Title: "Credit Card Transactions Fraud Detection using Machine Learning Algorithms"

Team Number : 007 **Team Name :** Sleuths BUAN 6341 : Thursday Class

Data Source :

For our group project, we are using the Credit Card Transactions Fraud Detection Dataset from Kaggle which is a simulated dataset containing legitimate and fraudulent credit card transactions from 1 January 2019 to 31 December 2020.

Link : https://www.kaggle.com/datasets/kartik2112/fraud-detection?select=fraudTrain.csv

Data Description:

The dataset is divided into training and test dataset and contains transactions of 1000 customers and 800 merchants. Each transaction includes customer details, the merchant and category of purchase, and whether the transaction was fraudulent.

Motivation:

The motivation behind choosing this topic was to put the supervised machine learning techniques learned in this class to practical application.

Project idea:

In order to put knowledge acquired in this class to practical application, we will be using classification supervised learning models to analyze the data and find whether a given credit card transaction is fraud or not. In order to understand the data and demographics we will be working with, we will be first analyzing types of purchases that are most likely to be instances of fraud, amount involved in the fraud per type of merchant and whether older customers are more likely to be victims of credit card fraud using data visualization tools. A discrete classification model such as a logistic regression supervised learning model will then be used to determine whether a transaction is fraud or not where discrete values of 1 and 0 will be used to indicate whether a transaction is fraud or is not fraud respectively. Performance evaluation of the model will be done using classification performance measures such as creating a confusion matrix, calculating precision, recall, accuracy and F1 score. Model comparison will also be done by creating K-nearest neighbors and a decision tree model for classification to see which of the three models performs the best on the given test dataset after training them on the given training dataset. Python code will be used to create all three supervised machine learning algorithms where the models will be first trained using the given training dataset and then their performance will be evaluated using the given test dataset. A conclusion will be made based on the outcome of the model performance which would be the proposed model and machine learning algorithm to determine credit card transactions fraud detection for a given demographics.