TITANIC SURVIVAL PREDICTION

PROJECT OVERVIEW

Project: Predicting Titanic Passenger Survival

****Overview****

I developed a machine learning model to predict the likelihood of survival for passengers on the Titanic using historical passenger data. This project involved various stages including data cleaning, feature engineering, exploratory data analysis, model selection, evaluation, and deployment using Python and key data science libraries. The goal was to build an accurate and efficient predictive model to provide insights into the factors influencing survival.

****Problem Statement****

The Titanic disaster is one of the most infamous maritime tragedies. The objective of this project was to predict which passengers were likely to survive the sinking of the Titanic based on their personal attributes and travel details.

****Tools & Technologies****

- **Python:** The primary programming language used.

- **Pandas & NumPy:** For data manipulation and analysis.

- **Scikit-Learn:** For building and evaluating machine learning models.

- **Matplotlib & Seaborn:** For data visualization and exploratory data analysis.

- **Flask:** For deploying the model as an interactive web application.

****Data Preparation****

- **Data Cleaning:** Handled missing values and inconsistencies in the dataset. For example, filled missing age values with the median age, and inferred missing embarkation ports.

- **Feature Engineering:** Created new features such as:

- **Family Size:** Combined the number of siblings/spouses and parents/children aboard.

- **Title Extraction:** Extracted titles (Mr., Mrs., Miss, etc.) from passenger names.

- **Cabin Deck:** Extracted the deck level from cabin numbers.

- **Encoding Categorical Variables:** Converted categorical variables (e.g., Sex, Embarked) into numerical format using one-hot encoding.

****Exploratory Data Analysis (EDA)****

- Visualized distributions and relationships between features using histograms, box plots, and heatmaps.

- Identified key factors affecting survival, such as passenger class, gender, and age.

****Model Development****

- Algorithm Selection: Evaluated multiple machine learning algorithms including:

  - Logistic Regression

  - Decision Trees

  - Random Forest

  - Gradient Boosting

- **Model Tuning:** Used cross-validation and grid search to optimize hyperparameters and improve model performance.

- **Evaluation Metrics:** Assessed models using accuracy, precision to ensure balanced and robust performance.

****Results****

-Performance: The final model achieved:

  - Accuracy: 78.21%

****Key Insights**:**

  - Passenger Class: Higher survival rates were observed for first-class passengers.

  - Gender: Females had a higher likelihood of survival compared to males.

  - Age: Younger passengers had a higher survival rate.

  - Family Size: Passengers with smaller families had a slightly higher chance of survival.

****Deployment****

- Interactive Web Application: Deployed the model using Flask, allowing users to input passenger details and get real-time survival predictions.

- User Interface: Designed a simple and intuitive interface for users to enter details such as age, gender, passenger class, etc., and obtain predictions.

****Key Learnings****

- **Data Preprocessing:** The importance of handling missing data and creating meaningful features to improve model accuracy.

- **Model Evaluation:** Understanding the trade-offs between different performance metrics and selecting the best model based on balanced performance.

- **Deployment:** Gaining practical experience in deploying a machine learning model and creating an interactive user interface for real-world applications.

****GitHub Repository****

Explore the complete project and code here:

https://github.com/sravanthi224/TITANIC-SURVIVAL-PREDICTION.git