****

**FACULTY OF TECHNOLOGY**

**COURSE: SOFTWARE DEVELOPMENT**

**CAT 1 SUMMARY**

**UNIT CODE:BSD 3204**

**REG\_NUMBER: 21/08078**

**Summary of What I Have Done & Goals Achieved**

**1. Data Collection & Loading**

I gathered relevant data from **credible sources** that align with the problem identified in **Assignment1**. Specifically, I used a publicly available **credit card fraud detection dataset** from Kaggle and loaded it into a **Jupyter Notebook** using Pandas.

**2. Data Exploration & Understanding**

I performed **data exploration** to understand the dataset's structure, patterns, and key insights:

* Checked for **missing values**, **data types**, and **feature distributions**.
* Analyzed **class distribution** to identify fraud vs. non-fraud transactions.
* Visualized fraud cases using **Seaborn and Matplotlib**.
* Investigated negative values in the dataset and found that they resulted from **PCA (Principal Component Analysis)** transformations.

**3. Data Preprocessing**

To prepare the data for model development, I carried out the following preprocessing steps:

* **Cleaning the data** by handling missing values.
* **Encoding categorical data** (if applicable) to convert non-numeric features into a machine-readable format.
* **Scaling and normalizing features** (e.g., transaction amounts) using **StandardScaler** to improve model performance.
* **Balancing the dataset** using **SMOTE (Synthetic Minority Over-sampling Technique)** to ensure fraud cases were not underrepresented.
* **Splitting the data** into **training and testing sets** to train machine learning models.

**4. Goals Achieved**

a) Gathered data from credible sources aligned with the problem statement.  
b) Successfully loaded and explored the dataset to extract key insights.  
c) Identified patterns and potential issues such as class imbalance and negative values.  
d) Cleaned and preprocessed the data to ensure it's suitable for model training.  
e) Uploaded the **Jupyter Notebook** to GitHub and created a submission document with the repository link.