

# Classification: Logistic Regression on Titanic Dataset

Predicting Passenger Survival using Logistic Regression  
by **Ridhwan S** — Data Analyst Intern | Coding Samurai  
April 2025

# Objective



## **Project Objective**

To build a **Logistic Regression model** that predicts passenger survival on the Titanic using features like age, sex, and class.

# Dataset Overview



## **Dataset: Titanic (Kaggle)**

- Rows: 891
- Columns: 12
- Target Variable: Survived (0 = No, 1 = Yes)
- Key Features:
  - Pclass, Sex, Age, SibSp, Parch, Fare, Embarked

# Tools & Libraries



## Tech Stack

- **Language:** Python
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- **Techniques:** Data Cleaning, Feature Engineering, Logistic Regression, Model Evaluation

# Data Cleaning



## **Cleaning Steps**

- Dropped irrelevant columns: PassengerId, Name, Ticket, Cabin
- Filled missing Age values with median
- Filled missing Embarked with mode
- Converted categorical features(Sex, Embarked) to numeric

# EDA Highlights

## **Exploratory Insights**

- Females had higher survival rates than males
- Higher class (1st) had better survival chances
- Younger passengers had slightly better survival odds

# Feature Engineering

## Engineered Features

- Encoded Sex: male  $\rightarrow$  0, female  $\rightarrow$  1
- One-hot encoded Embarked
- Selected features:
  - Pclass, Sex, Age, SibSp, Parch, Fare, Embarked\_Q, Embarked\_S

# Model Building



## **Logistic Regression**

- Split data: 80% Train / 20% Test
- Model: `LogisticRegression(max_iter=1000)`
- Library: Scikit-learn



# Evaluation Metrics



## Model Performance

- **Accuracy:** 81%
- **Confusion Matrix:**  
$$\begin{bmatrix} 90 & 15 \\ 19 & 55 \end{bmatrix}$$
- **Precision / Recall / F1-Score:**
  - Class 0: 0.83 / 0.86 / 0.84
  - Class 1: 0.79 / 0.74 / 0.76

# Conclusion

## **Key Takeaways**

- Logistic Regression is interpretable and fast for binary classification
- Good accuracy (81%) without hyperparameter tuning
- Future Scope: GridSearchCV, cross-validation, ensemble models

# Thank You

## Thank You!

 Let's connect on LinkedIn: [Profile](#)

 GitHub Repo: [Project 4](#)

 “Great models come from great questions.”