Credit Card Fraud Detection System

A Project Work Synopsis

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING IN BIG DATA AND ANALYTICS

Submitted by:

Ankush Singh Bisht	**21BCS6681**
Kunal Wadhwa	**21BCS6631**

Under the supervision of:

Prabjot Singh Bali

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI – 140413, PUNJAB

April, 2024

ABSTRACT

Credit card fraud is a serious criminal offense. It costs individuals and financial institutions billions of dollars annually. According to the reports of the Federal Trade Commission (FTC), a consumer protection agency, the number of theft reports doubled in the last two years. It makes the detection and prevention of fraudulent activities critically important to financial institutions. Machine learning algorithms provide a proactive mechanism to prevent credit card fraud with acceptable accuracy.

In this paper Machine Learning algorithms such as Logistic Regression, Naïve Bayes, Random Forest, K- Nearest Neighbor, Gradient Boosting, Support Vector Machine, and Neural Network algorithms are implemented for detection of fraudulent transactions. A comparative analysis of these algorithms is performed to identify an optimal solution.

In addition to developing a model, the project also aims to explore different features and techniques that can be used to improve the accuracy of the system. The ultimate goal of this project is to provide a tool that can be used by individuals and organizations to identify fake and prevent its spread, ultimately promoting more informed and responsible consumption of news media.

Table of Contents

1 INTRODUCTION

'Fraud' in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. Necessary prevention measures can be taken to stop this abuse and the behaviour of such fraudulent practices can be studied to minimize it and protect against similar occurrences in the future. Fraud detection involves monitoring the activities of populations of users in order to estimate, perceive or avoid objectionable behaviour, which consist of fraud, intrusion, and defaulting.

This is a very relevant problem that demands the attention of communities such as machine learning and data science where the solution to this problem can be automated. This problem is particularly challenging from the perspective of learning, as it characterized by various factors such as class imbalance. The number of valid transactions far outnumber fraudulent ones.

Problem Formulation:-

The problem of credit card fraud detection system is a complex one, and it requires the use of sophisticated algorithms and techniques. The challenge lies in the fact that a small percentage of activity can quickly turn into big dollar losses without the right tools and systems in place.

The aim of a credit card fraud detection system is to identify and prevent fraudulent activities related to credit card transactions. This system uses various techniques, including machine learning algorithms and statistical analysis, to analyze transaction patterns, identify unusual behaviors, and flag potentially fraudulent activities. The primary goal is to minimize losses for credit card companies, merchants, and consumers,

By detecting and preventing unauthorized transactions, stolen cards, and identity theft. Additionally, it helps maintain the integrity of the financial system and prevent criminal activities related to credit card fraud.

Purpose of Making credit card fraud Detection System

The purpose of making a credit card fraud detection system is to detect and prevent fraudulent transactions made using credit cards. Credit card fraud is a type of financial fraud in which someone uses another person's credit card information to make unauthorized purchases or cash withdrawals.

The system can also track and analyze historical transaction data to identify patterns of fraudulent activity and help prevent future occurrences.The benefits of a credit card fraud detection system include:

- **Reducing financial losses:** A credit card fraud detection system can quickly identify fraudulent transactions and prevent financial losses for both credit card issuers and consumers.

- **Enhancing customer trust:** By detecting and preventing credit card fraud, credit card issuers can build trust with their customers and enhance their reputation.

- **Improving operational efficiency**: Automated fraud detection systems can reduce the need for manual reviews and interventions, improving the efficiency of credit card processing and reducing costs.

- **Meeting regulatory requirements**: Many countries have regulations that require credit card issuers to implement fraud detection systems. A credit card fraud detection system can help companies meet these requirements and avoid penalties.

Promoting a More Informed Society

Promoting a more informed society in credit card fraud detection system can help people understand the importance of fraud detection and prevention, as well as the benefits and limitations of such systems. Here are some ways to promote a more informed society:

Education: Providing education and training on credit card fraud detection and prevention can help consumers better understand the risks associated with credit card transactions and how to protect themselves.

Transparency: Credit card issuers and financial institutions can be more transparent about their fraud detection systems, including the technologies used and the process for investigating suspected fraud.

Communication: Credit card issuers can communicate with their customers about suspected fraudulent activity, including providing alerts and notifications when unusual transactions are detected.

Protecting Individuals and Organizations

Protecting individuals and organizations in credit card fraud detection system involves implementing a range of measures that can help prevent, detect, and respond to fraud. Here are some ways to protect individuals and organizations:

Multi-factor authentication: Implementing multi-factor authentication can help prevent unauthorized access to credit card information and reduce the risk of fraud.

Real-time monitoring: Real-time monitoring of credit card transactions can help detect unusual activity and alert the card issuer or the customer to potential fraud.

Data encryption: Data encryption can help protect credit card information from unauthorized access or theft.

Fraud alerts and notifications: Fraud alerts and notifications can help individuals and organizations quickly detect and respond to potential fraudulent activity.

Fraud investigations: Having a process in place for investigating suspected fraud can help organizations quickly identify and respond to fraudulent activity, as well as help law enforcement agencies with their investigations.

Enhancing the Credibility of fraud detection

Enhancing the credibility of credit card fraud detection system is essential to building
trust and confidence among individuals and organizations that rely on these systems to protect their financial information. Here are some ways to enhance the credibility of credit card fraud detection system:

Robust security measures: Credit card fraud detection systems should be equipped with robust security measures that protect against unauthorized access, data breaches, and other forms of cyber-attacks.

Accuracy and reliability: Credit card fraud detection systems should be accurate and reliable, with a low rate of false positives and false negatives. This can be achieved through the use of machine learning algorithms, data analytics, and other advanced technologies.

Improving the Efficiency of News Verification

Improving the efficiency of credit card fraud verification is important to reduce the time and effort required to detect and prevent fraudulent transactions. Here are some ways to improve the efficiency of credit card fraud verification:

Real-time monitoring: Real-time monitoring of credit card transactions can help identify suspicious activity as soon as it occurs, allowing for immediate action to be taken to prevent fraudulent transactions.

Automated fraud detection: Automated fraud detection systems, such as machine learning algorithms, can help detect patterns and anomalies in credit card transactions that may indicate fraudulent activity, allowing for quick and accurate verification.

Multi-factor authentication: Implementing multi-factor authentication can help verify the identity of the credit card holder and reduce the risk of fraudulent transactions.

Mobile alerts: Mobile alerts can be sent to credit card holders to verify their transactions, providing a quick and efficient way to confirm legitimate transactions and prevent fraudulent ones.

Shared databases: Shared databases among financial institutions can help identify and track fraudulent activity across multiple accounts and systems, improving the efficiency of fraud verification and prevention.

Streamlined verification processes: Streamlining the verification process, such as by reducing the number of steps required for verification or using automated systems, can improve efficiency and reduce the time and effort required for verification.

In conclusion credit card fraud detection systems play a critical role in protecting individuals and organizations from financial fraud and cyber crime. The development of these systems has evolved with the advancements in technology and the increasing sophistication of fraudsters. To remain effective, credit card fraud detection systems must continue to evolve and adapt to new and emerging threats.

the development of credit card fraud detection systems is crucial to ensuring the security and integrity of financial transactions. By adopting advanced technologies, maintaining transparency and accountability, and continuously monitoring and improving, financial institutions can build trust and confidence among their customers and contribute to a safer and more secure financial ecosystem.

LITERATURE SURVEY

In this fraud detection system have proposed a system of credit card fraud is an important problem and has an appreciable price for banks and card issuer companies. Therefore, with this huge difficulty in transaction system, banks obtain credit card fraud very seriously, and have very complicated security systems check transactions and identify the frauds as quickly as possible one time it is dedicated. The aim of this survey is to achieve an overall review of different fraud detection methods and selects several innovative methods.

Datasets:

Data sets used in credit card fraud detection systems typically contain a large number of credit card transactions, some of which are fraudulent and others that are legitimate. The data sets are used to train and test machine learning models and other fraud detection algorithms to identify patterns and anomalies that are indicative of fraudulent activity.

Credit Card Fraud Detection Data Set: This data set is publicly available on Kaggle and contains more than 284,000 credit card transactions, of which 492 are fraudulent. The data set includes features such as the transaction amount, time, and type of transaction.

the selection of a data set for credit card fraud detection depends on factors such as the size of the data set, the number of fraudulent transactions, the features included in the data set, and the research question being addressed.

Techniques and past work done

There are various techniques and past work done in credit card fraud detection systems.

Here are some of the common techniques and past work:

Rule-Based Techniques: These techniques use a set of predefined rules to identify fraudulent transactions. These rules are based on thresholds and heuristics such as transaction amount, location, and time of day. While rule-based techniques are easy to implement and interpret, they may not be effective in detecting sophisticated fraud patterns.

Statistical Techniques: These techniques use statistical models to identify patterns and anomalies in credit card transactions. These models include regression analysis, clustering, and decision trees. Statistical techniques are effective in detecting complex fraud patterns but may require significant computational resources.

Machine Learning Techniques: These techniques use machine learning algorithms such as logistic regression, support vector machines, and neural networks to identify patterns and anomalies in credit card transactions. Machine learning techniques can learn and adapt to new fraud patterns, making them effective in detecting previously unseen fraud patterns.

Hybrid Techniques: These techniques combine multiple techniques such as rule-based and machine learning to improve the accuracy of fraud detection.

"Credit Card Fraud Detection using Genetic Algorithm and Neural Network" by Phua et al. (2010) - This paper proposes a credit card fraud detection system that combines genetic algorithms and neural networks to improve the accuracy of fraud detection.

"Credit Card Fraud Detection using Convolutional Neural Networks" by Liu et al. (2019) - This paper proposes a credit card fraud detection system using convolutional neural networks, achieving high accuracy in fraud detection.

Future Research Directions

Future research direction in credit card fraud detection system using machine learning could focus on addressing Some of the challenges they are following as

Explainable AI in Fraud Detection: The use of explainable artificial intelligence (XAI) techniques in credit card fraud detection could help improve the transparency and interpretability of fraud detection models. XAI techniques can provide explanations for the decisions made by a model, making it easier for investigators to understand and interpret the results.

Anomaly Detection with Graph Neural Networks: Graph neural networks (GNNs) can learn representations of data as graphs, allowing them to capture complex relationships and interactions between features. GNNs have shown promise in anomaly detection tasks, and future research could explore the use of GNNs for credit card fraud detection.

Real-Time Fraud Detection: Real-time fraud detection systems can help prevent fraudulent transactions before they occur, improving the efficiency of fraud detection. Future research could explore the use of real-time data streams and machine learning algorithms to detect fraudulent transactions as they occur.

Transfer Learning for Fraud Detection: Transfer learning is a technique that allows models to transfer knowledge learned from one task to another. Future research could explore the use of transfer learning techniques to improve the performance of credit card fraud detection models, particularly in cases where data sets are small or imbalanced.

3. PROBLEM FORMULATION & METHODOLOGY

Fraud detection is a challenging problem. The fact is that fraudulent transactions are rare they represent a very small fraction of activity within an organization. The challenge is that a small percentage of activity can quickly turn into big dollar losses without the right tools and systems in place.The Credit Card Fraud Detection Problem includes modeling past credit card transactions with the knowledge of the ones that turned out to be a fraud. This model is used to identify whether a new transaction is fraudulent or not. Our aim here is to detect 100% of the fraudulent transactions while minimizing the incorrect fraud classifications.

- **Imbalanced data:**The credit card fraud detection data has an imbalanced nature. It means that very small percentages of all credit card transactions are fraudulent. This makes the detection of fraud transactions very difficult and imprecise.

- **Different misclassification importance:** in fraud detection tasks, different misclassification errors have different importance.Misclassification of a normal transaction as fraud is not as harmful as detecting a fraud transaction as normal. Because in the first case the mistake in classification will be identified in further investigations.

- **Overlapping data:** many transactions may be considered fraudulent, while actually they are normal (false positive) and reversely, a fraudulent transaction may also seem to be legitimate (false negative). Hence obtaining a low rate of false positives and false negatives is a key challenge of fraud detection systems.

- **Lack of adaptability:** classification algorithms are usually faced with the problem of detecting new types of normal or fraudulent patterns. The supervised and unsupervised fraud detection systems are inefficient in detecting new patterns of normal and fraud behaviors, respectively.

- **Fraud detection cost**: The system should take into account both the cost of fraudulent behavior that is detected and the cost of preventing it. For example, no revenue is obtained by stopping a fraudulent transaction of a few dollars .

- **Lack of standard metrics**: there is no standard evaluation criterion for assessing and comparing the results of fraud detection systems.

METHODOLOGY USED - The methodology used for the credit card fraud detection system project using machine learning involved several steps. The following section describes the methodology in details:-

DATASET

The dataset for this research work is obtained from Kaggle, and it was generated using Sparkov Data Generation, a GitHub tool created by Brandon Harris. The dataset is a simulated credit card transaction containing legitimate and fraudulent transactions. It covers the credit card of 1000 customers doing transactions with a pool of 800 merchants. 14 The transactions presented by this dataset have 1048575 transactions in total, and the number of fraudulent transactions was recorded to be 6006 out of the total number of transactions. The dataset is highly imbalanced; the positive class (frauds) account for a tiny percentage of about 0.5727 of the complete transactions. The dataset contains 22 features such as" Amount," "Category," "is fraud," and so on, comprising different data types. It also includes both numerical and categorical features. Each transaction recorded per transaction date and time is contained in the feature "trans_date_trans_time" column. The 'Amount' feature column includes the transaction amount carried out, while the last feature in this dataset called "is Fraud" is the response variable that shows whether a transaction is a fraud or not. It takes 1 as a value if it is fraud and 0 if it is not.

DATA PREPROCESSING

Preprocessing data is required before implementing a machine learning algorithm, considering various models produce diverse specifications to the predictors, and data training can affect predictive production. Data preprocessing purposes are to clean and prepare the data to a spot that comprises more concise prejudice, checking for missing values, and more variation. Data contains both numerical and categorical, which means encoding the categorical data is necessary before using them for modeling. Outlier detection and removal was performed. We have the independent variables in the same range by performing feature scaling. To reduce feature skewness, a box-cox transformation was carried out. Resampling method such as undersampling and oversampling was performed on the imbalanced original dataset to avoid any form of bias and overfitting in our training model.

DATA CLEANING

The credit card dataset was imported using the python import command, and the data cleaning process was done. During data cleaning we perform two tasks; 1. Remove null values and missing values, and 2. Handle outliers. The dataset contains 1048575 transactions in total. There were no null values in the dataset. Also, our dataset does not have any missing value. Hence, next we look for outliers in the dataset. Outliers are known as the observations that are numerically distant from the rest of the data. The boxplot technique was adopted to detect the presence of outliers in all the independent features. An outlier is a data point located outside the box plot's whiskers.

ENCODING CATEGORICAL VARIABLES

After cleaning the dataset, we convert any categorical features to a numeric value as most machine learning algorithms perform better with numeric inputs. There are few ways to convert categorical values into numeric values with each approach having its own tradeoffs and impact on the feature set. In the study, we have used One-Hot Encoder to convert the categorical variables to numeric values. For a feature with two categories, the categories 18 are assigned a numeric value of 1 or 0.

FEATURE SCALING

This is another stage of the data preprocessing method used to normalize the range of independent variables within a dataset. Depending on the adopted scaling technique, it is centered around 0 or in the range of 0 and 1. If input variables have tremendous values applicable to the additional input variables, these large values can overlook or skew some machine learning algorithms. We have performed feature scaling using the Robust Scaler technique, also known as robust standardization.

DATASET RE-SAMPLING

Data resampling is a technique of inexpensively using a data sample to improve the accuracy and measure the unpredictability of a population variable. The nested resampling method has been used to carry out dataset resampling. The dataset used for this study was highly imbalanced; that is why we have carried out resampling methods like Undersampling and Oversampling.

UNDERSAMPLING

Since most of the instances in the dataset belong to the majority class, the dataset was under-sampled randomly, by reducing the numbers of instances of the majority class, which means that some essential data instances are not captured for training purposes in the data.

OVERSAMPLING

This method duplicates new or sometimes simulates examples in the minority class. It increases the instances, which makes the training of the model to perform better.

FEATURE CORRELATION AND SELECTION

Each of the features we obtain in the dataset might not be beneficial in building a machine learning model to execute the necessary prediction. Using some of the features might improve the prediction accuracy. So, feature correlation performs a tremendous purpose in creating a better machine learning model. Features with high correlation are more likely to be linearly dependent and have almost the same impact on the dependent variable. Therefore, when two features produce a high correlation, we can drop one of the two features. The heatmap for the correlation of the original dataset, and resampled dataset (both undersampled, and the oversampled) .

Machine learning algorithms

Logistic Regression:

Logistic regression works with sigmoid function because the sigmoid function can be used to classify the output that is dependent feature and it uses the probability for classification of the dependent feature.

This algorithm works well with less amount of data set because of the use of sigmoid function if value the of sigmoid function is greater than 0.5 the output will 1 if the output the sigmoid function is less than 0.5 then the output is considered as the 0. But this sigmoid function is not suitable for deep learning because the if deep learning when we back tracking from the output to input we have to update the weights to minimize the error in weight update. we have to do differentiation of sigmoid activation function in middle layer neuron then results in the value of 0.25 this will affect the accuracy of the module in deep learning.

Decision Tree:

Decision tree can be used for the classification and regression problems working for both is same but some formulas will change. Classification problem uses the entropy and information gain for the building of the decision tree model. entropy tell about how the data is random and information gain tells about how much information we can get from this feature.

Regression problem uses the gini and gini index for the building of the decision tree model. In classification problems the root node is selected by using information gain that the root node t id selected by using is having the high information again and low entropy. In Regression problems the root node is selected by using gini , the feature which is having the less gini is selected as the root here Depth of the tree can be determined
by using hyper parameter optimization, this can be achieved by Using grid search cv algorithm.

Random Forest:

The random forest randomly selects the features that is independent variables and also randomly selects the rows by row sampling and the number of decision tree can be determined by using hyper parameter optimization. For classification problem statement the output is the maximum occurrence outputs from each decision tree models inside the random forest. This is one the widely used machine learning algorithm in real word scenarios and in deployed models. And in most of the Kaggle computation challenges this algorithm is used to solve the problem statement.

NaÃ¯ve Bayes:

NaÃ¯ve Bayes is the machine learning algorithm for classification problem, which work on the property of Bayes theorem. It can be implemented by using features in data set independent feature as input and dependent feature as a output, the same thing what is behind the NaÃ¯ve Bayes theorem is applied here to calculate probability of the dependent feature with respect to independent features.

ANN Model:

Artificial neural networks in deeps learning can be used to replace the machine learning algorithms for better prediction, ANN is having different types of layers such as input layer, number middle layers having activation function for the action of neurons and the output layer having some kind of activation function like sigmoid and weight initialization and reinitialization in backward propagation for reducing the error between actual and predicted values.

Algorithm to choose :

Determining which algorithm is best for a fake news detection system project between decision tree and logistic regression algorithms depends on various factors such as the dataset, features, and performance metrics.

Decision tree algorithm is a good choice for a fake news detection system project due to its ability to handle both numerical and categorical data, and it can handle missing data. Decision tree also offers interpretability and can be useful for exploratory data analysis. However, decision trees can suffer from overfitting, where the tree is too complex and captures noise in the data rather than the underlying pattern. To avoid overfitting, pruning techniques can be used to simplify the tree.

On the other hand, logistic regression is a robust algorithm that can handle large datasets with many input features. It is also better suited for problems where the relationship between the input and output variables is linear. Logistic regression can handle non-linear relationships between the input and output variables to a certain extent. However, logistic regression assumes that the relationship between the input and output variables is linear, which may not always be the case.

In terms of performance, the choice of algorithm will depend on the specific metrics of interest. Some commonly used metrics for classification problems include accuracy, precision, recall, and F1 score. The algorithm that performs best on these metrics would be considered the best choice for the fake news detection system project.

Overall, both decision tree and logistic regression algorithms can be effective for a fake news detection system project, and the choice of algorithm depends on the specific characteristics of the dataset and the performance metrics of interest.

Overall, credit card fraud detection systems are essential for maintaining the integrity of the credit card system, protecting the financial interests of credit card companies and their customers, and maintaining customer trust. As technology continues to evolve, these systems will need to keep pace with new and emerging threats to ensure the continued safety and security of credit card transactions.

The future scope of credit card fraud detection systems is promising as the use of credit cards continues to grow and evolve. Here are some potential areas of development and improvement for these systems:

Artificial Intelligence (AI) and Machine Learning (ML): As AI and ML technologies continue to advance, credit card fraud detection systems can leverage these technologies to improve the accuracy and efficiency of fraud detection. AI and ML can help detect complex patterns and anomalies in data, and even learn from past fraud cases to prevent similar incidents in the future.

Blockchain Technology: The decentralized and transparent nature of blockchain technology can potentially provide a more secure and efficient way to process credit card transactions. Credit card fraud detection systems can leverage blockchain technology to provide real-time validation and authentication of transactions, reducing the risk of fraud.

Big Data Analytics: With the increasing amount of data generated by credit card transactions, big data analytics can help credit card fraud detection systems to process and analyze large amounts of data quickly and accurately. This can lead to more effective and efficient fraud detection, reducing false positives and false negatives.

Advanced Biometric Authentication: Biometric authentication technologies such as facial recognition, fingerprint scanning, and voice recognition can potentially provide a more secure way to authenticate credit card transactions. Credit card fraud detection systems can leverage biometric authentication to provide an extra layer of security and reduce the risk of identity theft.

Real-time Fraud Detection: Real-time fraud detection can provide an immediate response to potential fraudulent transactions, reducing the risk of financial loss.As technology continues to improve, credit card fraud detection systems can leverage real-time data processing and analysis to provide more accurate and efficient fraud detection.

4. OBJECTIVES :

The objective of a credit card fraud detection system is to identify and prevent fraudulent transactions made using credit cards. The system uses various techniques such as machine learning, data mining, and pattern recognition to detect suspicious activities and flag them for further investigation.

Higher accuracy of fraud detection. Compared to rule-based solutions, machine learning tools have higher precision and return more relevant results as they consider multiple additional factors. This is because ML technologies can consider many more data points, including the tiniest details of behavior patterns associated with a particular account.

2. Less manual work needed for additional verification. Enhanced accuracy leads reduces the burden on analysts. “People are unable to check all transactions manually, even if we are talking about a small bank,” Alexander Konduforov, data science competence leader at AltexSoft, explains. “ML-driven systems filter out, roughly speaking, 99.9 percent of normal patterns leaving only 0.1 percent of events to be verified by experts.”

3. Fewer false declines. False declines or false positives happen when a system identifies a legitimate transaction as suspicious and wrongly cancels it.

4. Ability to identify new patterns and adapt to changes. Unlike rule-based systems, ML algorithms are aligned with a constantly changing environment and financial conditions. They enable analysts to identify new suspicious patterns and create new rules to prevent new types of scams.

CONCLUSION & FUTURE SCOPE :

In conclusion, credit card fraud detection systems play a crucial role in safeguarding the financial interests of credit card companies and their customers. With the increasing use of credit cards for online and offline transactions, the risk of fraudulent activities has also increased. Fraudulent activities such as unauthorized transactions, identity theft, and account takeover can cause financial losses and damage the reputation of credit card companies.

Credit card fraud detection systems use advanced techniques such as machine learning, data mining, and pattern recognition to detect and prevent fraudulent activities. By analyzing historical and real-time data, these systems can identify patterns and anomalies that indicate fraudulent transactions. Once a suspicious activity is identified, the system can flag it for further investigation or block the transaction in real-time.

6.REFERENCES

[1] Joshi, Aruna & Shirol, Vikram & Jogar, Shrikanth & Naik, Pavankumar & Yaligar, Annapoorna. (2020). Credit Card Fraud Detection Using Machine Learning Techniques. International Journal of Scientific Research in Computer Science, Engineering, and Information Technology. 436-442. 10.32628/CSEIT2063114.

[2] KDnuggets. (2017). XGBoost, a Top Machine Learning Method. https://www.kdnuggets.com/2017/10/xgboost-top-machine-learning-methodkaggle-explained.html

[3] Lebichot, Bertrand & Braun, Fabian & Caelen, Olivier & Saerens, Marco. (2017). A graph- based, semi-supervised, credit card fraud detection system. 693. 721-733. 10.1007/978-3-319-50901-3_57.

[4] Meng, Cuizhu & Zhou, Li & Liu, Bisong. (2020). A Case Study in Credit Fraud Detection with SMOTE and XGBoost. Journal of Physics: Conference Series. 1601. 052016. 10.1088/1742-6596/1601/5/052016.

[5]Mohari, Ankit & Dowerah, Joyeeta & Das, Kashyavee & Koucher, Faiyaz & Bora, Dibya & Bora. (2021). A COMPARATIVE STUDY ON CLASSIFICATION ALGORITHMS FOR CREDIT CARD FRAUD DETECTION.

[6]More, Rashmi & Awati, Chetan & Shirgave, Suresh & Deshmukh, Rashmi & Patil, Sonam. (2021). Credit Card Fraud Detection Using Supervised Learning Approach. International Journal of Scientific & Technology Research. 9. 216-219.

[7]Niklas Donges. (2021). A complete guide to the Random Forest algorithm. https://builtin.com/data-science/random-forest-algorithm

[8]Paige Schaffer. (2018). Reducing the Impact of New Account and Credit Card Fraud on Financial Institution. https://www.cpomagazine.com/cybersecurity/reducing-the-impact-of-new-account-and-credit-card-fraud-on-financialinstitutions