FRAUD DETECTION – A SURVEY FROM A DATA MINING PERSPECTIVE

– Machine Learning

Created by Bîscă Emanuel (emanuel.bisca@stud.ubbcluj.ro) on December 2020

OVERVIEW

Nowadays, with the continuous progress of the financial technologies, the number of people wanting to earn undeserved profits has grown larger. The financial area is forced to improve their systems of detecting frauds, because these faults are costing billions of dollars, annually just in the United States. One answer when dealing with fraud detection is offered by data mining and its algorithms of knowledge discovery through pattern revealing.

After a detailed introduction, the paper contains an investigation on various methods with information about both the specific techniques and the limitations that break it. The results of two different experiments were presented at the end, in order to see the different accuracy of the selected algorithms being compared. The main goal was having an image of the existing techniques used in fraud detection and to improve the performances with ideas on further work.

Five different approaches were presented, with information on both the method itself and its constraints:

- Neural Networks-based fraud detection,
- Decision Tree-based fraud detection,
- Naive Bayes-based fraud detection,
- K - Nearest Neighbors-based fraud detection,
- Support Vector Machines-based fraud detection.

The experiment and its results

It has been showed, at the end of the research paper, that various algorithms are performing in different ways and return different accuracy on the same set of data. These are the conclusions drawn from them:

- The k - Nearest Neighbours model had given better results than all the other models, by all means.
- On the other hand, the Support Vector Machines' models has earned the title of best performance, with accuracy at the highest level and a zero false positive rate (FPR).

MOTIVATION

Why should researchers bother to invest time and energy in such projects?

1. It became a need of the financial area to have engineers trying to develop new techniques of detecting fraud, capable of preventing them from happening, because these frauds are costing billions of dollars, annually just in the United States.

2. There has been a tremendous increase of financial damages on account of electronic payment fraud, with the 7.6 billion dollars in 2010 reaching more than 21.8 billion dollars, five years later; so basically the damages has increased with 300 percents in 2015, related to 2010.

3. The progress of technology is giving us the opportunity and the equipment necessary in conducting attempts of reducing the fraudulent activities.

KEY WORDS

Machine Learning | Data mining | Fraud detection | Knowledge discovery | Pattern Highlight | Neural Networks | Decision Tree | k-Nearest Neighbor | Naive Bayes | Support vectors | Support Vector Machines