**CSE 472 : Machine Learning Sessional**

**Project name:**

# Credit Card Fraud Detection System

**Umama Rahman - 1705040**

**Sumaiya Azad - 1705048**

Level-4, Term-2

Date: March 1, 2023

Department of Computer Science and Engineering

Bangladesh University of Engineering and Technology

# Contents

# 1 Problem Definition

The digital payments market is soaring as the world shifts towards online and card-based payment methods at a faster rate. With such a shift comes the growing issue of cyber security and fraud, which is more common than ever. Hence, enhancing credit card fraud detection is a priority for all banks and financial organizations. In this project, we applied machine learning models to detect fraud in credit card transactions using our dataset.

# 2 Dataset and its analysis

Dataset link : Credit card fraud detection
This dataset presents transactions that occurred in two days, which contains 492 frauds out of 284,807 transactions. The dataset is highly unbalanced, the positive class (frauds) account for 0.172% of all transactions.

# 3 Proposed Solution

To detect fraud transaction we have applied XBNet : An Extremely Boosted Neural Network . This architecture attempts to combine gradient boosted trees with feed-forward neural networks. It updates the weights of each layer in the neural network in two steps:

1. Update weights by gradient descent.
2. Update weights by using feature importance, of a gradient boosted tree in every intermediate layer

In the last layer, this model uses one of the 3 different activation functions

- sigmoid
- softmax
- none

We applied this 3 different activation function in last layer for result generation .

We also applied following machine learning models :

- Decision Tree
- K-Nearest Neighbors
- Passive Aggressive Classifier
- Random Forest
- Stochastic Gradient Descent
- Support Vector Machine

At the end , we applied ensemble method to aggregate the prediction of each base model applied earlier individually .

# 4 Performance report

We computed accuracy , precision , recall , f1- score and area under precision-recall curve for different model

| model | accuracy | precision | recall | f1_scr | auc_pr |
|---|---|---|---|---|---|
| Decision Tree | 0.9993 | 0.9079 | 0.8774 | 0.892 | 0.7858 |
| K-Nearest Neighbors | 0.9994 | 0.9688 | 0.8605 | 0.9076 | 0.8298 |
| Passive Aggressive Classifier | 0.999 | 0.8875 | 0.7822 | 0.8265 | 0.6705 |
| Random Forest | 0.9995 | 0.9746 | 0.8843 | 0.9247 | 0.8593 |
| Stochastic Gradient Descent | 0.9991 | 0.9303 | 0.7958 | 0.8506 | 0.727 |
| Support Vector Machine | 0.9994 | 0.9189 | 0.8876 | 0.9027 | 0.8071 |
| xbnet(sigmoid) | 0.9994 | 0.9294 | 0.8944 | 0.9112 | 0.8244 |
| xbnet(softmax) | 0.9994 | 0.9201 | 0.8944 | 0.9069 | 0.815 |
| xbnet(none) | 0.9992 | 0.8883 | 0.891 | 0.8896 | 0.7799 |
| Ensemble | 0.9995 | 0.9526 | 0.8911 | 0.9196 | 0.8441 |

Figure 1: result

# 5 Result Analysis

Since our dataset is highly imbalanced , so precision and recall is important metrics to evaluate a model .

- Best Accuracy : Random forest model and ensemble method gave the highest accuracy
- Best Precision : Random forest
- Best Recall : XBNet with softmax and sigmoid activation function in last layer

**Random forest** model outperfromed for our dataset .

# 6 Discussion

Neural networks performs well in case of unstructured data but their performance is not up to the mark in case of tabular data . Tree based models is popular for tabular data . XBNet architecture combines these 2 models to create a robust architecture .
XBNet performed well for classification using softmax/sigmoid/none as the activation function in the last layer. We added ReLU activation function in the last layer of neural network for classification of transaction . But the result was not satisfying because precision and recall was around 50% . So we omitted that model . We applied various machine learning model for tabular data among them random forest outperformed other models having precision 97.46% , recall 88.43% and accuracy 99.95% .