**Project 4 Group 4**

**Anomaly Detection to identify Fraudulent Credit Card Transactions**

Amar R, Radu R, Salma A, Samanthi J, Farjana R

**Brief Introduction:**

Fraudulent transactions are a major problem for credit card companies. These transactions are hard to spot as they look a lot like normal and need to be identified among large number of genuine transactions.

We will be employing machine learning methodologies ([Supervised, unsupervised and Neural Network (ANN)] to identify whether the transactions are fraudulent or not using a real world dataset.

**Dataset-**

Anonymized normalised real-world dataset fromCredit Card Fraud Dataset (https://github.com/GuansongPang/ADRepository-Anomaly-detection-datasets)

**Explore the Dataset –** dataframe shape, pyplot, description of dataframe

**Model Test and Baseline Result –**

* Decision Tree (CART)
* k-Nearest Neighbors (KNN)
* Bagged Decision Trees (BAG)
* Random Forest (RF)
* Extra Trees (ET)
* Neural Network (ANN)

**Evaluate Models – using Accuracy,** Recall, Precision, Accuracy, and f1-score will be determined

**Some of the libraries to be used** –

* Scikit-Learn ML Library-

(a) Linear model - LogisticRegression,

(b) Ensemble - RandomForestClassifier,

(c) model\_selection - train\_test\_split & GridSearchCV,

(d) Preprocessing - StandardScaler, normalize, OneHotEncoder

(e) Metrics - classification\_report, precision\_score, recall\_score,

(f) Feature selction – SelectFromModel

(g) Decomposition - PCA,

(h) Manifold – TSNE,

(i) Cluster - KMeans, AgglomerativeClustering

* Scipy.cluster.hierarchy- Dendrogram, Linkage
* TensorFlow - Neural Network

* Keras-tuner – To automate the Neural Network to choose the best model and the best hyperparameters

* Pandas – EDA and data manipulation
* Matplotlib – Plots