Predicting Titanic Survivors Using Logistic Regression

Project 3 – Coding Samurai Internship

By Ridhwan S | Data Analyst

What's the Mission?

- Predict if a passenger survived the Titanic disaster.
- Use features like age, gender, and passenger class.
- Apply a supervised machine learning model.

Let’s Meet the Data

- Sourced from Kaggle Titanic Challenge.
- Total Rows: 891 | Columns: 12
- Target Column: Survived (0 = No, 1 = Yes)

💡 Feature Examples:

- Age, Sex, Pclass, Fare, SibSp, Embarked

Data Cleaning Steps

- Filled missing Age values with median.
- Dropped Cabin due to many nulls.
- Imputed missing Embarked values.
- Encoded categorical variables (Sex, Embarked).

Prepping for the Model

- Converted categorical to numeric.
- Selected most informative features:
  - Pclass
  - Sex
  - Age
  - Sibps
  - Fare
  - Embarked

Logistic Regression Model

- Used train_test_split with 80/20 split.
- Applied StandardScaler for feature scaling.
- Trained using Scikit-learn’s Logistic Regression.
- Binary classification problem.

Model Performance

- Accuracy: **81%**
- Confusion Matrix (Insert Plot Image)
- Classification Report Summary (precision, recall, F1)

What We Learned

- Gender was a strong predictor (Females survived more).
- Pclass had a direct impact on survival.
- Logistic Regression delivered reliable results with interpretability.

Ideas to Go Further

- Try Random Forest or XGBoost for comparison.
- Hyperparameter tuning with GridSearchCV.
- Handle class imbalance with SMOTE or class weights.

Let’s Connect!

Ridhwan S | Data Analyst

LinkedIn: Profile

GitHub: Coding Samurai Project-3