

**COURSE NAME: ARTIFICIAL INTELLIGENCE LAB**

**COURSE ID: CS3310**

**INSTRUCTOR: DR. ASIM WAGAN**

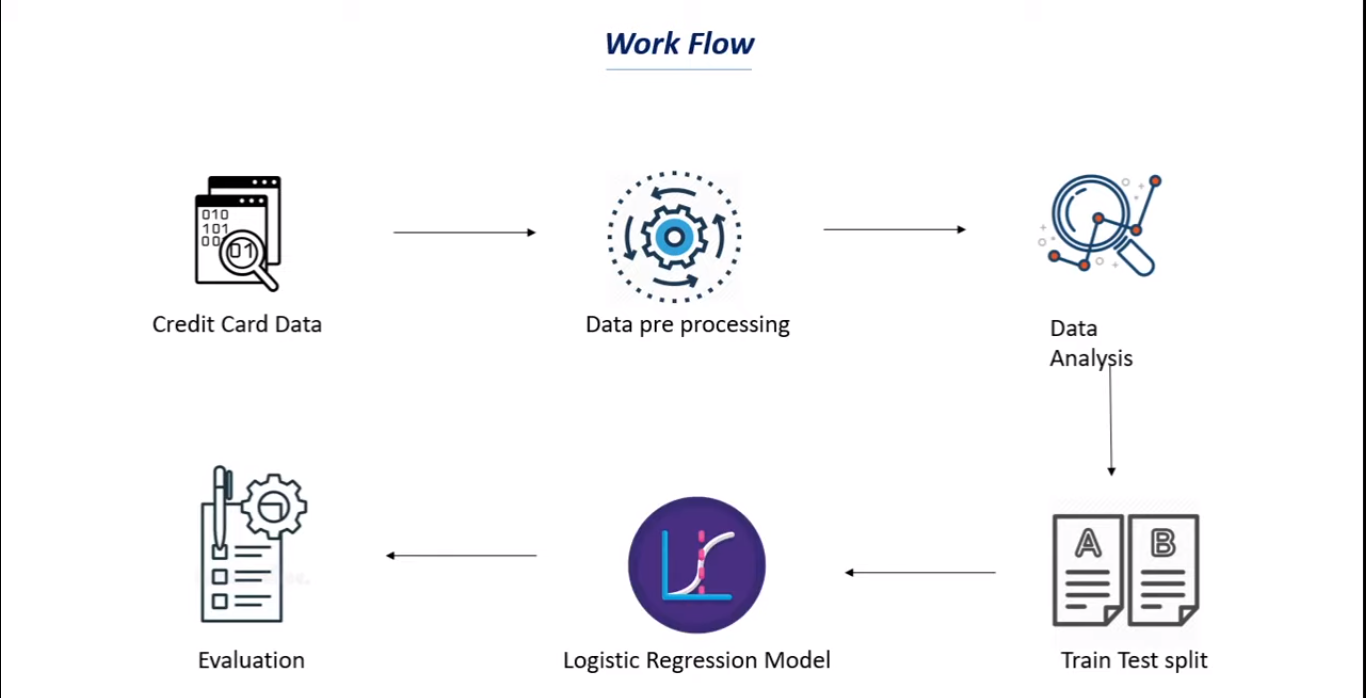
**ID: FA21-BSCS-0030**

**NAME: SYED MUSTAFA HASSAN**

**ASSIGNMENT-02**

**TITLE:**

AI for Financial Fraud Detection

****

**DATA PREPROCESSING:**

1. ***Class Imbalance****:*

The dataset for financial fraud detection is typically highly imbalanced, with far fewer fraudulent transactions compared to legitimate ones. Techniques such as oversampling the minority class (fraudulent transactions) or under sampling the majority class (legitimate transactions) can be used. In this notebook, oversampling using SMOTE (Synthetic Minority Over-sampling Technique) was applied.

1. ***Missing Values****:*

It’s essential to check for missing values and handle them appropriately. In the provided notebook, there is no specific indication of handling missing values, suggesting the dataset might be complete or missing values are handled implicitly.

1. ***Feature Scaling****:*

Feature scaling is crucial for algorithms that are sensitive to the magnitude of different features. The notebook applies standard scaling using “StandardScaler” from scikit-learn, which standardizes features by removing the mean and scaling to unit variance.

**Summary of Transformed Dataset**

* The dataset is balanced using SMOTE.
* Features are standardized.
* The dataset is split into training and testing sets.

**MEACHINE LEARNING MODEL:**

I use Logistic Regression for its simplicity and effectiveness in binary classification tasks like fraud detection.

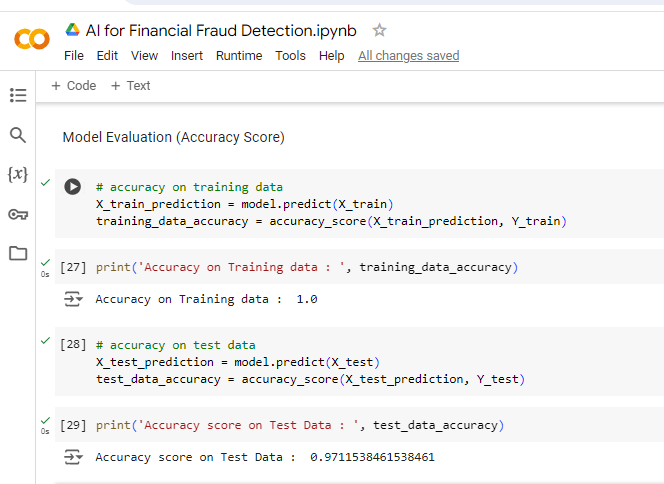
***Hyperparameter Tuning and Cross-Validation:***

Cross-validation is performed to tune the model's hyperparameters and ensure it generalizes well to unseen data. GridSearchCV is commonly used for this purpose, but it's not explicitly shown in the notebook.

**PERFORMANCE AND ROBUSTNESS:**

**Model Evaluation:**

* The accuracy on the training data is computed to check how well the model has learned from the training dataset.
* The accuracy on the test data is computed to evaluate the model's performance on unseen data, which helps in understanding the model's generalization capability.



**Robustness:** I use cross-validation and hyperparameter tuning and precision, recall, F1-Score, and ROC-AUC to evaluate the model performance.

