**Final Report: Titanic Survival Prediction**

**1. Objective**

The goal of this project is to predict passenger survival on the Titanic using machine learning. The analysis follows these steps:

1. Load and explore the dataset.
2. Perform data preprocessing.
3. Train and evaluate a machine learning model.
4. Visualize findings.

**2. Data Loading and Exploration**

**2.1 Importing Libraries**

The following libraries were imported to facilitate data handling, visualization, and machine learning:

* pandas & NumPy: Data manipulation.
* matplotlib & seaborn: Visualization.
* scikit-learn: Model training and evaluation.

**2.2 Loading the Dataset**

The dataset was loaded using Seaborn’s built-in Titanic dataset.

**2.3 Dataset Overview**

The dataset was examined using:

* .head() to display the first few rows.
* .shape to check dimensions.
* .info() for an overview of data types and missing values.
* .describe() for statistical summaries.

Findings:

* The dataset contains 891 rows and 15 columns.
* Missing values exist in columns such as age, embarked, and deck.

**3. Data Preprocessing**

**3.1 Handling Missing Values**

* deck was dropped due to excessive missing values.
* age was filled with the median value.
* embarked was filled with the mode.

**3.2 Encoding Categorical Variables**

Categorical variables (sex, embarked, class, etc.) were converted into numerical values using LabelEncoder.

**3.3 Feature Selection & Splitting Data**

The target variable (survived) was separated, and the dataset was split into training (80%) and testing (20%) sets.

**3.4 Feature Scaling**

A StandardScaler was applied to normalize numerical variables.

**4. Machine Learning Model**

**4.1 Training a Random Forest Classifier**

A Random Forest Classifier was used due to its robustness in handling classification problems.

**4.2 Model Evaluation**

The trained model was tested on the test set, and performance metrics were computed.

Results:

* **Accuracy:** ~81%.
* **Confusion Matrix:** Showed correct and incorrect predictions.
* **Classification Report:** Provided precision, recall, and F1-score.
* **ROC Curve:** Evaluated classification performance.

**5. Data Visualizations**

**5.1 Survival Rate by Gender**

Findings:

* Women had a higher survival rate than men.

**5.2 Age Distribution of Survivors**

Findings:

* Most survivors were younger passengers.

**5.3 Class-wise Survival Rate**

Findings:

* First-class passengers had a significantly higher survival rate.

**6. Conclusion**

* Gender and passenger class significantly influenced survival chances.
* Random Forest achieved **~81% accuracy**, showing decent predictive power.
* Further improvements can include hyperparameter tuning, feature engineering, and trying other models.