

CS698- Machine Learning Project Proposal

Fraud Detection Using Machine Learning

Shubham Koundinya
20718539
skoundinya@uwaterloo.ca

1 Introduction

Frauds detection and anomaly detection are one of the challenges applicable to a wide variety of industries including banking, finance, insurance, and IOT devices. Frauds are nuisances that should be encountered and identified appropriately. Machine Learning can play an important aspect in identifying these anomalies, whether in banking, finance or IOT based devices.

2 Project Description

The project will aim to identify anomalies using Machine Learning. Anomaly detection has many challenges, mainly owing to highly skewed data sets. The percentage of anomalous points are too less, as compared to non - anomalous points. A comparative study which shows the performance of different algorithms, specific to Fraud detection in Credit Cards has been provided in [1]. This paper discusses the performance of k nearest, naive bayes and logistic regression on a highly skewed data set. The paper uses a hybrid approach for sampling this data set. Another paper discusses the challenges of Machine Learning in Fraud Detection Systems, and has mentioned why anomaly and fraud based machine learning systems are difficult to bring into the real industry[2] Salima Omar et.al [3] gave an overview on major machine learning techniques used in anomaly detection.

The purpose of this project will be to identify anomalies in data set, with aim of identifying anomalies. First attempt will be to try to use some of the algorithms and techniques deployed by paper[1] to get some standard results for benchmarking. Further if the time and resources(need of High Computing Resources like GPU) permits, this project will attempt to use Deep Learning and Neural Networks for Fraud Detection and compare them with existing results from paper[1].

The various phases for the project will be:

1. Literature Survey and identify the Datasets to be used.
2. Data Preprocessing and cleansing.
3. Identify if the data is skewed and investigate on literature on how to handle these kind of datasets and implement in the project.
4. Apply Algorithms, hyper tune parameters and obtain the results from various Algorithms.

3 Challenges Expected

Some of the challenges that can be faced are:

1. Computing Resources such as need of GPU.

References

- [1] John O. Awoyemi, Adebayo O. Adetunmbi and Samuel A. Oluwadare, Credit card fraud detection using machine learning techniques: A comparative analysis, 2017 International Conference on Computing Networking and Informatics (ICCNi) (IJDKP), Dec 2017
- [2] Robin Sommer, Vern Paxson, Outside the Closed World: On Using Machine Learning For Network Intrusion Detection (0975-8887), 2010 IEEE Symposium on Security and Privacy, 4, July 2010.
- [3] Salima Omar, Asri Ngadi, and Hamid H. Jebur, Machine Learning Techniques for Anomaly Detection: An Overview, International Journal of Computer Applications (0975 – 8887) (0975-8887), vol 79, no. 2, October 2013.