Real Time Fraud Detection in Financial Markets

Kaarthik Elangovan

September 2013

Dissertation submitted in partial fulfilment for the degree of
Master of Science in Advanced Computing

Computing Science and Mathematics
University of Stirling

Abstract

Business Problem: The Financial Markets are seeing an increase in fraudulent activity every year and it costs millions of pounds for many institutions. In spite of various governing bodies enforcing strict regulations, fraudulent activities continue to happen. Most of this fraudulent activity are discovered many months or years later when the governing bodies conduct an audit on historical trades. Institutions/brokers who provide various trade services if found guilty of not detecting such activities end up paying huge penalties. The solution is to provide a software system that will detect fraudulent activity in a real time fashion to take action sooner than before and avoid such penalties.

Technology Problem: In today's digital world, there is a burst in data and standalone technologies are reaching their limits to process data and keep up with the speed with which data is being generated.
According to a survey conducted by IBM [18]
'Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone'
There is a need for big data technologies, which use the power of cluster computing using commodity hardware to process huge amounts of data at speeds which are impossible with single nodes. Financial markets have systems which generate huge amounts of data and financial services institutions are starting to migrate to the big data technology.

Objectives: To create a system that will use big data, cluster computing and in memory processing technologies and solve the business problem of detecting fraudulent activity on a real time basis.

Methodology: The system will use machine learning clustering algorithms to identify patterns in historical trades for each trader. The patterns/clusters will be used to determine any change in trading behaviour of each trader by comparing the clusters with the new trading activity of that trader on a real time basis and raise alerts to the business personal for them to determine the validity of alerts. The validation will be used to make the system learn again and make better decisions. The system should evolve on a regular basis.

Achievements: The built system is capable of learning from historical trades, identify patterns in the data. The system will create clusters for each trader and compare new trading activity against the clusters to give a verbose output on the differences real time.