AI & ML Internship – Task 7 Report

Title: Logistic Regression – Titanic Survival Prediction

Objective:

The objective of this task is to build a binary classification model using Logistic Regression to predict whether a passenger survived the Titanic disaster based on key features.

Dataset Used: Titanic Dataset (Seaborn / Kaggle)

Tools & Libraries:

Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

Methodology:

- Loaded the Titanic dataset and selected important features.
- Handled missing values using median and mode strategies.
- Encoded categorical features using One-Hot Encoding.
- Scaled numerical features using StandardScaler.
- Split the dataset into training and testing sets.
- Trained a Logistic Regression model.
- Evaluated the model using accuracy, precision, recall, F1-score.
- Generated confusion matrix and ROC curve with AUC score.

Evaluation Metrics:

- Accuracy: Measures overall correctness of the model.
- Precision: Measures correctness of positive predictions.
- Recall: Measures ability to identify actual survivors.
- F1-Score: Harmonic mean of precision and recall.
- ROC Curve & AUC: Measures classification performance across thresholds.

Conclusion:

This task provided practical understanding of Logistic Regression for binary classification. Proper preprocessing and evaluation metrics ensured reliable performance analysis. The model demonstrated effective prediction capability for Titanic survival.