Machine Learning Algorithm Using Large Data Analysis & Validate In Data Mining

S.Vigneshwari[1], S.Gayathri[2], J.Swapna[3], Ms.R .Saranya[4]
[1, 2, 3] Dept of Computer Science and Engineering
[4]Assistant professor, Dept of Computer Science and Engineering
[1, 2, 3, 4] P.S.R.Rengasamy College Of Engineering For Women,Sivakasi,TN

Abstract- *The number of online transactions is growing these days to a large number. Today online transaction is going on increasing. The growth of online fraud can be detected. They can be done on many models and methods for credit card fraud prevention and detection. Data Slicing using cluster the data. In this process, we mined the transaction of each data and it finally predicts the fraud detection of data. Our real the online bank transaction and bank relevant data's are generated this large amount of bank data's are maintained at bank documentation and server-side. Existing system the Decision tree process is to be classified and they won't predict any data to be cleared.*

Keywords- data slicing, model validation, model analysis

I. INTRODUCTION

When constructing a credit card fraud detection model, it is very important in the transactional data. Fraud is becoming increasingly and more complex. Online fraud detection. The major problem can occur in the transaction the frequent mining is data mining.

This process we implement fraud detection. It can detect fraud quickly. Fraud detection systems are trained using older transactions to decide about new ones. This training phase is time-consuming is very important. To reduce computation time one can reduce the number of previous transactions processed by minimizing the time window, use less complicated methods, etc. each of which might result in a reduction in accuracy, which means more missed fraud cases and more false alarms. the fraud detection system can occur in the minimum time.

II. LITERATURE SURVEY

Customer churns analysis used for banking sector using data mining techniques.

Oyeniyi, A.O., Adeyemo, A.B., Oyeniyi, A.O., Adeyemo, A.B. - 2015

A data mining model can be used to predict the data. The results can be obtained. The frequent patterns can be used in at least two different ways: first, one can be interested in the individual patterns and their occurrence frequencies; second, one can be interested in the whole collection

SVM can be extracted from data ming: an application to prediction in bank credit cards.

Farquad, M. A., et al. – 2009

Data mining (also called as Knowledge Discovery in Database) is a process that consists of applying data analysis and discovery algorithms that produce a particular enumeration of a pattern (or model) over the data [1]. Data mining has been efficiently used in a wide range of profiling practices, such as manufacturing [2], fraud detection [3]. An increasing number of customers has made the banks conscious of the quality of the services they offer. The problem of customers shifting loyalties from one bank to another has become common. This phenomenon, called 'churn' occurs due to reasons such as availability of latest technology, customer-friendly staff and proximity of geographical location, etc

III. ALGORITHM

SVM:

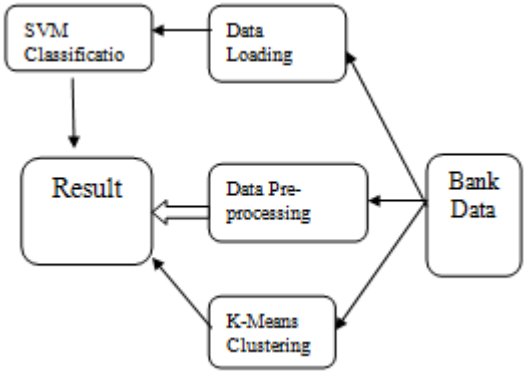
support vector machines are used for classification and regression analysis. SVM training algorithm can be easily classified two categories .it is a binary linear classifier. we display the payable loan and not payable loan.

K-means algorithm:

Machine Learning Algorithms Explained –

K-MeansClustering. **K-Means clustering** is an unsupervised **learning algorithm**. A **cluster** is a group of data points that are grouped due to similarities in their features

IV. ARCHITECTURE DIAGRAM



V. MODULES

Load Data:

In this module, a user browses and selects the dataset. The selected dataset will be viewed and loaded into the database. In this process, we are using bank dataset to analyses and produce the result. We accumulate and load the dataset. In this dataset the attributes are name, age, job, payment, status, etc., using this dataset only we want to find fraud detection.

Preprocessor Data:

A *preprocessor* is a program that processes it's as input data to produce output that is used as input to another program. Pre-processing as suppose dataset loading with after any data placement is to be null or unstructured data or unwanted data is to be removed. The output is should be in the *pre-processed* form of the input data, which is often used by some subsequent programs like compilers. Pre-process is removing null values or unstructured data from loading into certain datasets. Hereafter checking attributes is to be null, then only we check the unwanted data or null data.

Map and Reduce:

After the dataset has been pre-processed then our data will be map-reduce. The framework sorts the outputs of the result of the map, which are then inputted to the reduce tasks. Typically both the input and the output of the MapReduce job are stored in a file-system. The framework is used to takes care of scheduling tasks, monitoring them and re-executes the failed tasks. In this dataset split the repeated frequent data based on the loan.

Clustering the data:

Clustering is the grouping of a particular set of objects based on their characteristics, aggregating them according to their similarities of the object. K-means is also a partitioning technique of clustering that clusters the data set of n objects into k clusters with k known a priori. K-means minimizes is the sum of dissimilarities between points labelled to be in a cluster and a point designated as the centre of that cluster. *A cluster* is a group of objects that belongs to the same class. In other words, similar objects are grouped in one *cluster* and dissimilar objects are grouped in another *cluster* is known as clustering. Here a different set of attributes based load the certain dataset using clustering our data based on some valid attributes. The various algorithms that can differ a cluster to efficiently find them. In this dataset loan based cluster the data using the k-means algorithm.

SVM Classification:

Support Vector Machine is a supervised machine learning **algorithm** which can be used for both classification or regression challenges.SVM training algorithm can be easily classified into two categories. It is a binary linear classifier. we display the payable loan and not payable loan.

Fraud Detection:

After the SVM classification is identified then the process we want to detect the fraud. On transaction to find the deviation from normal behaviour. All these methods have the drawbacks of a typical behaviour-based method, including not tolerating major changes in user behaviour, having high FP, the need for a rather massive amount of transactions for a single user, and the issue of covering all possible. And generate accuracy.

VI. CONCLUSION

We have proposed Slice Finder as a tool for efficiently and accurately finding large, significant, and interpretable slices. The techniques are relevant to model validation in general, but also to model fairness and fraud detection where human interpretability is critical to understand model behaviour. Initially, the dataset has a lot of variation in the data values so the samples are normalized. After pre-processing it the data set is divided into clusters using unsupervised methods like k-means algorithms. Then the clusters are classified using SVM classification and finally detect fraud detection.

REFERENCES

[1] Alejandro Correa Bahnsen, DjamilaAouada, AleksandarStojanovic, jornOttersten, "Feature engineering strategies for credit card fraud detection", Expert Systems with Applications, Elsevier 2016.

[2] Tanmay Kumar Behera and SuvasiniPanigrahi, "Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering & Neural Network", Second International Conference on Advances in Computing and Communication Engineering, IEEE 2015**.**

[3] NedaSoltaniHalvaie and Mohammad Kazem Akbari, "A novel model for credit card fraud detection using Artificial Immune System", Applied Soft Computing 24 Elsevier (2014) 40–49.

[4] Andrea Dal Pozzolo, Olivier Caelen, Yann-A El Le Borgne, Serge Waterschoot.And Gianluca Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective", Expert Systems with Applications 2015.