**Credit Card Fraud Detection**

Credit card fraud happens when someone (mostly a fraudster) uses the stolen credit card or the credit card information to make unauthorized purchases. The reason I chose this topic is that this problem is real and so many people are getting scammed because of credit card fraudulent activities.

I am going to use Synthetic Financial Dataset for Fraud Detection.

Link for the dataset: <https://www.kaggle.com/datasets/ealaxi/paysim1>.

The dataset is very large with around 6 million entries and 11 columns. Due to computational limitations, I would be making use of 100,000 entries of the dataset.

The dataset is highly imbalanced, i.e., the number of fraudulent activities being very small in number. So, I would be incorporating SMOTE technique to up sample the minority class and down sample the majority class.

I plan to use the Learning Vector Quantization (LVQ) neural network to solve this problem. The reason I chose this algorithm is because LVQ is a classification algorithm, and I wanted to make use of it for a general dataset. I want to compare the performance of LVQ on this dataset with some classical classification algorithms like Decision Trees, KNN, and logistic regression.

I plan to write my own code for the LVQ algorithm and train this dataset using my code. For the classical algorithms, I intend to use sklearn python library. I also plan to use other supporting libraries like numpy, pandas, seaborn, and matplotlib.

I am going to reference our textbook ‘Neural Network Design by Martin T. Hagan’ to design my code for LVQ algorithm. Other than this, I am going to use sklearn documentation for other ML algorithms references.

Since the dataset is highly imbalanced, accuracy will not be an appropriate metric to evaluate the model. So, intend to use precision, recall, F1-score, AUC curve to evaluate the different models.

I plan to do the basic EDA, feature engineering of the dataset in the first week. In the second week, I plan to write the working code for LVQ and compare the results with the other classification models.