CREDIT CARD FRAUD DETECTION

LITERATURE REVIEW

Literature survey of 18 research papers related to credit card fraud detection:

1. Haibo He, Edwardo A. Garcia, “Learning from Imbalanced Data”, an IEEE Transactions on Knowledge and Data Engineering, 2009.

The paper addresses the critical challenge of handling imbalanced datasets in machine learning. Imbalanced data refers to situations where one class significantly outnumbers the other, potentially leading to biased model performance. This paper offers a comprehensive survey of techniques to address this issue, covering both oversampling and undersampling methods, as well as the importance of proper cross-validation techniques for robust model assessment. The authors thoroughly analyze various strategies, providing insights into their strengths, limitations, and suitability for different application domains. This paper serves as a valuable resource for researchers and practitioners seeking to navigate the complexities of imbalanced data and make informed decisions when selecting methods to address class imbalance in their machine learning projects.

2. Tejal Khairnar, Manasi Kale, "Credit Card Fraud Detection Using Machine Learning Techniques: A Survey", Procedia Computer Science, 2018

This survey paper provides an extensive review of the use of machine learning techniques for credit card fraud detection. It begins by emphasising the significance of this problem due to the financial and security implications of credit card fraud. The authors discuss the challenges of fraud detection, including the class imbalance issue where fraudulent transactions are rare compared to legitimate ones.

The paper surveyed various machine learning methods commonly applied in this field. These methods include decision trees, support vector machines, ensemble methods, neural networks, and more. The authors discussed the strengths and weaknesses of each technique, as well as their performance in terms of accuracy, precision, recall, and computational efficiency. Additionally, they provided insights into combining multiple techniques for improved fraud detection and the current state of the field, suggesting directions for future research.

3. Bhavesh Moghariya, Rajprakash Sharma, "Deep Learning for Credit Card Fraud Detection", Procedia Computer Science, 2018

The paper explores the application of deep learning techniques, particularly deep neural networks, in the context of credit card fraud detection. It discusses the advantages of using deep learning for this purpose, such as its ability to automatically learn complex patterns and features from large datasets. The authors present a detailed review of various deep learning models and architectures employed for credit card fraud detection, highlighting their strengths and potential challenges. They also provided insights into the performance and effectiveness of deep learning approaches compared to traditional methods and suggested areas for further research or improvement in this field.

4. Jaideep S, Rahul B, Venugopal K R, et al , "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective",Journal of Information Security, 2019

The paper delves into the various data sources and features used in credit card fraud detection, emphasising the importance of data quality and selection. It discusses data preprocessing techniques to handle issues like missing values, outliers, and class imbalance.

From a technique-oriented perspective, the authors survey a wide range of fraud detection methods, including rule-based approaches, machine learning algorithms, and deep learning techniques. They provide insights into the strengths and weaknesses of each method and discuss the effectiveness of ensemble methods and anomaly detection in this context. The paper concluded by summarizing the current state of credit card fraud detection, identifying areas for further research and development, and emphasizing the need for a multi-faceted approach that combines data and techniques to combat fraud effectively.

5. Alejandro Correa Bahnsen, Aleksandar Stojanovic, Djamila Aouada and Bjorn Ottersten , "Cost Sensitive Credit Card Fraud Detection using Bayes Minimum Risk", Interdisciplinary Centre for Security, Reliability and Trust University of Luxembourg, Luxembourg, 2013

The study begins by testing established algorithms, including logistic regression (LR), C4.5, and random forest (RF), for credit card fraud detection on differently skewed datasets. Under-sampling of legitimate transactions is conducted to balance class distribution. Five databases (S1, S5, S10, S20, and S50) with varying fraud percentages are created, highlighting the significance of an evenly distributed fraud rate. The thresholding optimization technique is then applied to LR, C4.5, and

RF, with RF demonstrating improved results. The Bayes minimum risk classifier is subsequently employed, initially using a cost matrix with fixed false negative (FN) cost, and then with adjusted probabilities to rectify overestimations. The RF-MR A model exhibits the most promising results, saving a significant amount of money while detecting high-value fraud cases. Finally, the study assesses the impact of varying the administrative cost parameter (Ca) on the algorithm's performance.

6. Abhinav Srivastava, Amlan Kundu, Shamik Sural, “Credit Card Fraud Detection Using Hidden Markov Model”, IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING , JANUARY-MARCH 2008

The study conducted large-scale simulation experiments due to the unavailability of real credit card fraud data sets and banks' reluctance to share their data. A simulator was used to generate mixed genuine and fraudulent transactions, allowing the analysis of the system's efficacy. By varying Hidden Markov Model (HMM) design parameters, such as the number of states, sequence length, and threshold value, the study determined an optimal combination. Parameters were selected based on a balanced trade-off between True Positive (TP) and False Positive (FP) rates. It was found that the sequence length of 15 and a threshold of 50% provided optimal performance, with little influence from the number of states. Fraud detection time was observed to increase linearly with sequence length. The results demonstrated the system's effectiveness in accurately identifying fraudulent transactions in a simulated environment.

7. Raghavendra Patidar, Lokesh Sharma, "Credit Card Fraud Detection Using Neural Network", International Journal of Soft Computing and Engineering (IJSCE) , June 2011

This paper investigates credit card fraud techniques and their impact on financial institutions, merchants, and customers. It examines fraud detection methods employed by VISA and MasterCard. Leveraging the robust learning and predictive capabilities of neural networks, the study focuses on utilising Backpropagation Network (BPN) for training. Genetic Algorithm (GA) is utilised to determine crucial parameters, such as weights, network type, and layer configuration, vital for the effective functioning of the neural network. This combined Genetic Algorithm and Neural Network (GANN) approach aims to successfully identify instances of credit card fraud, drawing inspiration from the notion that thorough training enhances the success probability, akin to an inherently talented individual.

8. Jon T. S. Quah and M. Sriganesh, "Real Time Credit Card Fraud Detection using Computational Intelligence", International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007

The approach employs Self Organizing Maps (SOM) in a multi-layered fraud detection system, incorporating initial authentication, risk scoring, and decision-making layers. SOM, an unsupervised neural network, aids in data classification and pattern detection, distinguishing genuine from fraudulent transactions. It serves to create 'legal cardholder' and 'fraudster' profiles, facilitating the identification of customer spending patterns. SOM's ability to map complex data into comprehensible formats allows the detection of new fraud patterns. It acts as a filtering mechanism, segregating transactions for further analysis, thereby reducing processing time and costs. By directly processing raw data, SOM minimizes the need for complex computations and database queries, optimizing system efficiency. Additionally, it enables the classification of data into distinct categories, aiding in the identification of specific transaction types. The subsequent layer, comprising a feed-forward neural network or rule-based system, analyzes potentially fraudulent transactions, learning fraud trends, and assigning risk scores based on predetermined thresholds. This integrated approach enhances fraud detection efficiency and accuracy.

9. Francisca Ogwueleka, "An Overview of Credit Card Fraud Detection Using Data Mining Techniques" , Department of Computer Science, University of Abuja-Nigeria , 2011

The study developed a Credit Card Fraud (CCF) detection model using neural networks, effectively detecting fraudulent transactions while maintaining a low probability of false positives. It highlighted the advantages of neural networks in learning from historical data and improving detection accuracy over time. The system's ability to operate in the background of existing banking software in real-time proved highly effective. The research contributed to the development of a CCF detection system employing four clusters, enhancing its efficiency in handling the Nigerian credit card fraud situation. By employing a unique approach that combined traditional data mining and neural network techniques, the study emphasised the reliability and accuracy of Artificial Neural Networks (ANNs) as a robust tool for operational fraud detection systems.

10. M Deekshith Kumar, Sowmya, Abdul Mubarak, M. S. Dhanush,"Credit Card Fraud Detection Using Bayesian Belief Network", International Journal of Research in Engineering, Science and Management, Volume-3, Issue-7, July-2020.

The proposed credit card fraud detection system integrates Bayesian Networks and the Naïve Bayes Classifier. It operates by training on collected datasets to distinguish between fraudulent and genuine transactions based on specific features, including class, time, and amount. The system employs a block diagram, illustrating its functioning with the mentioned methodologies. The Naïve Bayes Classifier assumes attribute independence, making it suitable for classification tasks. Bayesian Networks, known for their simplicity, are utilized for the same purpose.For fraud detection, the system relies on a predefined structure based on customer complaints, transaction probabilities, and a set ratio of genuine to fraudulent transactions. Data analysis is performed on a dataset of 284,807 transactions, with a minimal 0.2% identified as fraudulent. Principal Component Analysis (PCA) is applied to numerical attributes, while 'TIME' and 'AMOUNT' remain unchanged.The results and discussion section presents the system's performance on two transaction sets, demonstrating its ability to classify transactions accurately based on specific attributes. Overall, the proposed system exhibits effectiveness in handling fraud detection in credit card transactions, utilizing robust methodologies and data analysis techniques for efficient classification.

11. Adithya Tiwari, Akshat Mathur, Dheerendra Singh,"Credit Card Fraud Detection Using Machine Learning: A Survey", International Journal of Computer Applications,2018.

The paper serves as an extensive repository of knowledge, meticulously examining a diverse array of machine learning techniques that are deployed in the context of credit card fraud detection. It critically appraises the nuanced strengths and limitations inherent in each approach, offering a nuanced understanding of their respective applicability and effectiveness in combating sophisticated fraud schemes. Notably, it elucidates the intricate interplay between supervised, unsupervised, and semi-supervised learning paradigms, underscoring their unique contributions to the ever-evolving landscape of fraud detection methodologies. Moreover, the paper presents a comprehensive comparative analysis, shedding light on the performance metrics of various techniques, including their precision, recall, and F1 scores. This in-depth evaluation not only facilitates a holistic comprehension of the relative efficacy of these methods but also lays the groundwork for the identification of potential avenues for future research and development. By consolidating a wealth of empirical evidence and scholarly insights, the paper offers a valuable resource for researchers, practitioners, and stakeholders, fostering an informed and nuanced approach towards the ongoing battle against financial fraud.

12. Xavier Escalante, David M. Bossens, Pedro G. López,"Credit Card Fraud Detection with a Neural-Based Toolkit",Neural Computing and Applications, 2019

The paper introduces a comprehensive toolkit tailored specifically for credit card fraud detection, utilising the sophisticated capabilities of neural networks. It provides an in-depth exploration of the systematic construction of a robust neural-based model, emphasising its adaptability to dynamic fraud patterns and its potential to enhance the efficacy of existing detection systems. Furthermore, the authors highlight the seamless integration of the toolkit into the current fraud detection infrastructure, underscoring its potential to complement and fortify the capabilities of conventional methodologies. The toolkit's performance is rigorously assessed, demonstrating its capacity to discern intricate fraud instances with heightened accuracy and efficiency, thereby underscoring its significance in bolstering security measures within the financial domain.

13. Vaishnavi Nath Dornadula, Geetha S, “Credit Card Fraud Detection using Machine Learning Algorithms”, International Conference on recent trends in advance computing 2019, ICRATC 2019

This paper is about the challenge in applying machine learning methods is the unbalanced problem since the distribution of the transactions is skewed towards the genuine class and sampling methods are used to rebalance the datasets. The SMOTE(Synthetic Minority Oversampling Technique) oversampling method is used in this paper. SMOTE achieves balancing the dataset by generating synthetic examples in the neighbourhood of observed ones to oversample the minority class. It was found that the classifiers were performing better than before applying SMOTE. It was observed that there was an escalation in the precision values of Linear Regression, Decision tree and Random forest.

14. Sushmito Ghosh and Douglas L.Reilly “Credit Card Fraud Detection with a Neural Network”,1994.

A neural network based fraud detection system was trained on a large sample of labelled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month period of time which detected significantly more fraud accounts with significantly fewer false positives. The objectives of the Mellon Bank fraud detection feasibility study was to determine if the use of a wide variety of information characterising the transaction would be helpful in developing improved fraud detection capability. The neural network used in this fraud detection feasibility study is the P-RCE neural network.

The P-RCE is a three-layer, feed-forward network that uses only two training passes through the data set. The first training pass involves a process of prototype cell commitment in which examples from the training set are stored in the weights between the first and second layer cells of the network. A final training pass determines local a posteriori probabilities associated with each of these prototype cells. The P-RCE output layer consists of a single cell that outputs a numeric response that can be considered as a “fraud score”.

A neural network-based fraud detection system has been shown to provide substantial improvements in both accuracy and timeliness of fraud detection. The feasibility study demonstrated that due to its ability to detect fraudulent patterns on credit card accounts, it is possible to achieve a reduction of from 20% to 40% in total fraud losses, at significantly reduced caseload for human review.

15. Emanuel Mineda Carneiro, Luiz Alberto Vieira Dias, Adilson Marques da Cunha, “Cluster Analysis and Artificial Neural Networks”, 12th International Conference on Information Technology- New Generations, 2015

In this paper Cluster Analysis was applied to data normalisation. A Multilayer Perceptron (MLP) Artificial Neural Network classifier was chosen and as it needed the input values in the range from 0 to 1, the data was normalised. Cluster Analysis was used to normalise qualitative data. The Iterative Naive Bayesian Inference Agglomerative Clustering (INBIAC) algorithm was used to cluster the qualitative data as these classifiers are based on a very strong independence assumption. INBIAC is part of the Clustering Engine responsible for normalising the inputs to train a MLP. The results obtained from the use of Artificial Neural Networks and Cluster Analysis on fraud detection has shown that neuronal inputs can be reduced by clustering attributes.

16. Yashvi Jain, Namrata Tiwari, Shripriya Dubey, Sarika Jain,”A Comparative Analysis of Various Credit Card Fraud Detection Techniques”,International Journal of Recent Technology and Engineering, Volume-7, Issue-5S2, January 2019.

The paper has done an extensive review on the existing and proposed models for credit card fraud detection and has done a comparative study on these techniques based on the quantitative measurements. Artificial Neural Network combines the thinking power of the human brain with the computational power of a machine. It uses neurons as deciding sites and edges between the neurons to calculate the contribution of each neuron in the previous layer in the decision and the result at the current neuron.

Decision Tree is a computational tool for classification and prediction which contains internal nodes which is a test on an attribute, branch denotes an outcome of the test and leaf node contains the class label. Fuzzy Logic is used when there are continued truth values. It is a multivalued logic. Classifying the transaction based upon the monetary value associated with it is fuzzification. Rule based deals with drafting the rules based on customer behaviour. In Defuzzification, if a transaction does not comply with the predefined set of rules, it is stopped.

K-Nearest Neighbour is used for classification and regression predictive problems. Distance metrics gives the measure to locate nearest neighbours for the incoming transaction and the Distance rule helps to classify the transaction into class by comparing its features with that of the data in the neighbourhood. A validation error curve is plotted to find the value of K(number of neighbours to compare). Logistic Regression is used for classification problems for predicting binomial and multinomial outcomes, having the goal of estimating the values of parameter’s coefficients using sigmoid function. When a transaction is ongoing, it checks the values of its attributes and decides whether the transactions can proceed or not.

In the comparative analysis , it was found that neural networks had the highest accuracy whereas fuzzy logic and linear regression had the lowest accuracy. The neural network is expensive to train whereas logistic regression is not at all expensive to train. KNN, Fuzzy systems and decision trees are somewhat expensive to train.

17. S. Benson Edwin Raj, A. Annie Portia, “Analysis on Credit Card Fraud Detection Techniques”, International Conference on Computer, Communication and Electrical Technology- ICCET 2011, 18th & 19th March, 2011.

The paper presents a survey of various techniques used in credit card fraud detection mechanisms and evaluates each methodology based on certain design criteria. A hybrid approach of Dempster-Shafer theory and Bayesian learning combines evidence from current as well as past behaviours. In the rule based component , the suspicion level of each incoming transaction based on the extent of its deviation from the good pattern is determined and using the Dempster-Shafer theory, an initial belief is computed and its values are combined to obtain an overall belief. If a transaction is found suspicious, belief is strengthened or weakened based on its similarity with a genuine or fraudulent transaction using Bayesian learning. This hybrid approach has high accuracy and processing speed.

BLAST and SSAHA are sequence alignment algorithms and the hybridization of these two algorithms is used to analyse the spending behaviour of the customers where the alignment process is done using BLAST and SSAHA improves the speed of the alignment process. The profile analyzer determines the similarity of the incoming sequence of transactions with the past spending sequences and the tracked unusual transactions are passed to the deviation analyzer for possible alignment with past fraudulent behaviour. Based on the observations of both the analyzers, the final decision about a transaction is taken. Its performance is good with high accuracy.

The Hidden Markov Model is a double embedded stochastic process . It is initially trained with the normal behaviour of the card holder. If the incoming transaction submitted for verification is found to me malicious then it raises an alarm. If the transaction is not accepted by the Hidden Markov Model with sufficiently high probability, it is considered as malicious.

Fuzzy Darwinian Detection system uses genetic programming to evolve fuzzy logic rules to classify the transactions into suspicious and non-suspicious. The genotypes and phenotypes of the GP system consist of rules which match the incoming with the past sequence. GP evolves a series of variable length fuzzy rules which characterises the difference between classes of data held in a database. It has very high accuracy but produces low false alarms.

The commonly used neural network for pattern recognition is the feed-forward network. Incoming transactions pass from input layer through hidden layer to the output layer known as forward propagation. The training data is compared with the incoming transactions. The suspicious transactions are propagated backward through the neural network and classifies the transactions as suspicious and non-suspicious.

18. Sheo Kumar, Vinit Kumar Gunjan, Mohd Dilshad Ansari, and Rashmi Pathak,"Credit Card Fraud Detection Using Support Vector Machine",2022.

The paper is based on the Support Vector Machine(SVM) based machine learning algorithm that will be used to assess the preprocessed data. It separates the dataset using a hyperplane. SVM is used for classification and pattern analysis.SVM is a gradient boosting application which uses tree based learning algorithms and handles categorical features effectively . This algorithm prevents overfitting, fights against Gradient bias, categorical features support and provides good results with default parameters. The incoming transactions are classified based on which side of the hyperplane it falls into.The results showed that selecting SVM over all techniques could be better in obtaining a greater degree of completeness. SVM Kernel methods showed great performance in sensitivity,accuracy and specificity over traditional techniques.

Table: Report of individual contribution of project team members

Member \ contribution	Literature
Rounak Jain	Ref: 1-6
Shreya Suresh Jindrali	Ref: 7-12
Shikha Reji	Ref: 13-18