Credit Card Fraud Detection using Machine Learning and Neural Networks

**Abstract – *With the rise of online payment credit cards have had a huge rule in our daily life and economy for the past two decades and it is important task for companies to identify fraud and non-fraud transactions. Multiple methods have been suggested for this problem and they each have their own pros and drawbacks. In this paper we will apply machine learning algorithms and artificial neural networks on the real-world dataset that is taken from Kaggle [1]. Our main goal is to detect all fraud cases. Moreover, we will compare the results of different linear, ensemble, voting and other methods from open-source libraries as well as with methods done in previous papers in this field.***

**Keywords** *Credit Card Fraud, Supervised Machine Learning, Artificial Neural Networks, Imbalanced classification.*

**Introduction**

In the modern world, credit cards are important part of our life as people receive their salary, do their shopping, pay their bills with the help of credit cards. Only in one day there are more than 1 billion credit card transactions are made according to The Nilson Report. For the fraudsters who are eager to steal it can be another opportunity. There is plethora of methods which scammers use. Only in 2018, without even presence of card more than 400 million dollars were stolen. Credit card frauds are most common type of identity theft, occurring 41% of all identity theft reports. Moreover, for the most part police cannot investigate on the credit card fraud due to its international nature.

Credit card fraud detection’s goal is to decide if the given transaction is fraudulent or not according to the previous transaction data. Now the challenge in this type of dataset is that, when you want to train a model while measuring the accuracy the results will be higher than 90% even if the model labels all transactions as non-fraud and the reason for that is because these kinds of datasets are highly imbalanced. For example, in the data that we will use only 492 transactions are fraud and 284315 transaction are not fraud. This means roughly 0.17 percent of all transactions.

In this paper I used multiple supervised learning algorithms, deep learning models and compared their ROC\_AUC score, F1-Score, Precision and Accuracy on the real-world dataset.