ASSIGNMENT – 1 REPORT

**-Vaishika Agrawal (2201AI42)**

**1. Introduction**

The Titanic dataset is a classic dataset for binary classification, aiming to predict whether a passenger survived or not based on features such as age, sex, ticket class, and fare.  
In this assignment, I applied **Logistic Regression** to perform survival prediction.

**2. Dataset Description**

* **Rows:** 891 passengers
* **Columns:** PassengerId, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked
* **Target variable:** Survived (0 = Did not survive, 1 = Survived)

A screenshot of a computer

AI-generated content may be incorrect.

**3. Steps Performed**

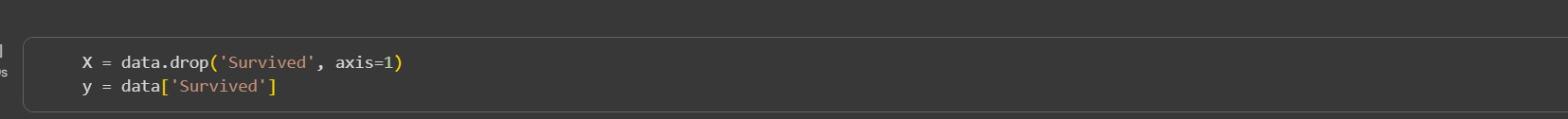
**3.1 Data Preprocessing**

* Handled missing values (‘Age’ filled with median, ‘embarked’ with mode).
* Converted categorical variables (‘Sex’ mapped to 0/1, ‘Embarked’ mapped to 0/1/2).
* Removed irrelevant columns (PassengerId, Name).

A black and white text

AI-generated content may be incorrect.

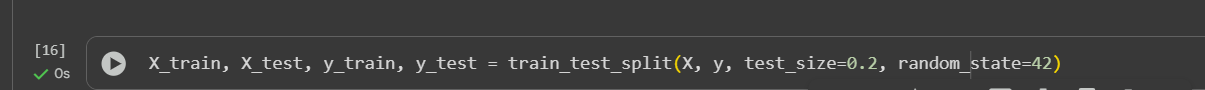
**3.2 Feature Selection**

****

Final features used for model training:  
Pclass, Sex, Age, SibSp, Parch, Fare, Embarked\_Q, Embarked\_S

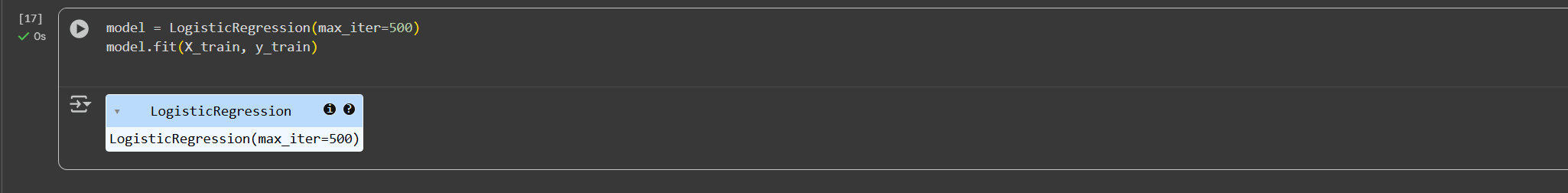
**3.3 Train-Test Split**

* Split data into training (80%) and testing (20%).



**3.4 Model Training**

* Applied **Logistic Regression** with max\_iter=500.



**3.5 Model Evaluation**

* Evaluated with **Accuracy, Confusion Matrix, Precision, Recall, and F1-score**.

**Colab notebook link**: [Assignment-1](https://colab.research.google.com/drive/1pl7EKvjc-VOS9zg4S2w6UkT6oV8Ms02u?usp=sharing)

**Final results**

A screenshot of a computer

AI-generated content may be incorrect.

**Accuracies**

Training Accuracy: 0.8061797752808989

Training Error: 0.1938202247191011

Testing Accuracy: 0.7988826815642458

Testing Error: 0.2011173184357542