A Comprehensive Survey of Machine Learning Techniques for Fraud Detection in Financial Transactions

Dr. S.Prakasam[1] M.Jayalathalakshmi[2],

[1]Associate Professor, Research Scholar[2]Department of Computer Science and Applications, Sri Chandrasekharendra Saraswathi Visa Mahavidyalaya, Kanchipuram, Tamilnadu, 631561, India.

Abstract

Worldwide banking systems face a significant issue with fraudulent transactions, which total $8.1 trillion annually. The issue of fraudulent transactions targeting several financial institutions is a prevalent one, and it's becoming more and clearer that sophisticated technology, such machine learning (ML), is required to stop these kinds of actions. When tactics that rely on fragmented and siloed data, rules-based approaches, or traditional point-solutions are not only expensive but also ineffective against these sophisticated bank frauds, machine learning emerges as the most effective strategy. Financial institutions can employ sophisticated algorithms driven by machine learning to cut down on manual inquiries. Because of the enormous volume of these transactions and the fact that many existing solutions do not concentrate on big data, the suggested model will use the most recent machine learning technology in conjunction with "Apache Spark" to work with big data. In order to flag a transaction as fraudulent or not with a likelihood score, the suggested model will look for patterns in the provided data set. The financial system will then be able to determine the best course of action. Additionally, we will discuss various machine learning algorithms that are used to identify fraudulent transactions and present a comparison analysis of them to demonstrate which works better.

Keywords: Machine Learning, Algorithms, Credit Card, Fraud Detection

1) Introduction

Nowadays, the globe is developing far more quickly than it did in the past, and the main causes of this include widespread trading, internet shopping, online money transfers, etc. The usage of real cards, or credit cards, is growing in significance as a result of these services. The likelihood of fraud is rising along with the use of credit cards. The term "fraud" in credit card transactions describes the unauthorized and unwanted use of an account by an individual who is not the account holder. There are many different methods that fraud can be committed, such as

• Theft of the credit card.

• Takeover of an account.

• Credit card frauds can be perpetrated without the actual card being accessible.

• Identity Theft: This occurs when a scammer gets a victim's credit card number or personal information.

• Seventy-one percent of all frauds are identity thefts.

Because these frauds are difficult to identify with outdated technology, we use cutting edge tools like machine learning (ML). Other methods, which are less successful than machine learning (ML), include restricting the number of fraudulent attempts a consumer can make during a transaction, flagging transactions of significant value, and so on. Because fraudulent transactions are so complicated, there is no set way to stop them, but ML can be used to recognize them. Therefore, by feeding the model with current fraudulent transaction data, we can train the model and, with a high degree of probability, flag the transactions as fraudulent or not. This is made possible by several machine learning algorithms.

Fraud detection can be accomplished in a number of ways, all of which seek to increase detection rates while lowering false alarm rates. Numerous techniques, including the Bayesian method, K-Nearest Neighbor, Support Vector Machine, and others, have been used for fraud detection. Supervised and unsupervised statistical fraud detection methods are the two main categories. In supervised fraud detection techniques, models are computed using samples of legitimate and fraudulent transactions to classify new transactions as fraudulent or legitimate. In unsupervised fraud detection, anomalies or atypical transactions are identified as potentially fraudulent transactions. These two fraud detection algorithms are able to predict the probability that a transaction is fraudulent. This project's goal is to conduct an extensive analysis of several fraud detection algorithms and use to create a system for detecting fraudulent transactions.

Fraud in finance is a deliberate act of deception or misrepresentation intended to result in financial or personal gain, often at the expense of others. It involves manipulating, concealing, or falsifying essential information to trick individuals, businesses, or institutions. Financial fraud is not limited by scale or scope; it can range from small-scale acts committed by individual perpetrators, like credit card fraud, to highly complex schemes orchestrated by organizations, such as securities fraud or corporate embezzlement. Financial fraud manifests in a multitude of forms, each characterized by unique tactics and requiring specialized strategies for identification and mitigation.

Identity theft: This form of fraud involves the unauthorized acquisition of personal data, such as names, social security numbers, and credit card details. Perpetrators use this stolen information for various illicit activities, including opening fraudulent accounts, obtaining loans, or making unauthorized transactions, all without the victim's consent. The repercussions for victims are severe, often including financial loss, credit score damage, and a lengthy.

2) Literature Review

Massimiliano Zanin et al. (2018) [1] In their submission, they consider the potential benefits of using complex networks to enhance the identification of credit card fraud. In particular, based on a recently discovered method, parenclitic networks are used to synthesize complex properties defining card transactions. Then, their utility is evaluated by comparing the rise in classification score obtained with the use of a conventional ANN approach, employing a large dataset of real transactions. They also show that an approach that combines data mining and sophisticated networks may sometimes perform better than a commercial solution.

Abhay Goel et al. (2020) [2] They focused their research on data preparation and analysis, as well as the use of various anomaly detection methods, including the Local Outlier Factor and Isolation Forest algorithm, on the PCA processed Credit Card Transaction data. Jupyter notebooks are also used in the development of this Python software. It has some helpful information about pre-processing datasets and using several anomaly detection methods, including isolation forest methodology and their local outlier factor.

Shailesh S. Dhok et al. (2012) [3] A Hidden Markov Model (HMM) has been suggested as a tool for identifying credit card fraud. The various stages involved in processing credit card transactions are described by the underlying stochastic process of an HMM. While item categories were regarded as HMM states, transaction amount ranges were used as observation symbols. In addition, they suggested a technique for estimating the first estimate of model parameters and the value of observation symbols based on cardholder spending patterns. The ability of the HMM to identify fraudulent activity in incoming transactions was also covered.

Masoumeh Zareapoor et al. (2015) [4] they looked into how well five state-of-the-art methods for predicting credit card fraud performed: Support Vector Machines (SVM), Nave Bayes (NB), KNN, and the Bagging ensemble classifier. With the exception of three of the five approaches that were employed in an experiment, the majority of their explanation is succinct. Three cutting-edge data mining algorithms were examined in their study, one of which was a

bagging ensemble classifier based on the decision three algorithm—a novel approach in the field of credit card fraud detection. A real-world dataset of credit card transactions serves as the foundation for their grading. They found that the bagging classifier based on decision three performs well with this kind of data since it is not dependent on attribute values.

Munira Ansari et al. (2021) [5] Developing a Credit Card Fraud Detection warning system to shield people from credit card online fraud is the main objective of their research. This detection model is derived from the Hidden Markov Model. Preserving the security of our transactions and information is the main objective of a credit card fraud detection system. With this method, fraudulent activity won't be able to continue on a stolen or counterfeit card before the cardholder discovers it. Subsequently, their technique is employed to ascertain the fraudulentness of a novel transaction.

Shailesh S. Dhok et al. (2012) [6] A Hidden Markov Model (HMM) has been suggested as a tool for identifying credit card fraud. The various stages involved in processing credit card transactions are described by the underlying stochastic process of an HMM. While item categories were regarded as HMM states, transaction amount ranges were used as observation symbols. In addition, they suggested a technique for estimating the first estimate of model parameters and the value of observation symbols based on cardholder spending patterns. The ability of the HMM to identify fraudulent activity in incoming transactions was also covered.

Singh, Shashank et al. (2021) [7] Their report also includes the method, pseudocode, explanation of implementation, and experimental results, along with a full description of how machine learning might be applied to enhance fraud detection findings. Although the approach gets above 99.6% precision, it only achieves 28% precision when a tenth of the data set is taken into account. However, the accuracy rises to 33% when the algorithm is given the entire dataset. This high accuracy rate is expected given the significant discrepancy between the quantity of real and legitimate transactions.

Ramya M. and others (2020) [8] they conducted research, experimented, and tweaked parameters to identify the best algorithms to fight four major types of frauds. All of these methods are used in the process. They believe that since the generated machine learning models are only moderately accurate, their focus should be on raising prediction levels to give a more accurate forecast.

Chan, Philip K. et al., 1997 [9] their research investigated several machine learning algorithms and meta-learning techniques using real-world data. In contrast to several documented tests on "standard" data sets, their endeavors in this field aim to replicate the actual environment and issues. The results show that the best True Positive rate and lowest False Positive rate are obtained by classifiers that are trained on a 50/50 distribution of fraud and non-fraud training data.

Zaslavsky, Vladimir, and others (2006) [10] This study suggests a novel approach to credit card fraud detection and transaction monitoring using the self-organizing map method. In an automated system with continuously changing data, it enables the automated Creighton of transaction monitoring rules in a learning process and their ongoing refinement.

The various types of Machine Learning algorithms in supervised learning are discussed:

Linear Regression:

Linear regression is an algorithm used to analyze the relationship between independent input variables and at least one target variable. This kind of regression is used to predict continuous outcomes — variables that can take any numerical outcome. For example, given data on the neighborhood and property, can a model predict the sale value of a home? Linear relationships occur when the data relationship being observed tends to follow a straight line overall — and as such, this model can be used to observe whether a data point is increasing, decreasing, or remaining the same relative to some independent variable, such as time elapsed or position. Machine learning models can be employed to analyze data in order to observe and map linear regressions. Independent variables and target variables can be input into a linear regression machine learning model, and the model will then map the coefficients of the best fit line to the data. In other words, the linear regression models attempt to map a straight line, or a linear relationship, through the dataset. Also, this algorithm is used to predict numerical values, based on a linear relationship between different values.

Neural Networks:

Neural networks are artificial intelligence algorithms that attempt to replicate the way the human brain processes information to understand and intelligently classify data. These neural network learning algorithms are used to recognize patterns in data and speech, translate languages, make financial predictions, and much more through thousands, or sometimes millions, of interconnected processing nodes. Data is “fed-forward” through layers that process and assign weights, before being sent to the next layer of nodes, and so on.Crucially, neural network algorithms are designed to quickly learn from input training data in order to improve the proficiency and efficiency of the network’s algorithms. As such, neural networks serve as key examples of the power and potential of machine learning models. In another way neural

networks simulate the way the human brain works, with a huge number of linked processing nodes. Neural networks are good at recognizing patterns and play an important role in applications including natural language translation, image recognition, speech recognition, and image creation.

Decision Trees:

Decision trees are data structures with nodes that are used to test against some input data. The input data is tested against the leaf nodes down the tree to attempt to produce the correct, desired output. They are easy to visually understand due to their tree-like structure and can be designed to categorize data based on some categorization schema. Decision trees are one method of supervised learning, a field in machine learning that refers to how the predictive machine learning model is devised via the training of a learning algorithm. Decision tree is the one of the best method to analyses fraud during the transactions. Decision trees can be used for both predicting numerical values (regression) and classifying data into categories. Decision trees use a branching sequence of linked decisions that can be represented with a tree diagram. One of the advantages of decision trees is that they are easy to validate and audit, unlike the black box of the neural network.

Random Forest:

Random forest models are capable of classifying data using a variety of decision tree models all at once. Like decision trees, random forests can be used to determine the classification of categorical variables or the regression of continuous variables. These random forest models generate a number of decision trees as specified by the user, forming what is known as an ensemble. Each tree then makes its own prediction based on some input data, and the random forest machine learning algorithm then makes a prediction by combining the predictions of each decision tree in the ensemble. In a random forest, the machine learning algorithm predicts a value or category by combining the results from a number of decision trees. A collective of decision trees is called a Random Forest. To classify a new object based on its attributes, each tree is classified, and the tree “votes” for that class.

3) Future Scope

After reviewing a number of articles, we have concluded that machine learning will be essential in the future for resolving financial issues that are illegal in nature. This review shows us that machine learning may be applied in a variety of ways to address certain financial problems more precisely and quickly. The idea of applying machine learning to prevent debt, or

more precisely, credit card debt in our situation, has been discussed for a while. The concept itself has been published as early as 1997; one of the reviews in this collection [9] addresses this concept by attempting to draw attention to the difficulties and several useful approaches**.**

Everything pertaining to a transaction in a dynamically changing environment essentially exists in some form and can be utilized in a dataset to train various Machine Learning models, since the efficiency of identifying patterns in datasets is increasing. The purpose of this review is to attempt Due to the significant risk involved, companies and large corporations are currently reluctant to share real world data. However, in the future, they might be more willing to do so, either by sharing real world data or by helping to form more accurate pseudo real world data, which will result in more accurate solutions on a larger scale. In the future, people will not only be able to detect fraud through banks and businesses; they will also have access to it in some form to protect their own transactions, which will likely be more beneficial to organizations in their own right. to determine the idea's viability, extent, and potential future applications.

4) Conclusion

The number of frauds committed is rising in tandem with the advancement of technology in many industries. Undoubtedly, credit card fraud is a type of criminal dishonesty that is becoming more and more common. We need to employ a more potent system that can distinguish between a fraudulent transaction and a legitimate one in order to combat/reduce the amount of frauds being performed. Because this method is more accurate and time-efficient, it will assist minimize the amount of human labor and effort required to detect fraudulent transactions.

This is the point in the project where several machine learning techniques are used to create a fraud detection system; the report goes into detail about this. The next step is to install a second monitoring system that monitors various credit card transactions and determines whether a transaction is likely to be fraudulent or safe based on the records it has. Models are updated often with new data and feature extraction, which leads to effective resource utilization. Because of this, the user will be able to choose from a range of security settings, which will help identify fraudulent transactions once the security has been compromised.

References

1. https://arxiv.org/pdf/1706.01953.pdf -Credit Card Fraud Detection through Parenclitic Network Analysis by Massimiliano Zanin, Miguel Romance, Santiago Moral and Regino Criado.

2. http://www.xajzkjdx.cn/gallery/489-april2020.pdf -Credit Card Fraud Detection Using Machine Learning by Abhay Goel, Agrim Mathur, Akanksha Tripathi, Dr. K.K.Aggarwal

3.https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.483.1235&rep=rep1&type=pdf -Credit Card Fraud Detection Using Hidden Markov Model by Shailesh S. Dhok.

4. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.259.16&rep=rep1&type=pdf -Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier by Masoumeh Zareapoora, Pourya Shamsolmoalia.

5.https://www.ijert.org/research/credit-card-fraud-detection-IJERTCONV9IS04018.pdf-Credit Card Fraud Detection by Munira Ansari, Siddhesh Jadhav, Hashim Malik, Zaiyyan Khan.

6. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.483.1235&rep=rep1&type=pdf -Credit Card Fraud Detection Using Hidden Markov Model by Shailesh S. Dhok.

7.https://www.ijert.org/research/credit-card-fraud-detection-IJERTCONV9IS04018.pdf Credit CardFraud Detection by Munira Ansari, Siddhesh Jadhav, Hashim Malik, Zaiyyan Khan.

8.https://www.ijert.org/improved-credit-card-fraud-detection-using-machine-learning – Improved Credit Card Fraud Detection using Machine Learning by M. Ramya, S. Ajith Kumar, K. Anandh Raja.

9.https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.218.7191&rep=rep1&type=pdf -Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results by Salvatore J. Stolfo, David W. Fan, Wenke Lee and Andreas L. Prodromidis

10.https://www.researchgate.net/publication/265746027_Credit_Card_Fraud_Detection_Using_Self- Organizing_Maps -Credit Card Fraud Detection Using Self-Organizing Maps by Vladimir Zaslavsky, Anna Strizhak.

11. Prachi and Narendra Kumar,(2019) " Fraud detection using Machine learning Algorithms on Financial Transaction Data " (ijrpr),4(8):2046-2060.

12. Abolfazi Mehbodniya, et.al.,(2021) "Financial Fraud Detection in HealthCare Using Machine Learning and Deep Learning Techniques " (Hindawi) Seurity and Communication

Networks, vol.2021, Article ID 9293877, 8 pages,2021.

13. Anuruddha Thennakoon,et.al.,(2019) "Real-time Credit Card Fraud Detection Using Machine

Learning"(IEEE), 9th International Conference on Cloud Computing, Data Science & Engineering(conference) 978-1-5386-5933-5/19/$31.00-2019.