracy.