Executive Summary: Titanic Survival Analysis

Project Name: Titanic Survival Prediction
Date: 08th March 2025
Name: Arun Chaturvedi
GitHub: https://github.com/abhay1446/abhay1446
LinkedIn: www.linkedin.com/in/arun-chaturvedi-62941b121

This project aimed to predict the survival chances of Titanic passengers using a **Random Forest Classifier**. The analysis involved several key steps, including data preprocessing, feature engineering, visualization, and model training.

Key Insights:

1. **Data Preprocessing:**
    - Handled missing values in critical columns such as Age, Cabin, and Embarked.
    - Extracted titles from passenger names to create a new feature, grouping them into common categories (e.g., "Royal", "Rare").
2. **Exploratory Data Analysis (EDA):**
    - Visualized survival rates based on gender, passenger class, and family size.
    - Showed that **women had a significantly higher survival rate than men**.
    - First-class passengers had a better chance of survival compared to lower classes.
3. **Feature Engineering:**
    - Created new meaningful features like **FamilySize** and **IsAlone**.
    - Mapped categorical values to numerical values for model training.
4. **Model Training and Performance:**
    - Used a **Random Forest Classifier**, achieving an accuracy of **83.8%** on the training set.
    - The model effectively captured important survival predictors like gender, ticket class, and fare.

Conclusion:

The analysis confirmed that gender, class, and ticket fare were strong determinants of survival. The trained model performed well, demonstrating the power of machine learning in predictive analytics.