Titanic Survival Prediction

A machine learning project to predict passenger survival on the Titanic using the Titanic dataset. This project preprocesses the data, engineers new features, trains multiple models, and optimizes the best-performing model to achieve high accuracy.

Project Overview

- **Data Preprocessing**: Handling missing values, encoding categorical variables, and normalizing numerical features.
- **Feature Engineering**: Creating meaningful features like `FamilySize` and `IsAlone`.
- **Model Training**: Comparing Logistic Regression, Random Forest, and XGBoost.
- **Hyperparameter Tuning**: Optimizing the Random Forest model with GridSearchCV.
- **Evaluation**: Reporting accuracy, precision, recall, and F1-score.

Requirements

- Python 3.8+
- Libraries:
  - pandas
  - numpy
  - scikit-learn
  - xgboost

Install dependencies:

```bash
pip install -r requirements.txt
```

How to Run

1. **Clone the repository**:
   ```bash
   git clone https://github.com/venkat-0706/Titanic-Survival-Prediction.git
   cd Titanic-Survival-Prediction
   ```

2. **Install dependencies**:
   ```bash
   pip install -r requirements.txt
   ```

3. **Add dataset**:
   Place `tested.csv` in the `data/` folder (or update the file path in `main.py`).

4. **Run the script**:
   ```bash
   python main.py
   ```

Results

Sample output:

```
Random Forest:
Accuracy: 0.8214
Precision: 0.8000
Recall: 0.7273
F1-Score: 0.7619
```

Final Test Accuracy: 0.8393

Author

Venkat-0706

GitHub: https://github.com/venkat-0706

LinkedIn: https://linkedin.com/in/venkat-0706