|  |  |
| --- | --- |
| **NAME** | **MUHAMMAD HARIS** |
| **SUBJECT** | **DATA SCIENCE** |
| **TOOL** | **GOOGLE COLAB** |

**Project Title: Titanic Survival Prediction using Machine Learning**

**1. Objective**

The objective of this project is to use Data Science and Machine Learning techniques to predict whether a passenger would survive the Titanic disaster based on their details such as age, gender, ticket class, and fare.

**2. Dataset Description**

The dataset used in this project contains the following columns:

* Passenger class (pclass)
* Gender (sex)
* Age (age)
* Number of siblings/spouses (sibsp)
* Number of parents/children (parch)
* Ticket fare (fare)
* Port of embarkation (embarked)
* Whether the passenger was traveling alone (alone)
* Survival status (survived)

**3. Tools and Libraries Used**

* Python
* Google Colab
* Pandas
* Scikit-learn
* Logistic Regression model

**4. Methodology**

1. Loaded the dataset using Pandas.
2. Cleaned missing values and encoded categorical variables.
3. Split data into training and testing sets.
4. Trained the Logistic Regression model on training data.
5. Evaluated model accuracy using confusion matrix and classification report.
6. Predicted the survival status of new passengers.

**5. Results**

The model predicted survival outcomes with good accuracy.  
It successfully identified key patterns affecting survival (e.g., gender, class, and fare).

**6. Conclusion**

This project shows how machine learning can help analyze and predict outcomes using real-world data. Logistic Regression provided clear insights into the survival factors in the Titanic dataset.