

Titanic Survival Prediction Using Machine Learning

Importing Libraries

- pandas → Data manipulation & analysis
- numpy → Numerical computations
- seaborn → Statistical visualizations
- matplotlib.pyplot → Charts & plots

Loading Data

- Load CSV files with pandas
- Separate Train & Test datasets
- Dataset contains passenger details – basis of survival prediction

Exploratory Data Analysis (EDA)

- Distributions: Age, SibSp, Parch, Fare
- Relationships: Explore links between features & survival
- Helps identify patterns, trends, and outliers

Data Cleaning

- Handle missing values (e.g., imputation)
- Remove irrelevant columns (PassengerId, Name, Ticket, Cabin)
- Feature Engineering – Create/transform variables for better prediction

Model Testing

- Models used:
 - Decision Tree, Random Forest, Extra Trees
 - XGBoost, LightGBM, Logistic Regression
- Best Accuracy: 73.8%

Test Submission

- Apply chosen model on test dataset
- Generate final output file (submission.csv)
- Evaluate model on unseen data

Conclusion

- Model achieved 73.8% accuracy
- Showed key survival factors (age, gender, class, etc.)
- Demonstrates ML application on historical data