**Titanic Survival Prediction**

**Objective**

The goal of this project is to build a machine learning model to predict whether a passenger survived the Titanic disaster based on features such as age, gender, ticket class, fare, and cabin information.

**Project Structure**

├── titanic-survival-prediction.ipynb # Main notebook with full implementation

├── README.md # Project documentation

├── requirements.txt # Python dependencies

└── dataset/ # Dataset file

**Steps to Run the Project**

1. **Clone the repository:**

git clone https://github.com/Abhi2122/Titanic-Survival-Prediction.git

cd titanic-survival-prediction

1. **Install dependencies:**

pip install -r requirements.txt

1. **Download dataset:**

Dataset can be downloaded from [Kaggle - Titanic Dataset](https://www.kaggle.com/datasets/brendan45774/test-file).

1. **Run the notebook:**

Open titanic-survival-prediction.ipynb in Jupyter Notebook / VSCode / Google Colab and run all cells.

**Project Highlights:**

* Handled missing values in **Age**, **Embarked**, and **Cabin** columns.
* Encoded categorical variables: **Sex**, **Embarked**, **Pclass**.
* Applied normalization to numerical features.
* Tried various models: XGBoost, Logistic Regression, Decision Tree, Random Forest.
* Evaluated using accuracy, precision, recall, F1-score.
* Achieved **82% accuracy** with XGBoost model.

**Performance Summary:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-score |
| XGBoost | 0.8212 | 0.8000 | 0.7568 | 0.7778 |
| Random Forest | 0.8156 | 0.8361 | 0.6892 | 0.7556 |
| Spport Vector Machine | 0.8156 | 0.8060 | 0.7297 | 0.7660 |
| Logistic Regression | 0.8101 | 0.7857 | 0.7432 | 0.7639 |
| Decision Tree | 0.7989 | 0.8276 | 0.6486 | 0.7273 |
| LDA | 0.7933 | 0.7681 | 0.7162 | 0.7413 |
| Naive Bayes | 0.7710 | 0.7200 | 0.7297 | 0.7248 |