CREDIT CARD FRAUD DETECTION

Machine Intelligence

BACHELOR OF TECHNOLOGY

Department of Computer Science & Engineering

V Semester Section E

SUBMITTED BY

Batch No: 3

Student name 1: Raunak Singh

SRN: PES2G20CS267

Student name 2: Rudraksh Sharma

SRN: PES2UG20CS287

PES UNIVERSITY

(Established under Karnataka Act No: 16 of 2013)

100 Feet Ring Road, BSK III Stage, Bengaluru-560085

REVIEW OF LITERATURE

1.Analysis of Credit Card Fraud Detection using Machine Learning techniques:

Proceedings of the Seventh International Conference on Communication and Electronics Systems (ICCES 2022) IEEE Xplore Part Number: CFP22AWO-ART; ISBN: 978-1-6654-9634-6.

This paper aims to analyse the frauds in credit card transactions using some machine learning algorithms. For training dataset, credit card datasets are collected. The user is then given credit card inquiries to test the dataset. After analysing a prior dataset, a deep neural network technique is used to classify user's current dataset, which has been provided. Techniques such as, Support Vector Machine (SVM), Deep Neural Network techniques, and Naive Bayes are all used in this research. Based on whale algorithm-optimized BP neural network, this research provides an improved credit card scam detection technology aimed at tackling difficulties of sluggish convergence rate and easy to become stuck in local optimum. In order to improve eight of BP network, we then utilize WOA method to generate an ideal beginning value, and then we use BP network technique to rectify incorrect value. The data set is divided into two sets, one for training purposes and one for testing purposes. Training takes up eighty percent of budget, whereas testing consumes just twenty percent of total. Training set is used to train model. Random Forest method is used to forecast outcome based on this collection of data. For performance parameter Accuracy has been used to analyse the performance of the algorithms used to detect the fraud. The paper detects the credit card transactions and show them in a form of a bar graph where the bar depicts the genuine transactions, how many transactions are fraud and some transactions which may or may not be genuine. Among the three techniques used DNN resulted with the best accuracy of 99% followed by SVM with 97% and NAIVE BAYES with 90%. The research paper used three very important techniques in the world of Machine Learning to detect, but there are certain limitations. The paper did not use any other important algorithms to analyse and detect. There could be some other algorithms which would result in better accuracy than the ones used in the paper. Also, for performance parameter this paper only focuses on Accuracy which is a good measure to analyse any data set but not enough.

2. Ensemble Learning based Credit Card Fraud Detection System:

Shri Govindram Seksaria Institute Of Technology And Science 23, Sir M. Visvesvaraya Marg, Vallabh Nagar, Indore, M.P. 452003 , India.

This paper aims at detecting the credit card fraud detections using Ensemble Learning approach. In this paper ensemble learning technique is employed by parallel applying Decision Tree, Logistic Regression, Naive base classifiers, and then best output is selected through hard voting. The experimental results conclusively proven that accuracy of Ensemble learning with hard voting achieve better accuracy as compared to other classifiers in detecting credit card fraud. The performance of the Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) algorithms applied on credit card dataset and the experimental result shows that applying data under-sampling, with RF outperforms SVM and LR. Decision tree (DT), logistic regression (LR), and random forest (RF) were used and The Bayes minimal risk (BMR) model was also employed. Gradient boosted tree (GBT), and deep learning were compared in and the experimental result proven that Deep learning has the best AUC value for the majority of feature sets. Across all feature sets,

Gradient Boosted Trees obtained the second highest AUC values. The experimental findings shows that the majority voting approach detects credit card fraud instances with a high degree of accuracy. The paper has focused on many different methods and has tried to come up with the best possible technique to detect frauds. For parameters Accuracy, Precision, Recall and F1 Score has been used. The paper has used Ensemble method which has limitation of having class imbalance and methods like SVM,NN,LR,BBS,AIS,GA which has less detection rate. From the Results observed Ensemble learning method through hard voting is high all the evaluation matrices that are respectively 92%, 92%,99%,87% rather than single classifier.

3. A new method for fraud detection in credit cards based on transaction dynamics in subspaces:

Institute of Telecommunications and Multimedia Applications, Universitat Polytechnic de València, Valencia, Spain.

This paper presents a new method for fraud detection in credit cards based on exploiting the dynamics of the card transactions. It hypothesizes different behaviour models in the use of the card between legitimate clients and fraudsters that are registered in the sequential pattern that follows the transactions. The method considers analyses in subspaces defined by two or three variables recorded in the transactions. From these subspaces, several dynamic features, such as transaction velocity and acceleration, are estimated as input vectors for a classification process. Linear and quadratic discriminant analysis and random forest are implemented as single classifiers. All the single classification results obtained for each of the subspaces are late fused to obtain an overall result using alpha integration algorithm. The proposed method was evaluated using a subset of real data with a very low fraud to legitimate transaction ratio. It demonstrates that the temporal dependence of card transactions exploited in different subspaces and fused to give an overall result improves the detection accuracy of fraud detection in credit cards. The dynamic features are processed in each of the subspaces using three single classifiers: linear and quadratic discriminant analysis (LDA and QDA) and random forest (RDF). The proposed method also includes two steps of decision fusion of the results provided by the single classifiers. The first decision fusion consists of combining the results of the classifiers at subspace level, i.e., obtaining a fused result for each of the subspaces. The second decision fusion consists of obtaining an overall result by combining all the fused results obtained for each of the subspaces. It then applies the alpha integration technique to fuse the scores (posterior probabilities) given to transactions by the single classifiers. Area Under the Curve has been used as parameter. Six variables were selected to form different 2D and 3D variable combinations, i.e., subspaces for analysis. Among those variables were, amount, transaction velocity, and country changes of the transaction. It was confirmed that the temporal patterns in the transactions in low dimension subspaces were able to distinguish the behaviour of fraudsters from that of legitimate customers. 3D subspaces yielded a better result than 2D subspaces. This is very complex approach and it uses only AUC as performance parameter.

4. Credit card fraud detection using machine learning algorithms

Jiang, Changjun et al. “Credit Card Fraud Detection: A Novel Approach Using Aggregation Strategy and Feedback Mechanism.” IEEE Internet of Things Journal 5 (2018): 3637-3647.

Multiple Supervised and Semi-Supervised machine learning techniques are used for fraud detection, but our aim is to overcome three main challenges with card frauds related dataset i.e., strong class imbalance, the inclusion of labelled and unlabelled samples, and to increase the ability to process a large number of transactions. Different Supervised machine learning algorithms like Decision Trees, Naive Bayes Classification, Least Squares Regression, Logistic Regression and SVM are used to detect fraudulent transactions in real-time datasets. Performance of Logistic Regression, K-Nearest Neighbour, and Naïve Bayes are analysed on highly skewed credit card fraud data where Research is carried out on examining meta-classifiers and meta-learning approaches in handling highly imbalanced credit card fraud data. Through supervised learning methods can be used there may fail at certain cases of detecting the fraud cases. A hybrid method is developed with a combination of Adaboost and Majority Voting methods.

5. Algorithm based detection

Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain, A Comparative Analysis of Various Credit Card Fraud Detection Techniques, Blue Eyes Intelligence Engineering and Sciences Publications 2019

Naive Bayes classifiers calculate the probability of a sample to be of a certain category, based on prior knowledge. They use the Naïve Bayes Theorem, that assumes that the effect of a certain feature of a sample is independent of the other features. That means that each character of a sample contributes independently to determine the probability of the classification of that sample, outputting the category of the highest probability of the sample. In Bernoulli Naïve Bayes the predictors are boolean variables. The parameters that we use to predict the class variable take up only values yes or no. The basic idea of

Naive Bayes technique is to find the probabilities of classes assigned to texts by using the joint probabilities of words and classes. Boosting is an ensemble modelling technique which attempts to build a strong classifier from the number of weak classifiers. This procedure is continued, and models are added until either the complete training data set is predicted correctly, or the maximum number of models are added. AdaBoost was the first really successful boosting algorithm developed for the purpose of binary classification. Adaboost is short for Adaptive Boosting and is a very popular boosting technique which combines multiple “weak classifiers” into a single “strong classifier”.

6. Comparative analysis of various credit card fraud detection algorithms

Lakshmi S V S S, Selvani Deepthi Kavila, Machine learning for credit card fraud detection system, International Journal of Applied Engineering Research ISSN 2018.

Fraud is any malicious activity that aims to cause financial loss to the other party. As the use of digital money or plastic money even in developing countries is on the rise so is the fraud associated with them. Frauds caused by Credit Cards have costs consumers and banks billions of dollars globally. Even after numerous mechanisms to stop fraud, fraudsters are continuously trying to find new ways and tricks to commit fraud. Thus, in order to stop these frauds, we need a powerful fraud detection system which not only detects the fraud but also detects it before it takes place and in an accurate manner. We need to also make our systems learn from the past committed frauds and make them capable of adapting to future new methods of frauds. In this paper we have introduced the concept of frauds related to credit cards and their various types. We have explained various techniques available for a fraud detection system such as Support Vector Machine (SVM), Artificial Neural Networks (ANN), Bayesian Network, K- Nearest Neighbor (KNN), Hidden Markov Model, Fuzzy Logic Based System and Decision Trees. An extensive review is done on the existing and proposed models for credit card fraud detection and has done a comparative study on these techniques on the basis

of quantitative measurements such as accuracy, detection rate and false alarm rate. The conclusion of our study explains the drawbacks of existing models and provides a better solution in order to overcome them. (Author-Yashvi Jain, Namrata Tiwari, ShripriyaDubey, Sarika Jain)