PROJECT : **Titanic - Machine Learning from Disaster**

MILESTONE – 02

Domain:

The project is focused on predicting the survival of passengers on the Titanic disaster, a well-known machine learning problem.

Datasets:

The primary dataset used is the Titanic passenger data available on Kaggle, which includes information like passenger class, gender, age, fare, etc.

Literature survey related to the topic :

1. Title: Tackling the Titanic Dataset with Machine Learning

Techniques/Tools: Python, scikit-learn.

Preprocessing : Data loading from Kaggle, data cleaning, feature engineering.

Models Used : Decision Tree, hyperparameter tuning with grid search.

Performance : Achieved validation accuracy of 0.82 after tuning; best parameters included n\_ estimators = 200.

Reference : [Medium Article]

(<https://medium.com/@sanjay_dutta/tackling-the-titanic-dataset-with-machine-learning-5ece2f77b4f2>)

2. Title: Predicting Titanic Survivors Using Machine Learning Techniques

Techniques/Tools : Python, various ML libraries.

Preprocessing: Handling missing values, converting features to numerical representations.

Models Used: Naïve Bayes, Support Vector Machine (SVM), Decision Tree.

Performance : Compared multiple models to assess predictive accuracy, utilizing a dataset of 891 individuals for training.

Reference: [Stanford Project]

(<https://cs229.stanford.edu/proj2012/LamTang-TitanicMachineLearningFromDisaster.pdf>)

3. Title: Titanic Disaster Prediction Based on Machine Learning Algorithms

Techniques/Tools : Python, Decision Tree, Random Forest for model building, feature importance visualization.

Preprocessing : Cleaned data, analyzed age, sex, and ticket class as significant features.

Models : Decision Tree and Random Forest used for survival prediction.

Performance : Decision Tree outperformed Random Forest in accuracy for predicting survival.

Reference : [Titanic Prediction Research]

(<https://www.researchgate.net/publication/370569058_Titanic_Disaster_Prediction_Based_on_Machine_Learning_Algorithms>)

4. Title: Machine Learning Approaches for Titanic Survival Prediction

Techniques/Tools : Python, pandas, scikit-learn.

Preprocessing : Data cleaning, feature normalization by replacing missing values with means.

Models Used : Naïve Bayes, SVM, Decision Tree analysis.

Performance : Evaluation of different techniques to establish a baseline for survival prediction.

Reference : [StanfordProject]

(<https://cs229.stanford.edu/proj2012/LamTang-TitanicMachineLearningFromDisaster.pdf>)

5. Title: Predicting Titanic Survivors Using Machine Learning Techniques

Techniques/Tools: Python, various ML libraries.

Preprocessing: Data cleaning, feature engineering.

Models Used: Multiple ML techniques including Decision Trees.

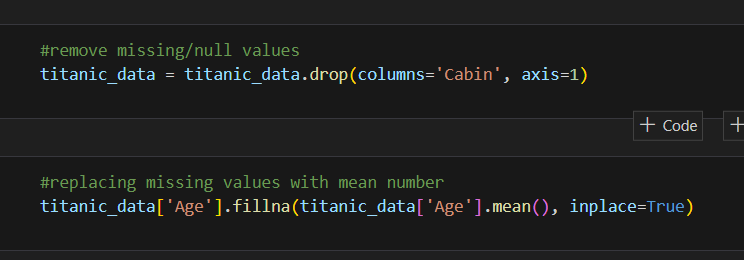
Performance: Utilized a dataset of 891 passengers to predict survival outcomes effectively.

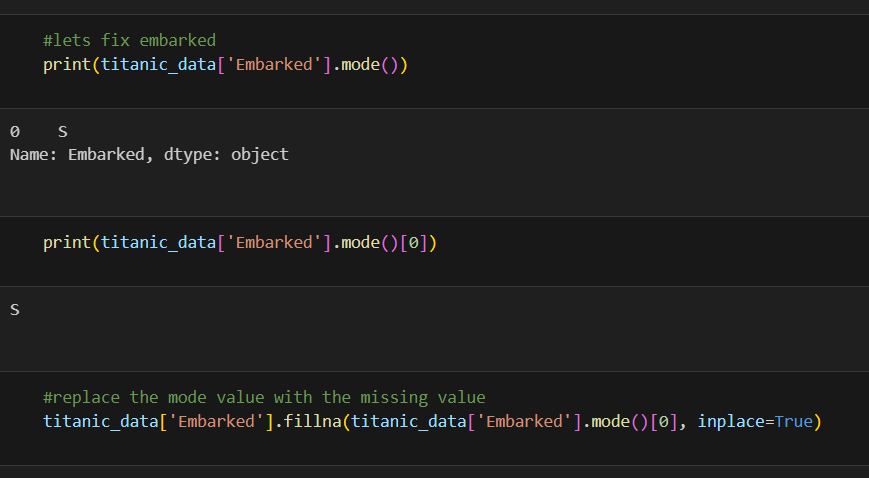
Reference : [Medium Article]

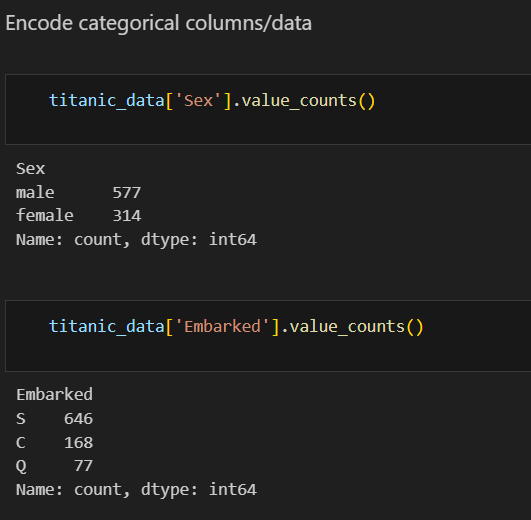
(<https://medium.com/@sanjay_dutta/tackling-the-titanic-dataset-with-machine-learning-5ece2f77b4f2>)

 Data Pre-processing :

* + Handling missing values (e.g. filling in missing ages with the mean)
  + Encoding categorical variables (e.g. converting gender to 0/1)
  + Splitting the data into training and testing sets

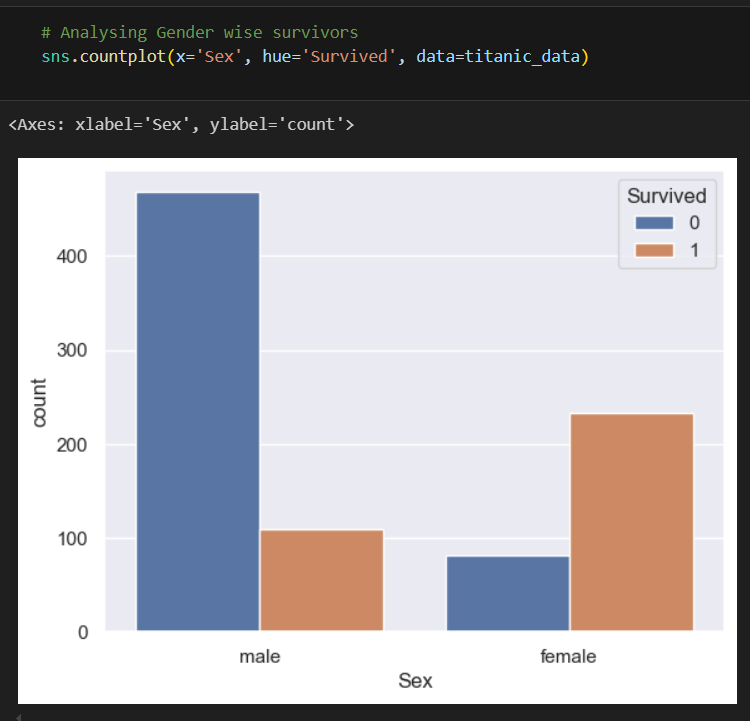


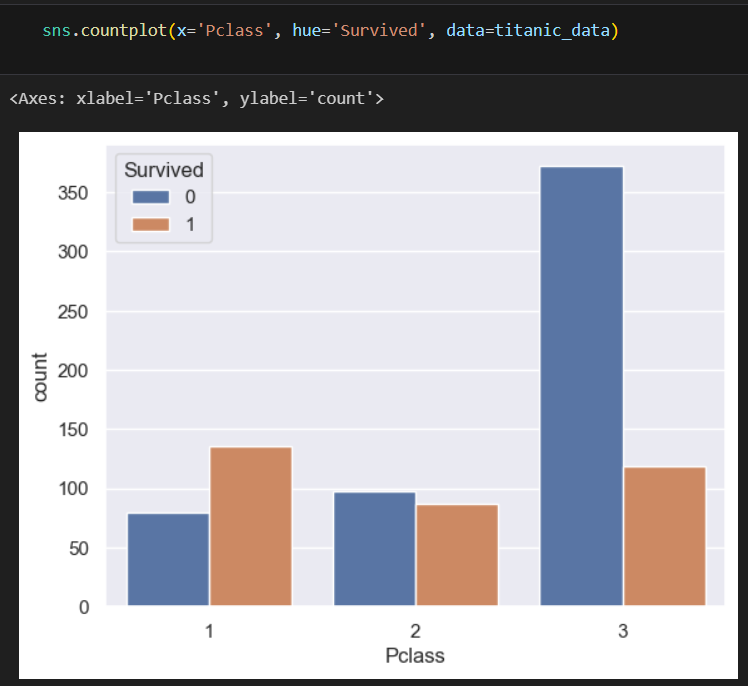




 Data Visualization / EDA using Python Lib :

- Matplotlib is used for displaying detection results in the notebook, helping visualize each detection step-by-step.





Model Creation and Testing :

-Using logistic regression to predict the binary outcome of survival

- Training the model on the training data

-After preprocessing the data and splitting it into X\_ train, X\_ test, y\_ train, and y\_ test, the assistant creates a logistic regression model using scikit-learn

-The assistant then makes predictions on the training data using the trained model:

