

ONLINE CREDIT CARD FRAUD DETECTION: USING BIG DATA TECHNOLOGIES

SIMI M.J.

Student

MSc computer science

St. Joseph's college (autonomous)

Irinjalakuda, India

Abstract: Now a days everyone uses credit cards for shopping. Credit cards are easy to carry and use. But there are some fraudulent activities in the use of online credit cards. The BIG DATA technologies can be used to detect the online credit card fraud. The term BIG DATA refers to the large amount of data which can be either structured or unstructured or both, which is difficult to store and process using traditional database and software technologies. Mainly credit card companies recognize fraudulent activities by analyzing the patterns and shopping behavior of client. This paper is intended to study with online credit fraud detection using BIG DATA technologies

Index Terms: apache hadoop, map reduce, apache spark, apache flink

I. INTRODUCTION

Online credit card payments become very important method in trading. Credit cards have both advantages and disadvantages. It is easy to carry and use. But it has a disadvantage that by the invention of credit cards the number of fraudulent criminal is increased. To detect the fraudulent activities the BIG DATA provides three major contributions: fraud detection algorithm, model fusion methods and big data technologies. Various BIG DATA techniques are available to detect frauds from online credit card usage by analyzing big data. The BIG DATA fraud detection techniques are: apache hadoop, map reduce, apache spark and apache flink.

The world is becoming digitalized and there are large amount of data to be processed. The problem is how to store, what to store, what kind of data to be deleted etc. The BIG DATA techniques are powerful to analyze large amount of data. The term BIG DATA refers to the large amount of data which can be either structured or unstructured or both, which is difficult to store and process using traditional database and software technologies. An example of big data is Petabyte or Exabyte which consist of billions of data collected from a huge amount of people.

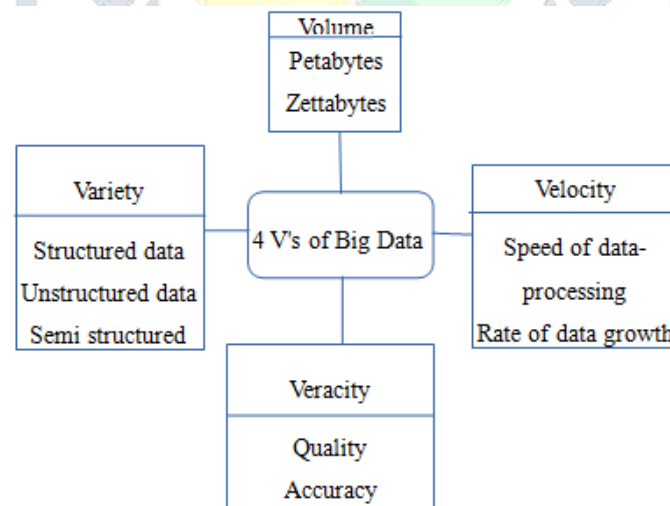


figure 1. big data properties [1]

II. REAL-TIME ONLINE FRAUD DETECTION: HYBRID FRAMEWORK

A. Apache Hadoop

It is an open-source used to store and analyze data from large data set. It has large amount of memory. Its main component is Hadoop Distributed File system (HDFS). It manages data from spreading into another server. Also it can store any type of data.

B. Map Reduce

It is a simple programming method to process data from large datasets in parallel. It take data from large dataset analyse and transform into useful data. It first divides the task into subtasks and these subtasks will be implanted by parallel processing. It has two functions: map and reduce. The map function is used to read inputs. These programs are usually runs in batch mode.

C. *Apache spark*

It is an open source cluster computing. It can be quickly programmed and runs fast. The spark means “resilient distributed datasets”. It is used to process data, to process machine learning algorithm. The HDFS can be used to store data.

D. *Apache flink*

It is an open source framework developed by apache software foundation. Its important part is distributed streaming dataflow engine written in java and scala. It executes dataflow programs parallel or pipelined manner. Pipelined runtime systems enable bulk/batch and stream processing programs execution. It can be used for fault tolerance and it is based on distributed checkpoints. The checkpoint is an automatic, asynchronous snapshot of the state of an application and position in source stream.

III. EXPERIMENT

A. *Test Environment*

We evaluate experiments with 11 nodes and each node holds the following hardware configurations: 2 Intel(R) Xeon(R) E5645 CPU with 6 cores per processor running at 2.40GHz

B. *Detection Algorithm*

We use two unsupervised algorithms (DBSCAN,HMM) in QF and one supervised algorithm in EF.

C. *Streaming Detection Performance*

It has 11-node Storm cluster. Increasing incoming transactions, the delay will increase significantly, due to a limited amount of computing resources.

IV. CONCLUSION

Online fraud detection is very challenging because now there are more transactions. Here, in this paper we design a hybrid frame work solution to solve the problem. We implement the framework with latest BIG DATA technologies which helps to find online fraudulent activities. This hybrid framework can be used in similar application

V. REFERENCE

- [1] Palak Gupta, Nidhi Tyagi, “An Approach towards Big Data –A Review”, International Conference on Computing, Communication and Automation (IEEE), 2015.

