

TASK-4

Given Titanic Dataset, for each passenger in the test set, use the model you trained to predict whether they survived the sinking of the Titanic.

Expected Result:

- **How you approach this problem**
- **What Algorithms you can use to solve this problem**
- **Code Completion**
- **Results Explanation with Errors and Probability Score.**

This Titanic survival prediction model follows a structured approach to data preprocessing, model training, validation, and prediction. First, we load the dataset and handle missing values by imputing median values for numerical features such as age and fare while filling categorical missing values with the most common occurrences. Next, we encode categorical variables like "Sex" and "Embarked" into numerical values to make them compatible with machine learning algorithms. We also engineer a new feature, "FamilySize," which combines the number of siblings, spouses, parents, and children aboard to improve prediction accuracy.

For model training, we select relevant features such as "Pclass," "Sex," "Age," "Fare," "Embarked," and "FamilySize." The dataset is then split into training and validation sets to assess model performance. A Random Forest Classifier with 100 estimators is trained on the data, achieving an accuracy of 82.68% on the validation set. The classification report indicates a balanced performance across both survival and non-survival classes, with precision and recall values close to each other. After validating the model, we use it to predict survival for passengers in the test dataset and also compute probability scores for each prediction. The final results are saved to a CSV file, which includes the passenger ID, survival prediction, and probability score, allowing for further analysis. This model effectively demonstrates the application of machine learning techniques to a real-world classification problem.