Weekly Project Update Report

(Manigandan Ramadasan)

This week, I implemented SVM (Support Vector Machine) which is known for its effectiveness in high-dimensional feature spaces, where the number of features is large. Since our dataset had more features (40). But due to homomorphic encryption, I had to choose a linear kernel which made me go with Linear SVM (which is particularly bad in anomaly detection) with Hinge loss function and Stochastic Gradient Descent optimizer.

I again did pre-processing and reduced the number of features to 30 based on Variance and collinearity. Instead of removing features solely based on spearman correlation coefficient, I formed clusters using Ward's linkage method (Agglomerative/Bottom-up Clustering) with highly correlated features and dropped the first feature of the cluster which helped to me avoid multicollinearity problem. Last time, my feature selection was only based on collinearity and attribute information gain which was achieved through random forest regressor.

After training models, which 30 features there was minor improvement in ANN and Logistic Regression but there was disimprovement in Logistic Regression and CNN. Even though there was improvement in SVM, the final results (80.10% Accuracy and 80.64 F1-Score) were not impressive as it was Linear SVM.

I also tried different parameters of Homomorphic Encryption (N and q values). I observed that as we reduce the N and q values, the decryption precision decreases and the loss increases whereas when I increase the hyperparameters values, the time taken to compute one instance takes more time and more memory (double of the smaller one). The decryption precision also changes but not as much as small parameters. So, I believe that choosing right parameters is crucial for the developed models to get correct decryption.

Next week, I will try comparison algorithm and decision tree as comparison algorithm gives incorrect results after tyring so many parameters. I will also be submitting the first draft of the report on or before Sunday (30/07/2023) for any further changes.