Credit Card Fraud Detection

INTRODUCTION

Now a day the usage of credit cards has dramatically increased. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. In this paper, we model the sequence of operations in credit card transaction processing using a Decision tree and Deep Neural Network show how it can be used for the detection of frauds. An both algorithms is initially trained with the normal behaviour of a cardholder. If an incoming credit card transaction is not accepted by the trained with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions. We present detailed experimental results to show the effectiveness of our approach and compare it with other techniques available in the literature.

Motivation

The prediction Model will describe you whether to invest in the proposal or not. Here, we choose to minimize the risk for investing, i.e. we aim to minimize investing in proposals for which the loan will not be paid back.

1.3 Scope of The Project

- In this proposed project we designed a protocol or a model to detect the fraud activity in credit card transactions.
- This system is capable of providing most of the essential features required to detect fraudulent and legitimate transactions.
- As technology changes, it becomes difficult to track the behaviour and pattern of fraudulent transactions.
- With the rise of machine learning, artificial intelligence and other relevant fields of information technology, it becomes feasible to automate this process and to save some of the intensive amount of labour that is put into detecting credit card fraud.

LITERATURE SURVEY

The Use of Predictive Analytics Technology to Detect Credit Card Fraud in Canada

(Kosemani Temitayo Hafiz, Dr. Shaun Aghili, Dr. Pavol Zavarsky)

This research paper focuses on the creation of a scorecard from relevant evaluation criteria, features, and capabilities of predictive analytics vendor solutions currently being used to detect credit card fraud. The scorecard provides a side-by-side comparison of five credit card predictive analytics vendor solutions adopted in Canada. From the ensuing research findings, a list of credit card fraud PAT vendor solution challenges, risks, and limitations was outlined.

1) BLAST-SSAHA Hybridization for Credit Card Fraud Detection

(Amlan Kundu, Suvasini Panigrahi, Shamik Sural, Senior Member, IEEE, and Arun K. Majumdar)

In this paper, we propose to use two-stage sequence alignment in which a profile Analyser (PA) first determines the similarity of an incoming sequence of transactions on a given credit card with the genuine cardholder's past spending sequences. The unusual transactions traced by the profile analyser are next passed on to a deviation analyser (DA) for possible alignment with past fraudulent behaviour. The final decision about the nature of a transaction is taken on the basis of the observations by these two analysers. In order to achieve online response time for both PA and DA, we suggest a new approach for combining two sequence alignment algorithms BLAST and SSAHA.

2) Research on Credit Card Fraud Detection Model Based on Distance Sum

(Wen-Fang YU, Na Wang)

Along with increasing credit cards and growing trade volume in China, credit card fraud rises sharply. How to enhance the detection and prevention of credit card fraud becomes the focus of risk control of banks. This paper proposes a credit card fraud detection model using outlier detection based on distance sum according to the infrequency and unconventionality of fraud in credit card transaction data, applying outlier mining into credit card fraud detection. Experiments show that this model is feasible and accurate in detecting credit card fraud.

3) **Fraudulent Detection in Credit Card System Using SVM & Decision Tree**

   (Vijayshree B. Nipane, Poonam S. Kalinge, Dipali Vidhate, Kunal War, Bhagyashree P. Deshpande)

With growing advancement in the electronic commerce field, fraud is spreading all over the world, causing major financial losses. In current scenario, Major cause of financial losses is credit card fraud; it not only affects trades person but also individual clients. Decision tree, Genetic algorithm, Meta learning strategy, neural network, HMM are the presented methods used to detect credit card frauds. In contemplate system for fraudulent detection, artificial intelligence concept of Support Vector Machine (SVM) & decision tree is being used to solve the problem. Thus by implementation of this hybrid approach, financial losses can be reduced to greater extend.

4) **Supervised Machine (SVM) Learning for Credit Card Fraud Detection**

   (Sitaram patel, Sunita Gond)

In this thesis we are proposing the SVM (Support Vector Machine) based method with multiple kernel involvement which also includes several fields of user profile instead of only spending profile. The simulation result shows improvement in TP (true positive), TN (true negative) rate, & also decreases the FP (false positive) & FN (false negative) rate.

5) Detecting Credit Card Fraud by Decision Trees and Support Vector Machines

(Y. Sahin and E. Duman)

In this study, classification models based on decision trees and support vector machines (SVM) are developed and applied on credit card fraud detection problem. This study is one of the firsts to compare the performance of SVM and decision tree methods in credit card fraud detection with a real data set.

6) Machine Learning based Approach to Financial Fraud Detection Process in Mobile Payment System

(Dahee Choi and Kyungho Lee)

Mobile payment fraud is the unauthorized use of mobile transaction through identity theft or credit card stealing to fraudulently obtain money. Mobile payment fraud is the fast growing issue through the emergence of smart phone and online transition services. In the real world, highly accurate process in mobile payment fraud detection is needed since financial fraud causes financial loss. Therefore, our approach proposed the overall process of detecting mobile payment fraud based on machine learning, supervised and unsupervised method to detect fraud and process large amounts of financial data. Moreover, our approach performed sampling process and feature selection process for fast processing with large volumes of transaction data and to achieve high accuracy in mobile payment detection. F-measure and ROC curve are used to validate our proposed mode

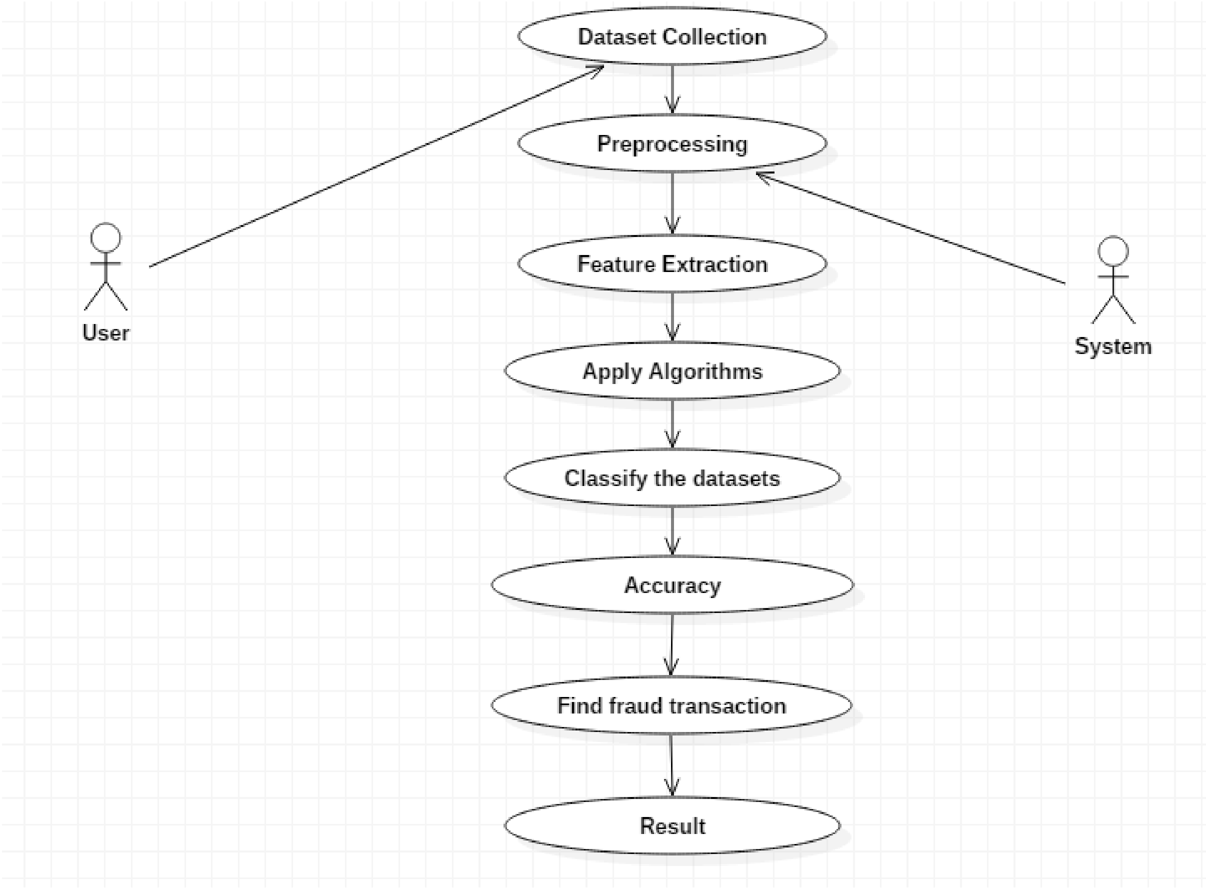
PROBLEM STATEMENT

Billions of dollars of loss are caused every year by the fraudulent credit card transactions. Fraud is old as humanity itself and can take an unlimited variety of different forms. The PwC global economic crime survey of 2017 suggests that approximately 48% of organizations experienced economic crime. Therefore, there is definitely a need to solve the problem of credit card fraud detection. Moreover, the development of new technologies provides additional ways in which criminals may commit fraud. The use of credit cards is prevalent in modern day society and credit card fraud has been kept on growing in recent years. Hugh Financial losses has been fraudulent affects not only merchants and banks, but also individual person who are using the credits. Fraud may also affect the reputation and image of a merchant causing non-financial losses that, though difficult to quantify in the short term, may become visible in the long period. For example, if a cardholder is victim of fraud with a certain company, he may no longer trust their business and choose a competitor.

METHODOLOGY

There are various fraudulent activities detection techniques has implemented in credit card transactions have been kept in researcher minds to methods to develop models based on artificial intelligence , data mining, fuzzy logic and machine learning. Credit card fraud detection is an extremely difficult, but also popular problem to solve. In our proposed system we built the credit card fraud detection using Machine learning. With the advancement of machine learning techniques. Machine learning has been recognized as a successful measure for fraud detection. A great deal of data is transferred during online transaction processes, resulting in a binary result: genuine or fraudulent. Online businesses are able to identify fraudulent transactions accurately because they receive chargebacks on them. Within the sample fraudulent datasets, features are constructed. These are data points such as the age and value of the customer account, as well as the origin of the credit card. There can be hundreds of features and each contributes, to varying extents, towards the fraud probability. Note, the degree in which each feature contributes to the fraud score is not determined by a fraud analyst, but is generated by the artificial intelligence of the machine which is driven by the training set. So, in regards to the card fraud, if the use of cards to commit fraud is proven to be high, the fraud weighting of a transaction that uses a credit card will be equally so. However, if this were to diminish, the contribution level would parallel. Simply put, these models self-learn without explicit programming such as with manual review. Credit card fraud detection using Machine learning is done by deploying the classification and regression algorithms. We use supervised learning algorithm such as Decision tree algorithm to classify the fraud card transaction in online or by offline. Random forest has better efficiency and accuracy than the other machine learning algorithms. Random forest aims to reduce the previously mentioned correlation issue by choosing only a subsample of the feature space at each split.

USE CASE DIAGRAM



REQUIREMENTS ANAYLSIS

SOFTWARE REQUIREMENTS

Python
Anaconda Navigator
Numpy
Pandas
Matplotlib
Sklearn
Seaborn

PYTHON

Python is a high-level, interpreted programming language known for its simplicity, readability, and versatility, widely used in various domains such as web development, data science, artificial intelligence, and more.

ANACONDA NAVIGATOR

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows you to launch applications and easily manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository. It is available for Windows, mac OS and Linux.

Why use Navigator?

In order to run, many scientific packages depend on specific versions of other packages. Data scientists often use multiple versions of many packages, and use multiple environments to separate these different versions.

The command line program conda is both a package manager and an environment manager, to help data scientists ensure that each version of each package has all the dependencies it requires and works correctly.

Navigator is an easy, point-and-click way to work with packages and environments without needing to type conda commands in a terminal window. You can use it to find the packages you want, install them in an environment, run the packages and update them, all inside Navigator.

PANDAS :

Pandas is a Python library used for data manipulation and analysis. It provides powerful tools for reading, writing, and processing structured data like tables or spreadsheets. With Pandas, you can easily clean, transform, and analyze data to extract valuable insights.

NUMPY :

NumPy is a Python library for numerical computing. It provides powerful tools for working with arrays, matrices, and mathematical functions. With NumPy, you can perform fast and efficient numerical operations, making it essential for scientific computing and data analysis.

MATPLOTLIB

Matplotlib is a versatile Python plotting library for creating static, interactive, and publication-quality visualizations.

SEABORN

Seaborn is a Python visualization library based on Matplotlib, specializing in statistical plotting with beautiful and informative visualizations.

SCIKIT-LEARN(SKLEARN)

Scikit-learn (sklearn) is a comprehensive machine learning library, while Seaborn is a high-level statistical data visualization library, both widely used in Python.

Conclusion :

Cost benefit analysis We have tried several models till now with both balanced and imbalanced data. We have noticed most of the models have performed more or less well in terms of ROC score, Precision and Recall. But while picking the best model we should consider few things such as whether we have required infrastructure, resources or computational power to run the model or not. For the models such as Random forest, SVM, XGBoost we require heavy computational resources and eventually to build that infrastructure the cost of deploying the model increases. On the other hand the simpler model such as Logistic regression requires less computational resources, so the cost of building the model is less. We also have to consider that for little change of the ROC score how much monetary loss of gain the bank incur. If the amount if huge then we have to consider building the complex model even though the cost of building the model is high.

We review the existing works on credit card fraud prediction in three different perspectives: datasets, methods, and metrics. Firstly, we present the details about the availability of public datasets and what kinds of details are available in each dataset for predicting credit card fraud. Secondly, we compare and contrast the various predictive modeling methods that have been used in the literature for predicting, and then quantitatively compare their performances in terms of accuracy.

REFERENCES :--

[1] P. Richhariya and P. K. Singh, "Evaluating and emerging payment card fraud challenges and resolution," International Journal of Computer Applications, vol. 107, no. 14, pp. 5 – 10, 2014.

[2] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602–613, 2011.

[3] A.DalPozzolo,O.Caelen,Y.-A.LeBorgne,S.Waterschoot,andG.Bontempi, "Learned lessons in credit card fraud detection from a practitioner perspective," Expert systems with applications, vol. 41, no. 10, pp. 4915–4928, 2014.

[4] C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection: classification of skewed data," ACM SIGKDD explorations newsletter, vol. 6, no. 1, pp. 50–59, 2004.

[5] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 1, pp. 63–77, 2006.

[6] S. Ertekin, J. Huang, and C. L. Giles, "Active learning for class imbalance problem," The 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 823–824, 2007.

[7] M.WasikowskiandX.-w.Chen,"Combatingthesmallsampleclassimbalance problem using feature selection," IEEE Transactions on knowledge and data engineering, vol. 22, no. 10, pp. 1388–1400, 2010. [8] S. Wang and X. Yao, "Multiclass imbalance problems: Analysis and potential solutions," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, no. 4, pp. 1119–1130, 2012.

[9] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," Statistical science, pp. 235–249, 2002.

[10] D. J. Weston, D. J. Hand, N. M. Adams, and C. Whitrow, “Plastic card fraud detection using peer group analysis,” vol. 2, pp. 45–62, 2008.

[11] E. Duman and M. H. Ozcelik, “Detecting credit card fraud by genetic algorithm and scatter search,” Expert Systems with Applications, vol. 38, no. 10, pp. 13057–13063, 2011.

[12] K. Ramakalyani and D. Umadevi, “Fraud Detection of Credit Card Payment System by Genetic Algorithm,” International Journal of Scientific & Engineering Research, vol. 3, no. 7, pp. 1–6, 2012.

[13] P. J. Bentley, J. Kim, G.-h. Jung, and J.-u. Choi, “Fuzzy Darwinian Detection of Credit Card Fraud,” pp. 1–4, 2007.

[14]A.Srivastava,A.Kundu,S.Sural,andS.Member,“CreditCardFraudDetection Using Hidden Markov Model,” IEEE Transactions on Dependable and Secure Computing, vol. 5, no. 1, pp. 37–48, 2008.

[15] S. Esakkiraj and S. Chidambaram, “Predictive Approach for Fraud Detection Using Hidden Markov Model,” International Journal of Engineering Research & Technology, vol. 2, no. 1, pp. 1–7, 2013.

[16] J. S. Mishra, S. Panda, and A. K. Mishra, “A Novel Approach for Credit Card Fraud Detection Targeting the Indian Market,” International Journal of Computer Science, vol. 10, no. 3, pp. 172–179, 2013.

[17] A. Brabazon, J. Cahill, P. Keenan, and D. Walsh, “Identifying Online Credit Card Fraud using Artificial Immune Systems,” IEEE Congress on Evolutionary Computation, pp. 1 – 7, 2010.

[18] N. Wong, P. Ray, G. Stephens, and L. Lewis, “Artificial immune systems for the detection of credit card fraud: an architecture, prototype and preliminary results,” Information systems, vol. 22, pp. 53–76, 2012.

[19] D. Sanchez, M. A. Vila, L. Cerda, and J. M. Serrano, “Association rules applied to credit card fraud detection,” ScienceDirect, vol. 36, pp. 3630– 3640, 2009.

[20] Y. Sahin, S. Bulkan, and E. Duman, “ A cost-sensitive decision tree approach for fraud detection,” Expert Systems with Applications, vol. 40, pp. 5916–5918, 2013.

[21] A. C. Bahnsen, A. Stojanovic, and D. Aouada, “Cost Sensitive Credit Card Fraud Detection using Bayes Minimum Risk,” 12th International Conference on Machine Learning and Applications, pp. 333–338, 2013.

[22] A. E. Pasarica, “Card fraud detection using learning machines,” The Bulletin of the Polytechnic Institute of Jassy, pp. 29 – 45, 2014.

[23] Y. Sahin and E. Duman, “Detecting Credit Card Fraud by Decision Trees and Support Vector Machines,” International Multiconference of Engineers and computer scientists, vol. 1, pp. 442–447, 2011.

[24]V.R.GanjiandS.N.P.Mannem,“Creditcardfrauddetectionusinganti-k nearest neighbor algorithm,” International Journal on Computer Science and Engineering (IJCSE), vol. 4, no. 06, pp. 1035–1039, 2012.

[25] S. Ghosh and D. L. Reilly, “Credit Card Fraud Detection with a NeuralNetwork,” Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, 1994, pp. 621–630, 1994.