IMPROVE ROBUSTNESS IN CREDIT CARD FRAUD DETECTION SYSTEM FOR ADVERSARIAL ATTACK USING MACHINE LEARNING

SALITHA DILSHAN HERATH

A dissertation submitted in partial fulfilment of the requirement for Bachelor of Engineering (Honours) degree in Software Engineering

Department of Computing

Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK

2021

Abstract

Many internet facilities have fascinated users to migrate to online banking and grow exponentially in the past few years. Payments done using credit cards are common and are quite crucial in numerous countries, and also simultaneously, frauds are increasing. Doing the transaction using a credit card becomes the norm with small businesses. When considering both legitimate users and fraudsters use mobile transactions. Therefore they become more exposed to large-scale systematic fraud. Credit card fraud has a significant negative effect because the economic impact incurred affects all parties involved.

Fraud can be identified by looking at previous transaction data and examining customer purchasing patterns if any deviation in spending behaviour from established trends may indicate a fraudulent transaction. Banks and credit card firms use several approaches to detect fraud, such as rule-based expert system or machine learning techniques. When considering the machine learning technique, supervised learning approaches are commonly used. These supervised learning approaches for fraud detection, on the other hand, have generally advanced with an assumption on that it is a benign environment. There are no adversaries attempting to get through the fraud detection system.

This research approach for improve robustness in credit card fraud detection, built a framework that can test with adversarial attacks environment and normal environment, also implement the defence mechanism for against to adversarial attacks. This concept using Europe and German credit card dataset were an experiment. These multiple experiments showed proposed approach improved the precession, f1-score, and recall than existing researches. This research considered fraudster's potential reactions to build a robust credit card fraud detection system using testing with different adversarial attacks and implementing defence mechanism for attacks.

Keywords: Supervised learning, Classification algorithms, Adversarial attacks, Adversarial defence