STUDY ON CREDIT CARD FRAUD DETECTION USING DATA MINING TECHNIQUES

DATA MINING (CS-330)

YELLUTLA JEEVANKUMAR
2019UGCS061R

TABLE OF CONTENTS

CREDIT CARD

Overview

Data mining is an activity of extracting some useful knowledge from a large data base, by using any of its techniques. Due to a rapid advancement in the electronic commerce technology, the use of credit cards has increased.

As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of credit card fraud also rising. Data mining techniques could be used to detect the credit card fraud detection. The main goal of this paper is compare the data mining techniques, such as Simple K-means, Hidden Markov Model, Bayesian Network, KNN algorithm and Outlier detection.

CREDIT CARD

INTRODUCTION :

Credit cards are the plastic cards issued by financial companies and institutions such as banks. It is a lot easier to carry payment cards over cash while on the move. We can perform online payments using credit cards. Use of credit cards for online purchases has dramatically increased and it caused an explosion in the credit card fraud. Credit card fraud includes illegal use of card or account information without the knowledge of the owner.

Credit Card Fraud:

Credit card fraud has been divided into two types:

Offline fraud and On-line fraud.

- Offline fraud is committed by using a stolen physical card at call centre or any other place.
- On-line fraud is committed via internet, phone, shopping, web, or in absence of card holder.

CREDIT CARD FRAUD VARIANTS

There are myriad ways to carry out credit card fraudnamely:

ID theft :

When an attacker obtains the personal information of a victim such as date of birth, gender, email id, he can easily get access to a new account using victim's details or even a step further by taking hold of the existing account. Identity theft constitutes 71% of the most common type of fraud.

Fake cards :

Card which is not authorized or not issued by financial institutions is termed as fake cards. Fake cards are developed by skimming the actual data of genuine card which was swiped over an EDC machine. This data is encoded from the magnetic strips and later used to create fake cards.

Stolen/lost cards :

A scenario where a card holder accidentally loses his card or his card has been stolen, if the cardholder fails to report it to the concerned bank there might be chances that the card can be misused by a criminal.

CNP fraud :

Card, not present fraud is a type of fraud where the criminal requires minimal information such as card number and expiry date. In such situation, the card need not be present while making the purchases online.

Clean frauds :

These frauds are not as clean as they sound. The purchases are made with stolen cards and later transactions are modified thus making it find a way around the FDS.

Friendly fraud :

In friendly fraud the actual cardholder himself makes the purchases and pays for the services using "pull" mode of payment with his credit/debit card. Later reports a complaint stating loss of card and claimsfor reimbursement.

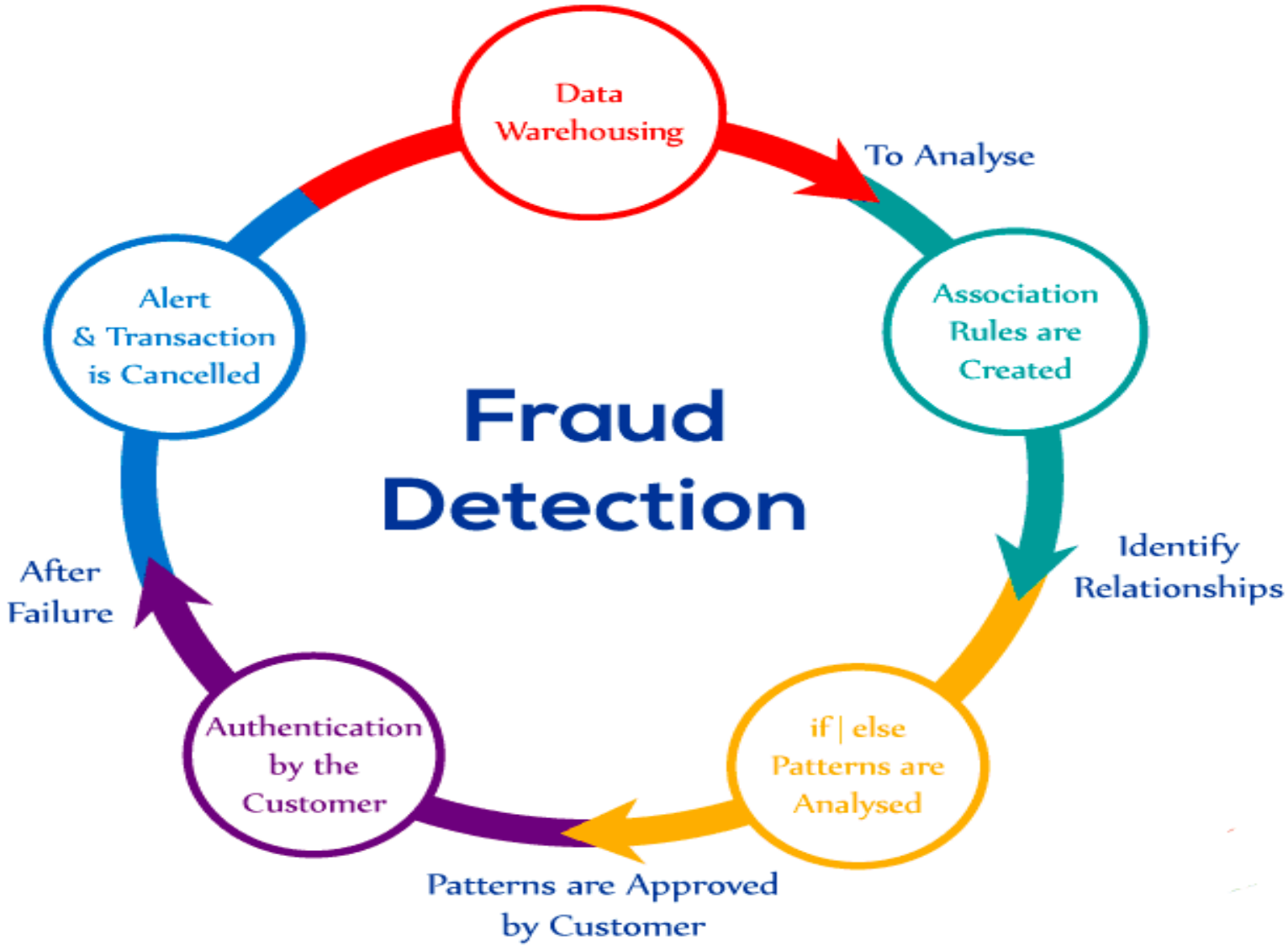
Affiliate fraud :

It is the most widely distributed fraud where either an individual logs into a website and makes purchases using a false account or a program is designed to carry out fraud activities

CREDIT CARD FRAUD DETECTION METHODS

The data mining includes the various techniques and their properties that can be used to detect credit fraud.

- Hidden Markov Model
- Neural Network
- Bayesian Network
- Genetic Algorithm
- K- nearest neighbour algorithm
- Support Vector Machine
- Decision Tree
- Fuzzy Logic Based System


Fraud Detection
Data Warehousing
To Analyse
Association Rules are Created
Identify Relationships
if | else Patterns are Analysed
Patterns are Approved by Customer
Authentication by the Customer
After Failure
Alert & Transaction is Cancelled

Hidden Markov Model:

Hidden Markov Model (HMM) is a statistical Markov modelling which the system being modelled is assumed to be a Markov process with unobserved (i.e. hidden) states. The hidden markov model can be represented as the simplest dynamic Bayesian network

Neural Network:

A neural network is a system of hardware and/or software patterned after the operation of neurons in the human brain. Neural networks -- also called artificial neural networks -- are a variety of deep learning technologies

Genetic Algorithm:

A genetic algorithm (GA) is a search heuristic that mimics the process of natural evolution. This heuristic is routinely used to generate useful solutions to optimization and search problems.

Support Vector Machine:

In machine learning, support vector machines (SVMs, also support vector networks[1]) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Support Vector Machines are based on the concept of decision planes that define decision boundaries.

Decision Tree:

A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node. It is one of the predictive modelling approaches used n statistics, data mining and machine learning.

Fuzzy Logic Based System:

Fuzzy logic deals with reasoning that is approximate rather than fixed and exact. In contrast with traditional logic theory, where binary sets have two valued logic, true or false, fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely false.

LITERATURE REVIEW

In the year of 2016, Maria R. Lepoivre et.al”Credit Card Fraud Detection with Unsupervised Algorithms”.
The objective of this paper is to develop an anti-fraud project by using a combination of two unsupervised algorithms.They using classification for generate a package. Then each package will be grouped by using
clustering concept. The model has been applied to manually implemented data containing on many bank accounts. PCA, SIMPLEKMEANS unsupervised classification scheme has been applied to classify the transactions. It directly classifies the transactions with a good precision and it can detect new fraudulent behaviors.
In the year of 2016, U.Rajeshwari and Dr.B.Sathish Babu” Real-time credit card fraud detection using Streaming Analytics”.

CONCLUSION

Our goal is to analyze different data mining techniques in a way that they help us to detect and predict the credit card fraud.

Analysis presented by different researcher's shows that different data mining techniques. Along with these techniques "Hidden Markov Model" is optimize the best solution for the fraud detection.

.REFERENCES

[1]. Maria R. Lepoivre et al" Credit Card Fraud Detection With Unsupervised Algorithms "Journal of advances in information technology Vol. 7, No. 1, February 2016.

[2]. U.Rajeshwari and Dr.B.Sathish Babu" Real-time credit card fraud detection using Streaming Analytics" 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016

[3]. B.Pushpalatha and C.Willson Joseph" Credit Card Fraud Detection Based on the Transaction by Using Data Mining Techniques" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 2, February 2017

[4]. N.Malini and Dr.M.Pushpa "Analysis on Credit Card Fraud Identification Techniques Based on KNN and Outlier Detection" 3rd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEEICB17).

[5]. VenkataRatnamGanji , " Credit card fraud detection using Anti-k Nearest Neighbor Algorithm", International Journal on Computer Science and Engineering (IJCSE) Vol. 4 ,06 June 2012