Text

Description automatically generated

**< A Python-based machine learning algorithm approach to improved credit card fraud detection using a Logistic regression model.>**

<Bhoomi Kaushal Sharma>

<B00733638>

Thesis Project for the partial fulfilment

of the requirements for the Master Degree

in Information Technology/E-Business Management / Advanced Computer Systems Development (delete one)

University of the West of Scotland

School of Computing

<22/12/2022>

Text

Description automatically generated

**DECLARATION OF ORIGINALITY**

I declare that this is an original study based on my own

work and that I have not submitted it for any other

course or degree.

A close up of a note

Description automatically generated with low confidence

Signature \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Text

Description automatically generated**

Library Form to Accompany MSc Project

To be completed in full

|  |
| --- |
| **Surname: Sharma** |
| **First Name: Bhoomi Initials: Mrs.** |
| **Banner No: B00733638** |
| **Course Code: COMP11024** |
| **Course Title: Masters Project 202223 T1** |
| **Project Supervisor: Dr. Raja Ujjan** |
| **Project Title: A Python-based machine learning algorithm approach to**  **improved credit card fraud detection using a Logistic regression model.** |
| **Session: T1 Date of Submission: 22/12/2022** |
| **Signature:** A close up of a note  Description automatically generated with low confidence |

**Please ensure that a copy of this form is included with your project before submission.**

# **Dedication**

To begin with, I want to thank you, my supervisor, Dr. Raja Ujjan, and my moderator Dr. Volkan Tunali, who have supported me during the whole project as well as guided me in the right direction to fulfill the l requirements of the project dissertation. Well, Dr. Raja Ujjain is a very knowledgeable person who gives me comfort during my whole study. In addition, he is a very kind and inspirational person, who respect a lot. I would also like to thank my module coordinator, Dr. Duane west for improving my project report and guiding me with perfect solutions. That encouraged me a lot. My Master’s degree journey taught me a lot like boosting my IT knowledge as well as learning new trends in technologies. Moreover, this university degree journey taught me how to deal with situations and how to react in different situations.

Additionally, from this module, I learned some new algorithms, new techniques, and new methods to develop the project dissertation. This will be very helpful in the future for getting a job as it is based on our best experiences. I saw some ups and downs during my study but still, it taught me a lot in each and every situation.

# **Abstract**

In this technological era, today’s world depends on digital equipment more than traditional equipment. Hence, most payments are done online through credit cards or debit cards or through applications. Due to this reason, there are a bunch of frauds detected in credit card transactions and this issue leads to major ethical problems as it affects financial industries, banking sectors, e-commerce websites, automobile companies, and so on. In addition, some credit card fraud detection refers to a physical, like someone losing the credit card in any place. Due to these circumstances, credit cards are used by someone else, and they do transactions with the help of credit card information, as a result, fraudulent activities are rapidly rising. Therefore, it is essential for any institution and any-commerce website to think about the process of dealing with this ethical issue.

However, there are few websites and some online payment methods are very secure using security settings, so, no one can access the information of someone’s credit card without the permission of the owner. Although, there are some risks involved to trust on this type of website or application. Because a few companies also leak information about the users’ credit cards to other companies or organizations for their own benefit. Due to this reason, no one can trust on the organization as well as on the online payment websites and or applications. Therefore, it will vital to mitigate this issue with the help of some techniques or methods.

There are ample of methods, algorithms, and techniques used to mitigate this problem, but some are not accurate and does not give proper result to detect credit card fraud detection. However, the system generates an enormous amount of legitimate and fraudulent signals at a direct cost, which is the existing problem with this credit card fraud detection. Additionally, examine the characteristics of this fraud and back on previous investigations into this problem. In order to deal with credit card fraud detection, some algorithms and proper approaches will be learned from the current techniques and used them in the correct way.

Furthermore, credit card fraud detection is the process of detecting fraud as well as normal transactions using some techniques, algorithms, and methods such as random forest, Logistic regression, decision tree, artificial neural networking, linear regression, naïve Bayes classifiers, supervise and unsupervised techniques, semi-supervised methods, support vector machine algorithms and so on. These are the main methods or algorithms used to detect credit card fraud detection and give various types of accuracy based on the model’s ability as well as the quality of performance. The past researchers found that some machine learning techniques were not set or does not give accurate results as they thought and then changes some methods or add some new technology to easily detect this credit card fraud detection. Much of the existing literature reviews show that there are numerous mistakes in some methods they do not add accurate methods like confusion matrix, scaling the data, analyze the data also some are not following the structure or the steps to detect false and true transactions, and as a result, it gives false value or does not give a satisfactory result of the process.

This paper uses the python-based machine learning algorithm in order to enhance the method for credit card fraud detection. This approach will generate a Logistic regression model as well as use existing datasets for analysis and visualization. Using this method will detect fraud and legitimate transactions. Firstly, the dataset needs to be analyzed and visualized first and then needs to be cleaned as the dataset ensures that it will interact well with the model. After that will develop and evaluate the Logistic regression model. Moreover, for implementing this Python-based machine learning algorithm will use Anaconda API (Application programming interface) and Jupyter platform to run the algorithms and methods. Quantitative results demonstrate that the Logistic regression model will give approximately 96% accuracy on training data which will be acceptable and after performing this model on test data it will give approximately 96% accuracy, therefore, the combination of the Logistic regression model and python-based machine learning algorithm will considerably detect normal and fraud transactions accurately and efficiently. This paper also identifies the model’s recall, precision, and F1 score to check the Logistic regression model performance on datasets and recognizes the perfect accuracy ratio.

Furthermore, this project visualized the data through the graphs and table to show some results of the proposed method.

**Keywords:** Credit card, Fraud transactions, detection, Machine learning algorithm, Python, Jupyter, Logistic regression model, Normal transactions, Anaconda, API (Application Programming Interface), Accuracy, analysis, visualization.

Table of Contents

1. Research Scope and Existing issues2
2. Introduction5
3. Literature Review6

3.1 RNN machine learning algorithms and LSTM classification methods7

3.2 ML, Neural networking model, classification, and clustering methods8

3.3 The hybrid methods9

3.4 Ensemble of deep sequential model’s methods10

3.5 A supervised and Unsupervised methods11

3.6 A machine learning algorithms methods12

3.7 A machine learning and data science methods12

3.8 Adaboost and majority voting methods14

3.9 A novel text2IMG mechanism methods15

3.10 A mixture of random forest and Adaboost Algorithms15

1. Research design and Methodology**22**

4.1 Dataset24

4.2 Some Python libraries use in this research25

4.3 About Logistic regression model26

* 1. Ethical Consideration27

1. Research design deployment**28**

5.1 Import libraries28

5.2 Load dataset28

5.3 Analyzing and visualizing dataset33

5.4 Create Logistic regression model36

1. Results Visualization**37**

6.1 Count the fraudulent and legitimate transactions through the matplotlib library38

6.2 Correlation matrix and heatmap40

1. Result discussion and analysis**42**
2. Critical Reflection**43**
3. Conclusion**44**
4. References**52**
5. Appendices**75**

# **List of Figures**

1. Figure 4.1: Research methodology20
2. Figure 4.1.1: datasets downloaded from the Kaggle website23
3. Figure 4.1.2: Datasets downloaded from the Kaggle website24
4. Figure 5.1.1: Libraries imported in Python28
5. Figure 5.2.1: Coding of loading dataset in Python28
6. Figure 5.2.2: Output after loading dataset in Python 28
7. Figure 5.3.1: Data exploration in Python29
8. Figure 5.3.2: Data exploration in Python30
9. Figure 5.3.3: Summary of data31
10. Figure 5.3.4: deal with missing values in Python31
11. Figure 5.3.5: encoding method in Python31
12. Figure 5.3.6: Ready dataset in Python32
13. Figure 5.3.7: Splitting dataset in Python32
14. Figure 5.3.8: Training dataset33
15. Figure 5.3.9: Testing dataset33
16. Figure 5.4.1: Logistic regression model created34
17. Figure 5.4.2: Accuracy of the model on testing dataset34
18. Figure 5.4.3: Accuracy of the model on training dataset35
19. Figure 6.1.1: Graph of fraud and normal transactions37
20. Figure 6.1.2: Count fraud and non-fraud transactions37
21. Figure 6.2.1: Coding for generating correlation matrix and heatmap 38
22. Figure 6.2.2: Correlation matrix38
23. Figure 6.2.3: Coding for generating heatmap38
24. Figure 6.2.4: Colored heatmap39

# **List Of Tables**

1. Table 3.1: Literature review lists19
2. Table 7.1: Results of the model performance41

# **Research Scope and existing issues:**

Initially, the world becomes modern due to some new trends of technology but as a result, here are some side effects on people’s lives for example, due to some online transactions frauds are rapidly rising in this era. Due to this reason, everyone loses their money or loses their financial budget. Some major frauds occur through malicious works done for fraudulent transactions which are used for personal as well as financial purposes over online or e-commerce platforms. Therefore, this needs to take time to deal with this issue. A number of companies and financial banks learn some methodologies to detect fraudulent transactions. They started to study various types of machine learning algorithms that are helpful to detect credit card fraud transactions. Owing to this, it is very helpful for users to learn about which one is a legitimate transaction and which one is a false transaction. Hence, humans know about this one is fraud or not and if it is fraud then they need to refund immediately into their account. So, it will mitigate this problem.

According to the survey, the fraudster does this with the help of user information that is uploaded online. Because of the online transactions, all the details of the credit card are saved online portal. Due to this, hackers collected online data from online websites through some techniques and do fraudulent transactions. Moreover, some frauds do this physically for an instance, if someone loses their card, and due to a reason, the fraudster takes this card and collects all information from the bank or online and does a fraudulent transaction.

Well, online websites nowadays store all the information about the user as they can easily process the transactions process. However, some websites are protected with some code or techniques, so no one can collect the other’s information. On the other hand, some are not much protected with the technique, hence it is very easy to collect information through the site.

According to the existing issues will need to find solutions to mitigate these existing problems. Well, some research is done on machine learning methods, artificial intelligence, and data science techniques. Some of them do not create accurate results but still, research will be going on and will improve the credit card fraud detection process. In this process firstly initialized the datasets which received from the financial industries, banking sector, or online. After that research needs to initialize the methods by whine gives accurate efficient result. Furthermore, according to B.C Soundarya, H.L Gururaj, and MJ Madhurya, the mixture of random forest, Logistic regression, support vector machine, and KNN algorithms gives various types of accuracy ratios and gives different results. So, for this reason, researchers need to compare all these algorithms and find out the best algorithm for credit card fraud detection. All the studies found that random forest and Logistic regression models give a higher accuracy ratio rather than the other machine learning algorithms.

# **Introduction:**

To begin with, technologies significantly booming all over the world and due to these trends, online services are also increasing such as online transactions, e-commerce systems, online payment systems, and so on. According to recent times, in this digitalization world, most payments are done online through mobile applications or net banking over the globe to reduce the time of the transaction as well as it is a very convenient and faster process of transferring money. Consequently, these lead to major issues and effects like fraudsters making fraud online using some hacking system and due to online transactions, all credit or debit card information stored online. Therefore, it is the easiest way to leak the data, and hackers to do fraudulent transactions without the owner’s permission.

Financial fraud is an ever-growing imperative ethical problem all over the globe as there are a bunch of frauds detected by online or offline. According to the current situation, online transaction is mostly done on online portal as users shares their information on that website. As a result, abuser leaks the details of card users and do fraudulent transactions using some technology. On the other hand, physical frauds do by if someone losses their credit card in any place, or stolen the credit card, etc. Hence, it is imperative that solved this problem based on the strong and perfect proposed method as it gives absolute accuracy and demonstrates legitimate and fraudulent credit card detection.

Moreover, the existing and main issue of this credit card fraud detection is the cost-effective and cost-sensitive issue in which the system generates a huge number of legitimate and fraud signals at a direct cost. The most popular credit card fraud transaction has significantly impacted financial industries, automobile industries, banks, and so on. A recent survey shows that approx. 27.8 billion dollars were lost due to credit card fraud in 2018, a 16.3% increase compared to 24.87 billion dollars lost in 2017, and it is estimated to reach 36 billion dollars by 2023 (H. Tingfei, C. Guangquan, and H. Kuihua, 2020). The motive for this fraud is crime intention. This act pursing committing fraud is basically embezzling money illegally which leads to losing personal and financial gain (Y. Sahin, S. Bulkan, and E. Duman, 2013). Credit card fraud detection is the use of information that is assigned to the credit card without the user’s knowledge. It is a very ethical issue all over the world and therefore there are numerous processes to solve this problem like data mining, and machine learning algorithmic approaches done for credit card fraud detection. But some methods or techniques did give an exact solutions. However, there is a new algorithm developed to resolve this issue using python and detect fraud credit card when it is going through the transaction.

Initially, there are several existing works done to detect fraud transactions using machine learning algorithms, artificial neural networks, support vector machines, decision tree algorithms, artificial immune systems, and discriminant analysis (Gadi M.F.A, Lago A.P.D, Wang X., 2008). In recent years machine learning algorithms received the main attention of researchers and this method gave a higher accuracy ratio and improved the performance of the process of detecting fraud credit cards.

In addition, by reviewing plenty of existing works the Logistic regression model can perform better than the other classifiers and models. The logistic regression model gives a higher accuracy than the other proposed models, but it takes more time than the other models and classifiers to detect credit card fraud. According to the existing issue, time does not mandatory for these processes, so this model gives efficient and accurate results of credit card fraud detection. Additionally, this model is based on a supervised learning technique that calculates the probability of the variable or data. The logistic regression model is used to solve classification issues. It basically predicts something in numerical values and gives accurate information to the users.

The main goal of this project is to create a credit card fraud detection model, which can determine whether transactions made by a credit card are authentic or fraudulent. Therefore, determining whether a new transaction completed by a consumer is fraudulent or not is primarily reliant on that transaction’s labels and data. From this process will be demonstrated the number of fraudulent transactions and tried to solve this issue.

It is essential to find out about anomalies in the system for some companies’ transactions with the use of credit cards. Additionally, the project will use the Logistic regression model for finding the accuracy of the model to identify whether this model is great to use for credit card fraud detection or not, and will use existing data to perform analyzing, visualizing, and splitting actions. After this process will illustrate the graphs and tables to show the results of credit card fraud detection.

Furthermore, this research project is based on improving this issue like detecting credit card fraud detections using the Logistic regression model. It will be very easy and give accurate results of fraud and legitimate transactions. This project will use a Python-based machine learning algorithm as it is a very efficient and fast detection process. Using this proposed method will separate the fraud and genuine transactions based on machine learning techniques. In this project will use the Logistic regression model for the main analysis of the data and utilize which one is fraud and a legitimate transaction. On top of that, the Logistic regression model is very fast to detect legitimate and fraudulent credit cards rather than other models. Moreover, this model is improving the visualization as well as the normalization process easy than other models or other algorithms. The major aim of this project is to build a credit card fraud detection model, which tells if the transactions made by a credit card are legitimate or fraudulent. Therefore, it is basically based on transaction and their labels as fraud or non-fraud detection processes if the new transaction made by the customer is fraud or not. Additionally, this research project gives three outputs such as the Logistic regression model’s F1 score, precision, and recall to analyze the performance of the model as well as the accuracy.

# **Literature Review:**

Credit card fraud detection is the criminal abuse activity in which someone uses the information of the users illegally online or offline for intended financial and personal usage. There are significantly increasing number of frauds leading to be major crime ratio and that ratio will be soared in the upcoming years. Therefore, it needs to be solved or mitigated this issue to obtain some methods and technique which detects credit card fraud when it will be going into payment. Furthermore, there are ample of frauds registered in the financial industries, banking sectors, and automobile sectors as well.

There are lots of existing methods related to this problem and mostly solved with machine learning algorithms, data science, artificial intelligence, and deep learning algorithm. Numerous literature reviews improved the identifying credit card fraud transactions. Moreover, it proved significantly successful though some had limitations and need to be improved for the future perspective to improve the detection method. Owing to this, the ratio the crime will be decreased in the near future.

## **3.1 RNN machine learning algorithm and LSTM classification method**

The RNN and LSTM classification-based machine learning method discussed that in a digital society, online payment systems like Paytm, google pay, banking app, etc as well as other online tools or applications have become rapidly booming all over the globe. Hence, it is a very difficult issue to provide trustworthy services online or physically as credit card information is stored online automatically by machines also, it’s an essential point to provide security in each and every platform as frauds are rising in online transactions through some techniques(Chen C., Zhang Y., Zhou J.).

The methodology of this research paper was based on Machine learning algorithm architecture. The first step is data processing in which the performance of the model was improved and collected more accurate results. After this step data were analyzed and visualized based on a machine learning algorithm and then some duplicate data were deleted from the datasets (J. Pun, Y. Lawryshyn). Then after, for imbalanced datasets using the resampling method data was ready for another find out the results. In the third step, data were divided into two parts which were the training dataset and testing dataset, in which on the training data the LSTM (Long short-term memory) machine learning algorithm model classification technique is applied. Well, RNN (Recurrent neural network) is a more efficient neural network method than the LSTM classifier because it is used in developing a deep learning method for the detection method. After these processes metrics methods are applied like TNR, FNR, TPR, and FPR for rate measurement as they detect fraudulent and legitimate transactions (Ekrem Duman, M. Hamdi Ozcelik).

It was concluded that the result of this methodology was given effective and accurate result. The LSTM classifier model and RNN machine learning algorithm improve the outcomes of the layers and interactions. This algorithm accurately detects credit card fraud detection and gives a quick result. Both machine learning algorithms are very strong and give an accuracy ratio based on recall, and precision.

Overall, this literature review analysed that using resampling data and then dividing it into two parts like testing and training datasets improves the model accuracy in the credit card fraud detection process and gives the accurate result of the project. Owing to this it is the best way to improve the dataset and then apply algorithms to it. This will easiest way to remove unbalanced data.

## **3.2 ML, Neural Networking model, classification, and clustering method**

The ML, neural networking model, classification, and clustering method based on “exploratory analysis of credit card fraud detection using machine learning techniques”. The main research problem of this project was that the online and e-commerce platforms significantly increased the fraudulent transactions, and their percentage went up in different types of payment methods. For instance, the rate of credit card fraud transactions increased from 0.06% in 2016 to 23% in 2018. Additionally, 10% of all frauds were considered as online methods by credit cards which have encountered huge financial losses (H. L Gururaj, M. J Madhurya).

The methodology of this research is based on the different types of machine learning algorithms and clustering techniques. This literature review worked on both supervised and unsupervised algorithms to detect fraudulent credit cards. Supervised techniques used classification of the algorithms such as decision tree, naïve Bayes algorithm, linear regression, neural network algorithm, and support vector machine. On the other hand, unsupervised techniques used clustering techniques like partition-based and hierarchical-based clustering. The main aim of this review is to deal with accuracy, efficiency, and performance between classifiers and clustering methods. This literature paper recognized all classifiers and chose one of the better classifiers to easily detect fraud transactions. Additionally, it used oversampling techniques and binary classifiers and observed the machine learning algorithms which were paired with other techniques. Owing to this, it identifies better results and efficiency of the research.

This paper concludes that the main result was calculated by using a confusion matrix method which is a machine learning algorithm. Most of the algorithms have been proven to be accurate and efficient although there could be irregular data and dissimilar datasets results occurred. For this reason, this research will improve some techniques and methods to eliminate this problem of the dataset. Need to learn about the resampling dataset methods that will be very helpful to give accurate output.

Overall, this literature review is useful in the credit card fraud detection process. From this research will need to implement the confusion matrix method to count how many transactions are frauds and how many transactions are normal. Owing to this will analyse how many transactions will be done in a time and improve the model’s efficiency. In addition, the confusion matrix is also helpful to find out the precision, recall, and F1 score. Because of these three criteria, the accuracy of the model will easily identify.

## **3.3 The hybrid Method**

The” hybrid method with dynamic weighted entropy for handling the problem of class imbalance with the overlap in credit card fraud detection”. The major research problem of this paper was most of the fraud transactions are overlapping with legitimate transactions. This problem is very challenging hardly recognize fraudulent credit card transactions with this issue. In addition, there were also issues with class and data unbalanced and due to this, overlapping between different types of classes does not recognize credit card fraud transactions (Japkowiez and Stephen,2002).

The proposed method of this research was based on two types of novel hybrid methods. The hybrid method consists of two steps divide step and conquer step. Well, the divide step includes an anomaly detection model such as an isolation forest (Liu, Ting, and Zhou, 2008), and an auto-encoder (Paffenroth and Zhou, 2017) for learning the dataset minority and improving the accuracy of the hybrid model. The Conquer step used random forest classifiers and random forest algorithms (Beriman, Basheer, and Hajmeer, 2000). These techniques eliminate the issue of overlapping the datasets like missing datasets values as well as repeated values. For this reason, it was rising the efficiency and accuracy score ratio for credit card fraud detection. This existing method used a dynamic weighted entropy technique for improving the overlapping and non-overlapping subsets of the datasets as well as the performance.

The result of this research was not really satisfactory because it was a very time-consuming process giving low efficiency. This research did not accurately detect credit card fraud detection. For this reason, this paper applied another DWE (Difficulty with Evacuation) parameter which was based on the weighted entropy method and its influence on the performance of the model and accuracy (Rodger, 2019). It minimizes the overlapping and misclassification issue and balances the dataset.

In a nutshell, this literature review helps to deal with missing data and overlapping data to easily detect fraudulent transactions. So, this technique is very helpful to deal with misclassification issues and give accurate results using a model. This process is time-consuming but overall, it detects fraud transactions efficiently.

## **3.4 Ensemble of deep sequential models’ method**

The literature review of the “ensemble of deep sequential models for credit card fraud detection” main research problem is the soaring amount of credit fraud transaction losses of billions of dollars in various financial industries and banking sectors. Owing to this, if it is not stopped then it becomes worse in the upcoming years, and all people lose their financial budget. Therefore, it needs to be necessary to follow the machine learning and artificial intelligence methods of credit card fraud detection. However, there were a number of challenges and difficulties that occurred in the research including skewness of the data, concept drifts, short-time response, cost-sensitive of the application, and pre-processing of the features (Behnsen A., Ottersten B. 2016).

The methodology of this research is based on running networking systems like LSTM (Long short-term memory), FFNN (Feedforward neural network), and GRU (Gated Recurrent Unit). This paper used these three deep learning processes in which they aggregate all of them and then decided which one was the best for the credit card fraud detection process. These three networks are mainly base classifiers. This research used voting mechanisms and give output based on the FFNN network deep learning method. The FFNN is the ensemble model that detects fraudulent transactions with the help of LSTM and GRU output. Additionally, this ensemble sequential deep learning method used a confusion matrix to find out the accuracy of the model, and it included Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) values which found the accuracy ratio.

As a result, According to Granitzer M., Jurgovsky J., and Portier P.E. in 2018, the LSTM model gave accurate and efficient output using the FFNN methods. Output considered using compared both GRU and LSTM values of the ratio. Both models gave great results but the LSTM model the gave highest accuracy ratio than the GRU model. Well, LSTM is the second baseline classifier whereas the LSTM model is the first baseline classifier. These two models perform on two datasets one was the European dataset and another was the Brazilian dataset.

Overall, this deep sequential ensemble learning technique is very helpful but it’s a time-consuming method and does not give significant results to detect fraudulent credit cards. On the other hand, some techniques like confusion matrix, and comparing model is the best way to identify the accuracy ratio and give the appropriate result for credit card fraud detection.

## **3.5 A supervised and unsupervised methods**

This literature review is based on “combined supervised and unsupervised learning in credit card fraud detection”. The main research problem was it is difficult to analyse customer behaviours and its difficult to recognize the ability of the fraudster who wants to take advantage of doing fraudulent transactions using novel pattern techniques. (A.C Bahensen, D. Aguada). Therefore, it is essential to deal with this ethical problem using machine learning techniques.

The proposed method of this literature review was a combination of a hybrid method of supervised and unsupervised learning techniques. Both methods are machine learning algorithm techniques. The supervised methods build a model on the datasets to detect fraud transactions and non-fraud transactions and give an accuracy score. Unsupervised methods used the RNN networks, LSTM networks, and GRU networks to prevent fraudulent transactions and normal transactions. These approaches give superior performance according to the other single machine learning model. Additionally, this existing work is done by ensemble methods like deep machine learning to learn about simple voting mechanisms as well as give results of supervised and unsupervised outlier scores. This whole process was done on both testing as well as training datasets and try to improve model’s performance. Based on the model performance this research paper count accuracy and then compares both supervised and unsupervised techniques. After the comparison model gives two accuracy ratios and decided which one was good.

The result of this approach was not very convincing and did not give a proper accuracy ratio. However, the unsupervised outlier classifier model properly detects false and true transactions rather than the supervised techniques. Both hybrid methods worked on training as well as testing datasets. Overall, this research paper gave great and efficient output but not satisfactory results. Therefore, it needs to be improved using another machine-learning algorithm.

## **3.6 Machine Learning algorithm methods**

This research was based on “credit card fraud detection machine learning methods”. Nowadays, numerous companies are providing the best services to their customers and owing to this they need to store information of the users online on their websites. But it needs to be secure so that no one can access this customer’s data as the purpose of the fraud. On the contrary, some companies sell fake items or things to customers to take the information of the user and do fraudulent transactions on online websites. It is the main problem of this research and needs to be eliminated these issues (Dejan Varmedja, Mirjana Karanovic, 2019).

The methodology of this project was developed classical algorithms that are very useful. All algorithms must be run on training datasets and generate accuracy of the fraud or legitimate credit card transactions. This paper analysed the different types of machine learning algorithms such as Random Forest, naïve Bayes, multilayer perceptron, and Logistic regression, and compares all these algorithms and finalizes which gives the best accuracy for credit card fraud detection. Initially, the Logistic regression algorithm uses for prediction in this paper and gave an accurate accuracy ratio that mainly depends on training datasets. On the other hand, the Naïve Bayes algorithm is basically used for supervised learning techniques based on the Bayes theorem. Well, the random forest algorithm is used for both classifications as well as for regression problem which is based on the decision tree algorithm. On the other side, multilayer perceptron worked on an artificial neural network that consists of three layers of nodes a hidden layer, an output layer, and an input layer to identify the weighted sum of the bias which is based on datasets.

The conclusion of this method was to use the confusion metrics method for analysing the result of machine learning algorithms like gradient boosting, LR (Logistic Regression), RF (Random Forest), decision tree, and vector machine to give accurate rates on the precise basis for credit card fraud detection. However, according to the result, the Logistic regression algorithm gave the highest accuracy ratio than the other algorithms which was used in this paper. The output of the accuracy ratio was decided on the confusion metrics technique in which they used calculation of the three parameters such as precision, recall, and F1 score.

Overall, this machine algorithm gave satisfactory results and easily detect fraudulent transactions. In addition, according to the output the Logistic regression model outperforms and gave the best result. This model proves to a state of the art in this research.

## **3.7 Machine learning and data science methods**

This paper was based on “credit card fraud detection using machine learning and data science”.

Credit card fraud transactions define as an unauthorized person using someone else’s credit card for personal reasons or any other reason. The major issue of these papers was the unwanted usage of the user’s account by the fraudster illegally without the owner’s permission. It is vital to minimize this ethical problem and necessary to implement some steps to mitigate this problem from the world. In addition, there are some machine learning algorithms with data science methods that tackled this issue (Aditya Saini, S P Maniraj, Shadbad Ahmed, Swarna Deep Sarkar, 2019).

The method for this project was to use the latest machine learning algorithms to detect abnormal credit card fraud detection such as Local outlier factor and Isolation Forest algorithms. These two algorithms are mainly part of the Python sklearn library which was imported into Python. This research paper used a dataset downloaded from the Kaggle website. These two algorithms were implemented with the help of machine learning algorithms as well as data science techniques. Well, this paper used some graphs to analyse and visualize the dataset as well as the fraud transactions. In the end, compared the Local outlier factor and the Isolation Forest algorithm’s output for found the accuracy ratio.

In nutshell, the result was based on printed numbers of False Positive value that calculated the precision, recall as well as F1 score of two algorithms. This value was compared with the actual value and then identified that its accuracy ratio was 99% but the dataset was imbalanced when got the accuracy ratio.

Overall, this method is very fruitful though without analysing and visualizing data it gives mixed fraud and legitimate transactions. These machine learning and data science methods have proven successful as they detect fraudulent transactions.

## **3.8 Adaboost and majority voting methods**

This research was conducted by Kuldeep Randhawa, Chu Kiong Loo, and Manjeevan Seera based on “credit card fraud detection using Adaboost and the majority voting in 2018”. The problem statement of this research was credit card fraud is an illegal activity that is done digitally or physically. This is done for criminal or personal purposes. Physical fraud is done through offline transactions and digital transactions are done by the internet or some online payment methods. This research paper adopted Adaboost and the majority voting methods for detecting fraudulent transactions when it is going through the transaction.

This research paper used twelve machine learning algorithms with the help of the Adaboost and the majority voting model. The twelve algorithms are random tree, random forest, decision tree, logistic regression, Naïve Bayes, deep learning method, neural network, MLP (Multi Layer perceptron) network, gradient boosted tree, decision stump, linear regression, and SVM (support vector machine). These all algorithms did different tasks to detection of fraud transactions. The Adaboost was used for improves the performance of the algorithms and then combines all outputs using boosted classifiers. The majority classification method is used for data classification in which it’s involved with the two algorithms.

Well, the limitations of this research were the need for a good understanding of abnormal fraudulent cases. As compared to the data mining technique the performance of the classification was poor during credit card fraud detection using the Logistic regression method. the vector machine algorithm was not the easiest method because of the transferring of the input data. Additionally, the decision tree algorithm is limited on datasets like it is overfitting on the training data so, for this reason, retraining of the dataset was required. The neural network needs high computational power for real-time operations as well as required for retraining the dataset.

This paper had given results on the measurement of the MCC (Matthews Correlation Coefficient) metric methods performance. And using majority voting the MCC score achieved 0.823 was the best score using this technique. Moreover, using the Adaboost and voting methods the MCC score was 0.942 which was very efficient and gave accurate output.

Overall, this paper will learn about how to measure accuracy scores using algorithms or some techniques. It has many limitations for detecting fraud transactions accurately but also it is very time consuming process.

## **3.9 A novel text2IMG mechanism method**

This research paper was based on the “Novel text2IMG mechanism of credit card fraud detection a deep learning approach “(Abdullah Alharbi, Hafiz Tayyab Rauf, 2018). The problem statement of this project was that significantly increased fraud in credit cards through online or offline which led to the use of fake ids of users for personal usage or for financial purposes. It is a very ethical issue in the world, therefore, need to apply some techniques to detect fraudulent transactions. Credit card fraud detection is a very challenging ask as there are a number of common transactions done every day rather than fraud transactions. There were data mining and machine learning techniques to identify credit card fraud detection, but in this paper, there are two new methods developed for detecting fraudulent transactions.

The methodology of this research was based on image processing and deep learning approaches with the help of machine learning algorithms. The method was the text2IMG conversion approach that converts text data into image format. This method is called the digital lattice technique. The digital lattice was then iterated over a mask, and a final image was prepared with various intensity levels. The given data in image format was then fed into the proposed CNN (Convolutional neural network) and classical ML (machine learning) methods via pipeline features. This research paper used the inverse frequency method that is based on machine learning algorithms to solve the problem of class imbalance and for that, it assigned the class weights. In addition, this research also used GMM (Gaussian mixture model) for the recognized probability density function (PDF)of the model.

In conclusion, the classification proved that it detect fraudulent credit card accurately on the basis of both deep learnings as well as machine learning algorithms. Well, this paper required more data to improve the model evaluation process. Moreover, the proposed method used a novel text2IMG as a computer vision technique was proven successful and achieved a great result. However, machine learning based classification methods does not reduce the time complexity but reduce the space complexity in the credit card fraud detection process.

Overall, it can be seen that this research paper give satisfactory result but need more data for converting the data into an image as computer vision.

## **A mixture of random forest algorithm and Adaboost algorithm:**

This research paper on credit card fraud detection using machine learning algorithms compares the two algorithms and gave the best result. The problem statement of this research was due to rise the of online transactions and e-commerce platforms, credit card fraud detection increases around the globe, and most of the users are lost their financial status due to these circumstances. Therefore, it is crucial to deal with these issues and apply some strict steps to solve this problem in the world.

The methodology of this paper was its applied machine learning algorithms such as the random forest algorithm and Adaboost algorithm. First of all, this research classified the fraud and normal transactions from the datasets using two algorithms AdaBoost and random forest. After this step, these two algorithms were compared and give the accurate result of credit card fraud detection. This process includes first splitting the data, model training, model deployment, and evolution of the model. After evaluation, it gave the accuracy ratio of these both algorithms and find out which had a higher ratio. This accuracy ratio finds out in the plotting graphs and tables. For this whole process, this paper loaded the dataset and divided it into two parts one is the training dataset, and another is the testing dataset.

In a nutshell, using confusion matrix curves it is initialized that the random forest algorithm and Adaboost algorithm gave different outputs after being applied to testing and training datasets. After the comparison was finished it is analysed that random forest gave a higher F1 score, precision, and recall ratio than the Adaboost algorithm. Additionally, the random forest algorithm gave faster and more effective results in credit card fraud detection.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Article | Date | Research Problem | Methodology | Significant Result |
| “The Autonomous credit card fraud detection using machine learning approach” | 2022 | Fraud detection is a cost-effective problem in which the system obtained many false fraud signals for the transaction. | This article used the LSTM-RNN architecture. Which used existing datasets and visualize and analyse data and then resample the data. | Using the RNN algorithm and LSTM classification, the model effectively predicts credit card fraud. |
| “The Exploratory analysis of credit card fraud detection using machine learning techniques.” | 2022 | The online websites and e-commerce platforms predict that the fraud rates as per market ratio and frauds rose from 0.06% in 2016 to 23% in 2017. | The proposed methods are based on ML, Neural Networking models, classification, and clustering techniques. | These techniques and algorithms are proven as methodical and error-free if either implemented alone or with a combination of two. |
| “The hybrid method with dynamic weighted entropy for handling the problem of class imbalance with the overlap in credit card fraud detection.” | 2021 | One imperative reason for the problem is a class imbalance and owing to this the overlap between imbalance classes makes it difficult to identify credit card fraud detection. | To overcome this problem applied a novel hybrid method using the idea of the Divide-and-Conquer method in a hybrid method. It consists of two steps, Divide and Conquer. | In a nutshell, the DWE parameter provides decision-making rules that encounter the execution of the model. Maximize the number of majority and minority subsets in the overlapping classes and gives the perfect accuracy on datasets. |
| “Ensemble of deep sequential models for credit card fraud detection” | 2021 | The problem of research because of the automatic online fraud detection system should get a large volume of losses in financial industries as well as some other industries. | The proposed method use a deep sequential of five approaches such as probabilistic graphical model, machine learning, deep learning, ensemble, artificial immune system, and Linear fisher discriminant analysis. | The result of two real actual datasets illustrated that the proposed ensemble model based on two LSTM was not accurate. Additionally, the evaluated solo LSTM model significantly performs faster than others. |
| “Combining unsupervised and supervised learning in credit card fraud detection” | 2018 | The use of fraudulent patterns that can be equipped from online websites it’s a very ethical issue. | Using supervised and unsupervised techniques for the existing dataset the data is distributed in two parts one is training and the other is a test dataset. | This article suggests the execution of a hybrid approach with the help of unsupervised and supervised learning techniques it makes good scores for the detection of the fraud classifier and credit card transactions. |
| “Credit card fraud detection machine learning methods” | 2019 | Online sales are fraud where fraudsters sell fake things or copy items to take payment without buying or delivering the items. | Using classical algorithms such as gradient boosting, support vector machine, decision tree, LR, and RF is very accurate. | Using metrics methods for the result of machine learning algorithms are error free, and give perfect accuracy ratio. |
| “Credit card fraud detection using machine learning and data science” | 2019 | Fraud in credit card transactions is based on an unauthorized person using the owner’s account for fake transactions which is a very major issue. | The use of the latest machine learning algorithms to detect abnormal activities as well as detect fraud or normal transactions separately. | The algorithms did reach over 99% accuracy as well as detect the fraud transaction effectively. |
| “Credit card fraud detection using Adaboost and majority voting” | 2018 | More losses because of fake transactions affect industries, where they brace all the costs.. | This paper uses twelve machine learning algorithms for detecting fraud transactions. In addition, they also use Adaboost and majority voting methods for hybrid models. | The using MCC method it was easily predicted the false transactions using majority voting. The best MCC score was 0.823. |
| “A novel text2IMG mechanism of credit card fraud detection: A deep learning approach.” | 2022 | This fraud significantly increases and has an effect on fake transactions using unwanted usage of user information, which has become an increasing security issue. | The proposed method uses image processing in which the text is converted into an image. Additionally, using deep and machine learning-based approaches. | The classification results prove that the image compression method was detect the accurate result of normal and fraud transactions as well as gives 100% accuracy. |
| “Credit card fraud detection using machine learning algorithms” | 2020 | The main problem of this paper is raising credit card fraud transactions and its effects on e-commerce websites and online payment methods | This paper used two algorithms such as random forest and Adaboost algorithms. This compared both algorithms and analysed them based on the accuracy ratio. | In a nutshell, it was initialized that random forest gave an accurate and higher ratio than the Adaboost algorithm after being applied to testing and training datasets. |

Table 3.1: Literature review lists

# **Research Design and Methodology:**

Initially, credit card fraud detection is the process of detecting legitimate and fraudulent transactions.

It basically tells users that transactions will be fraudulent or normal. This research is based on a python machine-learning algorithm using a Logistic Regression model. The research method uses secondary datasets which are collected from Kaggle and using this dataset will find out the legitimate as well as fraudulent transactions. Additionally, this research also measures the accuracy and performance of the Logistic regression model that will help to detect credit card fraud transactions. This research basically uses the transactions as well as the labels to detect the fraudulent card using the Logistic Regression model. The below figure shows the exact step-by-step process of the project.

Graphical user interface

Description automatically generated

Fig 4.1: Research methodology

By Bhoomi Sharma

Well, the above figure shows the process of credit card fraud detection. This research will be based on a python machine-learning algorithm by Logistic regression model that improves the accuracy as well as the model performance. This proposed method is applied to the datasets which downloaded from the Kaggle website. It is a very old dataset hence; it needs to be analysed and visualized the dataset before model development. Moreover, it is basically a secondary dataset so do not need any ethical consideration. This research will investigate or recognize normal and fraudulent transactions as well as it will separate and calculate the false and true transactions.

The research methodology section gives an overview of the whole proposed methods that will be used in this research. The whole process will be conducted based on various machine-learning techniques. Well, the whole proposed method explains in the below trailed paragraph.

1. To begin with, this project first collects the dataset which is downloaded from the Kaggle website. Secondly need to analyse whether the dataset is imbalanced or not. The dataset is in CSV format, which is 143 MB data size. Furthermore, this project must employ Python as a computer language and it is a very easy to use, interpreted, and object-oriented programming language. After the collection of the dataset need to install Anaconda API (Application Programming Interface) and then install Jupyter for coding ad running the deployment of the project. Initially, Jupyter is the Python programming language platform.
2. Secondly, after this process need to install the libraries as it is a member of the Python programming language. Without libraries Python, machine learning algorithms can not work properly. Libraries worked as a multi-dimensional group. In addition to this, It will be used for collecting information and modification. However, these libraries are very efficient to install and use as well.
3. Numpy
4. Pandas
5. Matplotlib
6. Seaborn
7. SKLearn
8. Thirdly, After the installation of libraries needs to be imported the dataset on the Jupyter platform. And then perform explanatory data analysis on the dataset for getting a balanced dataset. This process includes cleaning the data and making it ready to train the model and it is very helpful to improve the model’s performance.
9. After this process needs to separate the fraud and genuine transactions and then analyse and visualize the fraud as well as the legitimate transactions. After this method data goes for analysing and visualizing process for retrieving the missing values. This process used an encoding algorithm and feature scaling algorithm to get ready data to develop the Logistic regression model.
10. Then, data will be divided into two parts first is the training dataset and the second is the testing dataset. After this data will be split into labels and features. Training and testing dataset gives a higher accuracy ratio. It will be calculated by the confusion matrix algorithm.
11. Finally, need to create the Logistic regression model to get an accuracy ratio it will detect the fraud transaction more efficiently. After this process, data will be fitted into this model, and then will perform the model evaluation techniques that are based on the machine learning algorithm. In order to perform this evaluation method model, test the data on training as well as the testing phase. Both training and testing datasets give model accuracy ratios using the confusion matrix algorithm.
12. For finding the accuracy ratio need to find out firstly True Positive value, False Positive value, False negative value, and True Negative value of the confusion matrix method. These four variables identify the precision, recall, and F1 score ratio which analyse the model’s accuracy ratio.
13. Lastly, after evaluating this process will analyse and visualize the data using charts and tables. This process will include some results such as showing normal and fraudulent transactions and calculating the fraud and legitimate transactions. It will help to show the accurate output of the research and also conclude the whole proposed methodology. Well, to perform this task will use the Matplotlib library and use correlation matrix and heatmap techniques to show graphical representations.
14. Additionally, in the last this research project utilizes fraud transactions, therefore, will add some screenshots of the Python-based machine learning algorithm coding step by step. Moreover, these coding screenshots show the implementation of the project.

## **4.1 Dataset:**

This research uses datasets that are collected from the Kaggle site, and it consists of CSV file. The dataset consists of transactions that occurred in two days, where there are 150 frauds and 51440 legitimate transactions shown. The main point of this dataset is confidential transactions, and it contains only numerical values. It’s coded like V1, V2, V3…. V28. Additionally, the dataset includes time which shows the seconds of transactions done within two days whereas the amount column shows the monetary values of the transactions. The Class indicates the binary value 0 and 1 values in which 0 consists of legitimate transactions and 1 consists of fraudulent transactions. There are fewer fraudulent transactions than normal transactions. Initially, this dataset is highly unbalanced in that there are 17% of labelled transactions.

Hence, it is crucial to preprocess the dataset and it plays an important role to improve model accuracy and performance.

Graphical user interface, application, table, Excel

Description automatically generated

Figure 4.1.1: Dataset downloaded from the Kaggle website

Table

Description automatically generatedFigure 4.1.2: Dataset downloaded from the Kaggle website

Moreover, a dataset is highly unbalanced so need to balance the dataset first. Initially, data will be initialized and then visualize. For this process first will find out missing values and it contains 20 missing values. After this process, the dataset will be encoded, and then count the fraudulent and legitimate transactions and shown in the Matplotlib library using correlation matrix and heatmap algorithms.

Secondly, need to split the datasets into two parts testing and training datasets, and apply the feature scaling method to differentiate the dataset using preprocessing technique using the standard scalar method. Lastly, this dataset will be set and, on this dataset, will apply the Logistic regression model algorithm.

These datasets are collected and analysed from the worldline and machine learning group during research. This dataset is very old, and it has so many transactions done online as well offline so for that using a confusion matrix needs to analyse data like how many transactions are done per second and how many transactions are genuine or how many are frauds. This dataset is very helpful to find out the model accuracy ratio and also recognize the model’s ability to detect fraud and legitimate transactions.

## **4.2 Some Python Libraries use in this research:**

To create a credit card fraud detection process, need to use some Python libraries.

* NumPy
* Panadas
* Matplotlib
* Seaborn
* Sklearn

These libraries are very efficient to use as well as a need to install in the Jupyter platform using the pip function. The Matplotlib and seaborn library act as a backbone of data visualization in Python. Additionally, the Matplotlib library is mainly used for plotting graphs with the help of other libraries like NumPy and Pandas. It is a powerful tool for visualizing the data.

Pandas is a Python library for data analysis as well as acts as a wrapper over the NumPy and Matplotlib. to allow to access a bunch of NumPy’s and Matplotlib’s methods with less coding.

NumPy is a leading scientific computing library in Python. It is also used for mathematical functions in Python. This library mostly works with arrays for coding like c programs.

Sklearn is the most useful and robust library for the machine learning algorithm. It provides a selection of efficient tools for machine learning as well as statistical modeling including classification, regression, clustering, and dimensionality reduction.

On the other hand, for the installation process the use command prompt to install these libraries like:

Pip install NumPy pandas matplotlib seaborn sklearn.

## **4.3 About Logistic Regression Model:**

The logistic regression model is a type of supervised and unsupervised learning technique, and it is a well-known machine learning algorithm, which is used to predict the probability of values or variables. This model predicts the output of a categorized variable which means that it has only two possible classes, one is true and the other is false. Basically, this model mostly improves performance and gives accurate results.

This is one of the most efficient, reliable, and simplest machine learning algorithms used for different types of classification problems like fraud detection, spam detection, cancer detection, COVID cases detection, etc. Moreover, this model gives 95% accuracy in the detection process and gives the perfect output of the detection process of a machine learning algorithm. This accuracy find out from the existing research.

For creating a Logistic regression model first need to import the dataset and then do normalization to set unbalanced data and then split the data. After this process creates the Logistic regression model using a machine learning algorithm in Python. After creating this model need to evaluate the model on testing data as well as training data to check the accuracy of the model.

In addition, the Logistic regression model’s accuracy recognize based on the precision, recall and F1 score ratio which will identify using confusion matrix algorithm. Well, precision ratio get from the true positive value divided by the sum of true positive and false positive value. On the other hand, recall identify by the true positive value divided by the sum of true positive and false negative value.

F1 score recognize by the multiplication of the recall and precision multiply by 2 and divided by the sum of recall and precision.

Moreover, Logistic regression model gives accuracy ratio from these three parameters which will be created by the confusion matrix. The Logistic regression is significantly used algorithms as many researchers add this model for prediction of the data and it recognize the accurate result and give higher ratio than the other algorithms. It is used for classifying the main predictions of the variable and its mostly use classification algorithm. But in this paper will use the confusion matrix algorithm for creation of model as well as find out accuracy ratio.

## **4.4 Ethical Consideration:**

The research used the existing datasets that were downloaded from the Kaggle website. For this dataset, no ethical approval is needed because there is uninvolved user data through Practical implementation or surveys. Basically, the research project will be performed on existing data, and it has no privacy concerns regarding misuse of the user data performed in this project. This research will be conducted with usable datasets which are static data, therefore, no need to be ethical approval for developing this project.

# **Research Design Deployment:**

## **Import Libraries:**

To begin with, first of all, import the libraries of python in the Anaconda API in the Jupyter platform.

Graphical user interface, text

Description automatically generated

Figure 5.1.1: libraries imported in Python

## **5.2 Load Dataset:**

After installation needs to load the dataset which downloaded from Kaggle which is in a .csv file.

Graphical user interface, text, application, email

Description automatically generated

Figure 5.2.1: coding of load dataset in Python

The output is shown below.

Graphical user interface, text, application

Description automatically generated

Figure 5.2.2: output after loading the dataset in Python

**5.3 Analysing And visualizing the Dataset:**

In this step analyze the dataset and perform some operations to ready unbalanced data into balanced data for the model and then visualize the dataset. Moreover, this step includes encoding and dealing with some missing values for visualization.

Text, letter

Description automatically generated

Figure 5.3.1: data exploration in Python

Initially, data exploration has two highlighted traits like one single variable and the other revealing patterns as well as relationships between variables. The above figure indicates the relationship and patterns of the datasets.

Table

Description automatically generated

Text

Description automatically generated

Figure 5.3.2: data exploration in Python

A screenshot of a computer

Description automatically generated

Figure 5.3.3: Summary of data

Graphical user interface, text, application, email

Description automatically generated

Figure 5.3.4: deal with missing values in Python

In this process, data is first analyzed whether there are any null values if it is true then sum all the null values that show 20 on the above figure. And solve the missing value issue improves the model performance.

Graphical user interface, text, application

Description automatically generated

Figure 5.3.5: encoding method in Python

Well, encoding is the process of machine learning technique in which it converts categorical data into numerical values so this could easily fit in the Logistic regression model. The above figure shows that encoding is successfully done with a 0 value and data is set now. After this step needs to split the data set for creating the Logistic regression model. This process is done on the x and y-axis.

Graphical user interface, text, application

Description automatically generated

Figure 5.3.6: ready dataset in Python

Graphical user interface, text, application, email

Description automatically generated

Figure 5.3.7: splitting dataset in Python

After this step scaling the dataset to separate the test dataset and train data set. This will help to analyze and visualize the fraud and legitimate transactions as well as the Logistic regression model accuracy.

Graphical user interface, text, application, email

Description automatically generated

Figure 5.3.8: training dataset

Text, table

Description automatically generated with medium confidence

Figure 5.3.9: testing dataset

Using the feature scaling method dataset automatically separates into testing and training parts using the standard scalar library. It’s called preprocessing of data.

## **5.4 Create Logistic Regression Model:**

Now create the Logistic regression model to find out the accuracy ratio, and precision call ratio and improves the credit card fraud detection process.

Graphical user interface, text, application, email

Description automatically generated

Figure 5.4.1: logistic regression model created

Logistic regression is one of the best and most well-known classification algorithms in machine learning techniques. It gives recall, precision, and F1-score for initializing the accuracy of the model. After creating a model, we need to evaluate the model on testing and training datasets to check the model’s accuracy using the confusion matrix method of the algorithm.

Graphical user interface, text, application, email

Description automatically generated

Figure 5.4.2: accuracy of the model on testing data

Graphical user interface, text, application, email

Description automatically generated

Figure 5.4.3: accuracy of the model on training data

As the above figure, it shows that created Logistic regression model gives 99% accuracy on training as well as testing datasets. It means that this model is giving good accuracy to find legitimate and fraudulent transactions. The Logistic regression model is very accurate and efficient for credit card fraud detection.

Furthermore, the confusion matrix consists of an N \* N matrix which is used for the evaluation of the classification model in the machine learning algorithm. This matrix compares a relevant value with those predicted by the logistic regression model. It improves the performance of the model. These targeted values are positive and negative. There are four targeted values first is a true positive (TP) which predicts fraudulent transactions, the second is a false positive (FP) which indicates the number of normal transactions, the third one is a false negative (FN) indicates the fraudulent transaction that acts as a normal, and the fourth one is true negative (TN) indicates as the normal transaction.

Moreover, recall estimates the fraud transaction ratio:

Recall = TP / TP + FN in which TP = True Positive, FP = False Positive, and FN = False Negative

Precision identifies the percentage of the fraud transactions which is detected correctly:

Precision = TP / TP + FP

The F1 score measures the harmonic mean of the recall and precision

F1-score = 2 \* Precision \* Recall / Precision + Recall

# **Results Visualization:**

Initially, in this visualization section need to analyze and visualize the datasets in the matplotlib library in the Jupyter platform. This section divides into two parts first need to count how many transactions are fraudulent and how many transactions are normal. After this process needs to apply heatmap and correlation matrix functions to analyse the time and amount of the transactions.

## **6.1 count the fraudulent and legitimate transactions through the Matplotlib library.**

Graphical user interface, application

Description automatically generated

Figure 6.1.1: graph of fraud and normal transaction

Graphical user interface, text, application, email

Description automatically generated

Figure 6.1.2: count fraud and non-fraud transaction

From the above figure, it will be seen that there are 51400 normal transactions occurred and 150 fraudulent transactions occurred. According to this, it is easy to separate normal and fraudulent transactions. This is the count method that initializes from the graph that the blue bar is the legitimate transaction by credit card and the smallest bar is counting the fraudulent transactions.

Normally, this counting process is done by two libraries seaborn and matplotlib. Using the seaborn library easily checks non-fraud and fraud transactions. And matplotlib library is used for generating the graphs to easily visualize and analyse the dataset.

## **6.2 Correlation matrix and heatmap:**

After analysis of the datasets, need to plot the correlation matrix and heatmap graphs to recognize the data analysis. Additionally, this process needs to plot a colored heatmap to represent the data as well as to study the correlation between the out-predicting variable and classes. Both graphs are shown below.

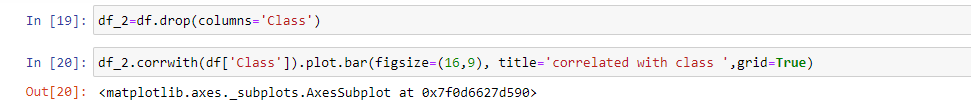


Figure 6.2.1: coding for generating correlation matrix and heatmap

Chart, waterfall chart

Description automatically generated

Figure 6.2.2: correlation matrix

Graphical user interface, text, application

Description automatically generated

Figure 6.2.3: coding for generating heatmap

Chart

Description automatically generated

Figure 6.2.4: colored heatmap

Moreover, the correlation matrix and heatmap define the plotting of the graph to show the strength of the relationship between the numerical variables which is represented by the column. In addition, correlation matrix and heatmap plots are used for detecting the linear and nonlinear relationship between variables. It is the graphical representation of the data that indicates the information about the variable in a colored form of a graphical view.

The above two graphs show the statistical analysis using a heatmap and correlation matrix. Using this analysis, it can analyse the time and amount of all transactions done by the credit card. In these graphs, time and amount are carried out in numerical because of the confidential purpose. They are coded as V1, V2, V3…. V28 and the time attribute show the seconds that occurred between the transactions. Amount attribute leads to a cost of transactions that saw in numerical. Lastly, the class attribute shows the legitimate and fraudulent transactions such as 1 as a fraud transaction and 0 as a normal transaction.

From the above two graphs, its initialized that the fraud transaction ratio is 40% and the ratio of the genuine transaction is higher than the fraud up to 90%. According to these two graphs, there are 51440 normal transactions seen and 150 fraudulent transactions. Correlation matrix and heatmap are helpful to analyse and visualize datasets and balanced datasets to unbalanced datasets.

# **Results discussion and analysis:**

In a nutshell, this research used a Logistic regression model to improve the credit card fraud detection method using a machine learning algorithm. After implementation of the model, it gives 99% accuracy to detect fraudulent and legitimate transactions that were good, and it improves the efficiency of the model as well as easily detects fraud transactions. Before implementation first analysed and visualized the datasets and set all the data into two parts such as the training dataset and the testing dataset. And then built a logistic regression model using the Python algorithm technique.

In this process first initialized the data was unbalanced so for that reason found the missing values and then encoded the dataset. After that split the dataset to perform a Logistic regression model algorithm. In this process, it gave 51440 fraudulent transactions and 150 legitimate transactions using matplotlib and seaborn library. Secondly, initialized how many transactions were done in a time and checked which one was a fraud and which one was a non-fraud transaction using correlation matrix and heatmap functions. According to the results, there are minimal transactions that detect fraud like 150 and the maximum number of transactions is 51440. Initially, this research analysed the using Logistic regression model it is easy to detect the fraud transaction and improve the detection process.

Thirdly, separated the datasets into two parts a training dataset and a testing dataset to analyse the accuracy of the model. After creating a model tested this model on these two datasets. Testing datasets generated 99% accuracy of the model and training datasets generated the same 99% accuracy of the model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dataset | Model | Recall | Precision | F1-score |
| Training | Logistic regression | 0.99 | 0.99 | 0.9 |
| Testing | Logistic regression | 0.99 | 0.99 | 0.9 |

Table 7.1: Results of the model performance

According to the above table, it can be seen that the Logistic regression model gives a recall and precision ratio of 0.99 which same in both training and testing datasets. In addition to this, F1-score is 0.9 which is very good, and therefore, model accuracy is best to detect fraud transactions. All three algorithms such as recall, precision, and accuracy score are calculated by the confusion matrix and equation of these three algorithms.

By analysing and visualizing the process accuracy of the model is very high and it should be a combination of the algorithms and matrix. Initially, the Logistic regression model is the traditional way of the machine learning algorithm which enables to detection of credit card fraud detections with a good accuracy score.

Furthermore, using a confusion matrix method it identifies the F1 score, precision score, and recall score that is shown above in the table. After executing the confusion matrix techniques the true positive value on the testing dataset is 10282, the false positive value is 9, the true negative value is 14, and the false negative value is 13. On the contrary, the true positive value of the training dataset is 227421, the false positive value is 33, the true negative value is 150, and the false negative value is 241. According to these results, the above table is generated and all ratios are the same on both the training and testing dataset. Hence, it is true that the Logistic regression model gives the exact result of credit card fraud detection as well as gives higher accuracy of the dataset.

# **Critical Reflection:**

This research project is based on a machine learning algorithm using Python coding with the help of a Logistic regression model to improve the credit card fraud detection process. Using this model, it was found the fraudulent and normal transactions were accurate and efficient. This project includes some algorithms to analyse the model accuracy such as the confusion matrix technique, then analysing and visualizing the datasets in a graph using correlation matrix and heatmap technique of algorithms. These are very strong algorithms to improve the credit card fraud detection process as well as help the model in the detection of credit card fraud effectively and reliably.

However, there are some future requirements to improve in this process to deal with the time duration. Though the Logistic regression model gives a good accuracy score, but in more time, therefore it will need some quick process to eliminate the time issue to detect credit card fraud transactions. Moreover, this model’s accuracy score is higher than the other model and classifiers, hence it is an efficient and good model for this research. It will help in the future to mitigate fraudulent transactions as it will be very useful for users as well as some industries such as banks, online websites, automobile industries, etc.

Furthermore, credit card fraud detection is a very essential part of human life as now every payment depends on an online basis. Owing to this, it will need to solve credit card fraud transactions. One paper by K. Venkatesan and K. Kumar suggests that credit card fraud detection in the banking industry is a python-based machine learning and compare with the accuracy of Logistic regression model techniques. It should be able to locate or anticipate fraud detection.

According to the existing issues and gaps the Logistic regression model gave 92 to 95% accuracy but now using machine learning algorithms its accuracy is increased and gives accurate results of the credit card fraud detections. Additionally, this model gives 99% of accuracy in this research project and gives efficient and reliable results from the datasets. In the future, this model will be very useful to detect fraud transaction and helps to mitigate the existing problems and gaps, as well as its, improves the process of this python-based machine learning algorithm.

# **Conclusion:**

It can be concluded that there are a number of methods, classifications, techniques, and algorithms used for credit card fraud detection but comparison to all these Logistic regression model using machine learning algorithm definitely improves the performance as well as the accuracy and easily detect the fraudulent credit card. If considering the precision, recall, and F1 score ratio the Logistic regression model gives the same output on both sides of training as well as the testing dataset. This research considers that credit card fraud detection is a cost-effective task, and this process was very long to detect credit card fraud transactions because of the unbalanced dataset. Finally, the main issue of credit card fraud detection is the performance and ability of the model accuracy that was proven and successfully done in this process.

From a future perspective, it will be initialized that the Logistic regression model does faster than now in this paper and will be given efficient results. Additionally, first will see the model capabilities, and see some datasets that will be perfect and not included with missing values.

# **References:**

Alfaiz, N.S. and Fati, S.M. (2022). Enhanced Credit Card Fraud Detection Model Using Machine Learning. *Electronics*, 11(4), p.662. doi:[10.3390/electronics11040662](file:///C:\Users\bhoom\Downloads\10.3390\electronics11040662).

Darwish, M. and belal (2022). APPLICABILITY OF SMART CONTRACTS FOR REAL ESTATE: A LITERATURE REVIEW. *Future Computing and Informatics Journal*, [online] 7(1). doi:[10.54623/fue.fcij.7.1.2](file:///C:\Users\bhoom\Downloads\10.54623\fue.fcij.7.1.2).

Fang, Y., Zhang, Y. and Huang, C. (2019). Credit Card Fraud Detection Based on Machine Learning. *Computers, Materials & Continua*, 61(1), pp.185–195. doi:[10.32604/cmc.2019.06144](file:///C:\Users\bhoom\Downloads\10.32604\cmc.2019.06144).

Garg, V., Chaudhary, S. and Mishra, A. (2021). ANALYSING AUTO ML MODEL FOR CREDIT CARD FRAUD DETECTION. *International Journal of Innovative Research in Computer Science & Technology*, 9(3). doi:[10.21276/ijircst.2021.9.3.5](file:///C:\Users\bhoom\Downloads\10.21276\ijircst.2021.9.3.5).

Ileberi, E., Sun, Y. and Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data*, 9(1). doi:[10.1186/s40537-022-00573-8](file:///C:\Users\bhoom\Downloads\10.1186\s40537-022-00573-8).

Madhu, S. (2022). REVIEW on FRAUD DETECTION in CREDIT CARD TRANSACTIONS USING MACHINE LEARNING TECHNIQUES. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 06(01). doi:[10.55041/ijsrem11522](file:///C:\Users\bhoom\Downloads\10.55041\ijsrem11522).

Marabad, S. (2021). CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING. *ASIAN JOURNAL OF CONVERGENCE IN TECHNOLOGY*, 7(2), pp.121–127. doi:[10.33130/ajct.2021v07i02.023](file:///C:\Users\bhoom\Downloads\10.33130\ajct.2021v07i02.023).

Naik, H. and Kanikar, P. (2019). Credit card Fraud Detection based on Machine Learning Algorithms. *International Journal of Computer Applications*, 182(44), pp.8–12. doi:[10.5120/ijca2019918521](file:///C:\Users\bhoom\Downloads\10.5120\ijca2019918521).

SONI, K.B. (2021). Credit Card Fraud Detection Using Machine Learning Approach. *Applied Information System and Management (AISM)*, 4(2), pp.71–76. doi:[10.15408/aism.v4i2.20570](file:///C:\Users\bhoom\Downloads\10.15408\aism.v4i2.20570).

Warghade, S., Desai, S. and Patil, V. (2020). Credit Card Fraud Detection from Imbalanced Dataset Using Machine Learning Algorithm. *International Journal of Computer Trends and Technology*, 68(3), pp.22–28. doi:[10.14445/22312803/ijctt-v68i3p105](file:///C:\Users\bhoom\Downloads\10.14445\22312803\ijctt-v68i3p105).

Yaqoob, M.A. (2022). Credit Card Fraud Detection Using Hybrid Approach of Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, 10(6), pp.3154–3159. doi:[10.22214/ijraset.2022.44468](file:///C:\Users\bhoom\Downloads\10.22214\ijraset.2022.44468).

Gao, J., Zhou, Z., Ai, J., Xia, B. and Coggeshall, S. (2019). Predicting Credit Card Transaction Fraud Using Machine Learning Algorithms. *Journal of Intelligent Learning Systems and Applications*, 11(03), pp.33–63. doi:[10.4236/jilsa.2019.113003](file:///C:\Users\bhoom\Downloads\10.4236\jilsa.2019.113003).

Kumar, T. (2021). Comparison of Logistic Regression and Decision Tree method for Credit Card Fraud Detection. *International Journal for Research in Applied Science and Engineering Technology*, 9(5), pp.680–683. doi:[10.22214/ijraset.2021.34241](file:///C:\Users\bhoom\Downloads\10.22214\ijraset.2021.34241).

Naik, H. and Kanikar, P. (2019). Credit card Fraud Detection based on Machine Learning Algorithms. *International Journal of Computer Applications*, 182(44), pp.8–12. doi:[10.5120/ijca2019918521](file:///C:\Users\bhoom\Downloads\10.5120\ijca2019918521).

patel, S. and Gond, S. (2014). Supervised Machine (SVM) Learning for Credit Card Fraud Detection. *International Journal of Engineering Trends and Technology*, 8(3), pp.137–139. doi:[10.14445/22315381/ijett-v8p225](file:///C:\Users\bhoom\Downloads\10.14445\22315381\ijett-v8p225).

Poojitha, S. and Malathi, K. (2022). An Innovative Method to Enhance the Accuracy of Credit Card Fraud Detection Using Logistic Regression Algorithm by Comparing Random Forest Algorithm. *ECS Transactions*, 107(1), pp.14205–14218. doi:[10.1149/10701.14205ecst](file:///C:\Users\bhoom\Downloads\10.1149\10701.14205ecst).

Porkess, R. and Mason, S. (2011). Looking at debit and credit card fraud. *Teaching Statistics*, 34(3), pp.87–91. doi:[10.1111/j.1467-9639.2010.00437.x](file:///C:\Users\bhoom\Downloads\10.1111\j.1467-9639.2010.00437.x).

Prakash, Dr.B. (2018). ATM Card Fraud Detection System using Machine Learning Techniques. *International Journal for Research in Applied Science and Engineering Technology*, 6(4), pp.5124–5129. doi:[10.22214/ijraset.2018.4836](file:///C:\Users\bhoom\Downloads\10.22214\ijraset.2018.4836).

Venkata Suryanarayana, S., N. Balaji, G. and Venkateswara Rao, G. (2018). Machine Learning Approaches for Credit Card Fraud Detection. *International Journal of Engineering & Technology*, 7(2), p.917. doi:[10.14419/ijet.v7i2.9356](file:///C:\Users\bhoom\Downloads\10.14419\ijet.v7i2.9356).

Aman (2021). Credit Card Fraud Detection using Machine Learning and Data Science. *International Journal for Research in Applied Science and Engineering Technology*, 9(VII), pp.3788–3792. doi:[10.22214/ijraset.2021.37200](file:///C:\Users\bhoom\Downloads\10.22214\ijraset.2021.37200).

Bhat K, S.K. (2020). Credit Card Fraud Detection using Machine Learning Methods. *International Journal for Research in Applied Science and Engineering Technology*, 8(6), pp.1436–1440. doi:[10.22214/ijraset.2020.6233](file:///C:\Users\bhoom\Downloads\10.22214\ijraset.2020.6233).

Bodepudi, H. (2021). Credit Card Fraud Detection Using Unsupervised Machine Learning Algorithms. *International Journal of Computer Trends and Technology*, 69(8), pp.1–3. doi:[10.14445/22312803/ijctt-v69i8p101](file:///C:\Users\bhoom\Downloads\10.14445\22312803\ijctt-v69i8p101).

Chou, T. and Lo, M. (2018). Predicting Credit Card Defaults with Deep Learning and Other Machine Learning Models. *International Journal of Computer Theory and Engineering*, 10(4), pp.105–110. doi:[10.7763/ijcte.2018.v10.1208](file:///C:\Users\bhoom\Downloads\10.7763\ijcte.2018.v10.1208).

Dandotia, S. and Tiwari, S.K. (2021). Detection of credit card fraud transactions using machine learning algorithms techniques with data driven approaches: a comparative study. *International Journal of Innovative Research and Growth*, 10(11). doi:[10.26671/ijirg.2021.11.10.101](file:///C:\Users\bhoom\Downloads\10.26671\ijirg.2021.11.10.101).

Gupta, M., Swati, Patel, S., Kumar, S. and Chauhan, G. (2020). ANOMALY DETECTION IN CREDIT CARD TRANSACTIONS USING MACHINE LEARNING. *International Journal of Innovative Research in Computer Science & Technology*, 8(3). doi:[10.21276/ijircst.2020.8.3.5](file:///C:\Users\bhoom\Downloads\10.21276\ijircst.2020.8.3.5).

ja, P. and sha, Dr.A. (2020). Review on Credit Card Fraud Detection using Machine Learning Algorithms. *International Journal of Computer Trends and Technology*, 68(6), pp.22–27. doi:[10.14445/22312803/ijctt-v68i6p104](file:///C:\Users\bhoom\Downloads\10.14445\22312803\ijctt-v68i6p104).

Kumar, G., Kumar, S. and Prakash, A.A. (2021). Credit Card Fraud Detection using Machine Learning. *International Journal of Engineering and Advanced Technology*, 10(4), pp.124–126. doi:[10.35940/ijeat.d2344.0410421](file:///C:\Users\bhoom\Downloads\10.35940\ijeat.d2344.0410421).

Madhu, S. (2022). REVIEW on FRAUD DETECTION in CREDIT CARD TRANSACTIONS USING MACHINE LEARNING TECHNIQUES. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 06(01). doi:[10.55041/ijsrem11522](file:///C:\Users\bhoom\Downloads\10.55041\ijsrem11522).

Mijwil, M.M. and Salem, I.E. (2020). Credit Card Fraud Detection in Payment Using Machine Learning Classifiers. *Asian Journal of Computer and Information Systems*, 8(4). doi:[10.24203/ajcis.v8i4.6449](file:///C:\Users\bhoom\Downloads\10.24203\ajcis.v8i4.6449).

Penmetsa, S.D. and Mohammed, S. (2021). Ensemble Techniques for Credit Card Fraud Detection. *International Journal of Smart Business and Technology*, 9(2), pp.33–48. doi:[10.21742/ijsbt.2021.9.2.03](file:///C:\Users\bhoom\Downloads\10.21742\ijsbt.2021.9.2.03).

Sharma, S., Mittal, P. and G, G. (2018). An Approach to Detect Credit Card Frauds using Attribute Selection and Ensemble Techniques. *International Journal of Computer Applications*, 180(21), pp.1–6. doi:[10.5120/ijca2018916482](file:///C:\Users\bhoom\Downloads\10.5120\ijca2018916482).

Venkata Suryanarayana, S., N. Balaji, G. and Venkateswara Rao, G. (2018). Machine Learning Approaches for Credit Card Fraud Detection. *International Journal of Engineering & Technology*, 7(2), p.917. doi:[10.14419/ijet.v7i2.9356](file:///C:\Users\bhoom\Downloads\10.14419\ijet.v7i2.9356).

VOICAN, O. (2021). Credit Card Fraud Detection using Deep Learning Techniques. *Informatica Economica*, 25(1/2021), pp.70–85. doi:[10.24818/issn14531305/25.1.2021.06](file:///C:\Users\bhoom\Downloads\10.24818\issn14531305\25.1.2021.06).

Alfaiz, N.S. and Fati, S.M. (2022). Enhanced Credit Card Fraud Detection Model Using Machine Learning. *Electronics*, 11(4), p.662. doi:[10.3390/electronics11040662](file:///C:\Users\bhoom\Downloads\10.3390\electronics11040662).

Alonso Robisco, A. and Carbó Martínez, J.M. (2022). Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction. *Financial Innovation*, 8(1). doi:[10.1186/s40854-022-00366-1](file:///C:\Users\bhoom\Downloads\10.1186\s40854-022-00366-1).

Alphy, M. (2020). A Literature Review on Different Types of Machine Learning Methods in Web Mining. *International Journal of Psychosocial Rehabilitation*, 24(1), pp.1761–1769. doi:[10.37200/ijpr/v24i1/pr200276](file:///C:\Users\bhoom\Downloads\10.37200\ijpr\v24i1\pr200276).

Altabrawee, H. (2018). Performance evaluation of fourteen machine learning algorithms on credit card default classification. *Journal of Education College Wasit University*, 1(26), pp.461–474. doi:[10.31185/eduj.vol1.iss26.103](file:///C:\Users\bhoom\Downloads\10.31185\eduj.vol1.iss26.103).

Ankara, N. and Sahinturk, H. (2019). Performance evaluation of ensemble learning algorithms on unbalanced credit scoring data sets. *Pressacademia*, 9(9), pp.180–185. doi:[10.17261/pressacademia.2019.1089](file:///C:\Users\bhoom\Downloads\10.17261\pressacademia.2019.1089).

Arri, H.S. (2022). Real-Time Credit Card Fraud Detection Using Machine Learning. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 06(04). doi:[10.55041/ijsrem12659](file:///C:\Users\bhoom\Downloads\10.55041\ijsrem12659).

Atiyah, O.S. and Thalij, S.H. (2022). Evaluation of COVID-19 Cases based on Classification Algorithms in Machine Learning. *Webology*, 19(1), pp.4878–4887. doi:[10.14704/web/v19i1/web19326](file:///C:\Users\bhoom\Downloads\10.14704\web\v19i1\web19326).

Aziz, S.F.A. (2020). Students’ Performance Evaluation Using Machine Learning Algorithms. *College Of Basic Education Research Journal*, 16(3), pp.977–986. doi:[10.33899/berj.2020.166006](file:///C:\Users\bhoom\Downloads\10.33899\berj.2020.166006).

Bekhet, S. and Alghamdi, A.M. (2021). A COMPARATIVE STUDY FOR VIDEO CLASSIFICATION TECHNIQUES USING DIRECT FEATURES MATCHING, MACHINE LEARNING, AND DEEP LEARNING. *Journal of Southwest Jiaotong University*, 56(4), pp.745–757. doi:[10.35741/issn.0258-2724.56.4.63](file:///C:\Users\bhoom\Downloads\10.35741\issn.0258-2724.56.4.63).

Darwish, M. and belal (2022). APPLICABILITY OF SMART CONTRACTS FOR REAL ESTATE: A LITERATURE REVIEW. *Future Computing and Informatics Journal*, [online] 7(1). doi:[10.54623/fue.fcij.7.1.2](file:///C:\Users\bhoom\Downloads\10.54623\fue.fcij.7.1.2).

Dimitri, A. and Talamo, M. (2018). The use of data mining and machine learning in nanomedicine: a survey. *Frontiers in Nanoscience and Nanotechnology*, 4(3). doi:[10.15761/fnn.1000s1004](file:///C:\Users\bhoom\Downloads\10.15761\fnn.1000s1004).

DROTÁR, P. and SMÉKAL, Z. (2014). COMPARATIVE STUDY OF MACHINE LEARNING TECHNIQUES FOR SUPERVISED CLASSIFICATION OF BIOMEDICAL DATA. *Acta Electrotechnica et Informatica*, 14(3), pp.5–10. doi:[10.15546/aeei-2014-0021](file:///C:\Users\bhoom\Downloads\10.15546\aeei-2014-0021).

F., S. (2018). Machine-Learning Techniques for Customer Retention: A Comparative Study. *International Journal of Advanced Computer Science and Applications*, 9(2). doi:[10.14569/ijacsa.2018.090238](file:///C:\Users\bhoom\Downloads\10.14569\ijacsa.2018.090238).

Fang, Y., Zhang, Y. and Huang, C. (2019). Credit Card Fraud Detection Based on Machine Learning. *Computers, Materials & Continua*, 61(1), pp.185–195. doi:[10.32604/cmc.2019.06144](file:///C:\Users\bhoom\Downloads\10.32604\cmc.2019.06144).

Gaikwad, S., Patel, S. and Shetty, A. (2020). Comparative Study of Machine Learning Techniques in the Medical Field. *International Journal of Computer Science and Mobile Computing*, 9(12), pp.59–67. doi:[10.47760/ijcsmc.2020.v09i12.008](file:///C:\Users\bhoom\Downloads\10.47760\ijcsmc.2020.v09i12.008).

Grefenstette, J.J. (1988). Credit assignment in rule discovery systems based on genetic algorithms. *Machine Learning*, 3(2-3), pp.225–245. doi:[10.1007/bf00113898](file:///C:\Users\bhoom\Downloads\10.1007\bf00113898).

Guendouz, M. and Amine, A. (2022). A Comparative Study of Machine Learning Techniques for Android Malware Detection. *International Journal of Software Innovation*, 10(1), pp.1–13. doi:[10.4018/ijsi.309719](file:///C:\Users\bhoom\Downloads\10.4018\ijsi.309719).

Hasan, Z. and Roy, N. (2021). Trending machine learning models in cyber‐physical building environment: A survey. *WIREs Data Mining and Knowledge Discovery*, 11(5). doi:[10.1002/widm.1422](file:///C:\Users\bhoom\Downloads\10.1002\widm.1422).

Hegde, R. (2022). Machine Learning Based Scholarship and Credit Pre-Assessment System. *Engineering and Technology Journal*, 07(01). doi:[10.47191/etj/v7i1.03](file:///C:\Users\bhoom\Downloads\10.47191\etj\v7i1.03).

Hüllermeier, E. (2011). Fuzzy machine learning and data mininga. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4), pp.269–283. doi:[10.1002/widm.34](file:///C:\Users\bhoom\Downloads\10.1002\widm.34).

Jadhav, S., Chandran, P. and Vijaykumar, S. (2019). Comparative Study of Chronic Kidney Disease Prediction using Machine Learning Techniques. *International Journal of Computer Sciences and Engineering*, 7(6), pp.501–506. doi:[10.26438/ijcse/v7i6.501506](file:///C:\Users\bhoom\Downloads\10.26438\ijcse\v7i6.501506).

Kibria, Md.G. and Sevkli, M. (2021). Application of Deep Learning for Credit Card Approval: A Comparison with Two Machine Learning Techniques. *International Journal of Machine Learning and Computing*, 11(4), pp.286–290. doi:[10.18178/ijmlc.2021.11.4.1049](file:///C:\Users\bhoom\Downloads\10.18178\ijmlc.2021.11.4.1049).

Kumar, Mr.V.N. (2022). Machine Learning Based Solution for an Effective Credit Card Fraud Detection. *Emperor Journal of Applied Scientific Research*, 04(09). doi:[10.35338/ejasr.2022.4907](file:///C:\Users\bhoom\Downloads\10.35338\ejasr.2022.4907).

Lavrač, N., Motoda, H. and Fawcett, T. (2004). Editorial: Data Mining Lessons Learned. *Machine Learning*, 57(1/2), pp.5–11. doi:[10.1023/b:mach.0000035515.33403.6b](file:///C:\Users\bhoom\Downloads\10.1023\b:mach.0000035515.33403.6b).

Lu, M. and Li, F. (2020). Survey on lie group machine learning. *Big Data Mining and Analytics*, 3(4), pp.235–258. doi:[10.26599/bdma.2020.9020011](file:///C:\Users\bhoom\Downloads\10.26599\bdma.2020.9020011).

Marabad, S. (2021). CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING. *ASIAN JOURNAL OF CONVERGENCE IN TECHNOLOGY*, 7(2), pp.121–127. doi:[10.33130/ajct.2021v07i02.023](file:///C:\Users\bhoom\Downloads\10.33130\ajct.2021v07i02.023).

Martens, D., Baesens, B. and Fawcett, T. (2010). Editorial survey: swarm intelligence for data mining. *Machine Learning*, 82(1), pp.1–42. doi:[10.1007/s10994-010-5216-5](file:///C:\Users\bhoom\Downloads\10.1007\s10994-010-5216-5).

Mijwil, M.M. and Salem, I.E. (2020). Credit Card Fraud Detection in Payment Using Machine Learning Classifiers. *Asian Journal of Computer and Information Systems*, 8(4). doi:[10.24203/ajcis.v8i4.6449](file:///C:\Users\bhoom\Downloads\10.24203\ajcis.v8i4.6449).

Naik, H. and Kanikar, P. (2019). Credit card Fraud Detection based on Machine Learning Algorithms. *International Journal of Computer Applications*, 182(44), pp.8–12. doi:[10.5120/ijca2019918521](file:///C:\Users\bhoom\Downloads\10.5120\ijca2019918521).

Nick, W., Asamene, K., Bullock, G., Esterline, A. and Sundaresan, M. (2015). A Study of Machine Learning Techniques for Detecting and Classifying Structural Damage. *International Journal of Machine Learning and Computing*, 5(4), pp.313–318. doi:[10.7763/ijmlc.2015.v5.526](file:///C:\Users\bhoom\Downloads\10.7763\ijmlc.2015.v5.526).

Oghenekaro, L.U. and Ugwu, C. (2016). A Novel Machine Learning Approach to Credit Card Fraud Detection. *International Journal of Computer Applications*, 140(5), pp.45–50. doi:[10.5120/ijca2016909316](file:///C:\Users\bhoom\Downloads\10.5120\ijca2016909316).

Oo, M.M. (2018). Comparative Study of MFCC Feature with Different Machine Learning Techniques in Acoustic Scene Classification. *International Journal of Research and Engineering*, 5(7). doi:[10.21276/ijre.2018.5.7.1](file:///C:\Users\bhoom\Downloads\10.21276\ijre.2018.5.7.1).

Rosshan, T.C. (2022). REVIEW on FRAUD DETECTION in CREDIT CARD TRANSACTIONS USING MACHINE LEARNING TECHNIQUES. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 06(01). doi:[10.55041/ijsrem11521](file:///C:\Users\bhoom\Downloads\10.55041\ijsrem11521).

S., M. (2020). Survey Paper on Fraud Detection in Medicare Using Machine Learning. *International Journal of Psychosocial Rehabilitation*, 24(5), pp.4170–4174. doi:[10.37200/ijpr/v24i5/pr2020130](file:///C:\Users\bhoom\Downloads\10.37200\ijpr\v24i5\pr2020130).

Santosh, T. and Ramesh, D. (2020). Machine Learning Approach on Apache Spark for Credit Card Fraud Detection. *Ingénierie des systèmes d information*, 25(1), pp.101–106. doi:[10.18280/isi.250113](file:///C:\Users\bhoom\Downloads\10.18280\isi.250113).

Soysal, M. and Schmidt, E.G. (2010). Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison. *Performance Evaluation*, 67(6), pp.451–467. doi:[10.1016/j.peva.2010.01.001](file:///C:\Users\bhoom\Downloads\10.1016\j.peva.2010.01.001).

Venkata Suryanarayana, S., N. Balaji, G. and Venkateswara Rao, G. (2018). Machine Learning Approaches for Credit Card Fraud Detection. *International Journal of Engineering & Technology*, 7(2), p.917. doi:[10.14419/ijet.v7i2.9356](file:///C:\Users\bhoom\Downloads\10.14419\ijet.v7i2.9356).

Wang, X., Zong, K. and Luo, C. (2022). Credit risk detection based on machine learning algorithms. *International Journal of Financial Services Management*, 1(1), p.1. doi:[10.1504/ijfsm.2022.10050887](file:///C:\Users\bhoom\Downloads\10.1504\ijfsm.2022.10050887).

Yu, N., Li, Z. and Yu, Z. (2018). Survey on encoding schemes for genomic data representation and feature learning—from signal processing to machine learning. *Big Data Mining and Analytics*, 1(3), pp.191–210. doi:[10.26599/bdma.2018.9020018](file:///C:\Users\bhoom\Downloads\10.26599\bdma.2018.9020018).

Zhang, D., Bhandari, B. and Black, D. (2020). Credit Card Fraud Detection Using Weighted Support Vector Machine. *Applied Mathematics*, 11(12), pp.1275–1291. doi:[10.4236/am.2020.1112087](file:///C:\Users\bhoom\Downloads\10.4236\am.2020.1112087).

# **Appendices**

Methodology of project process done in Jupyter on Python programming language.­

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Chart, waterfall chart

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Chart

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 1 Date/Time: 6th October at 1 PM**

**Agenda for meeting: meeting with a supervisor as well as the moderator to discuss**

**process of the project report.**

**Discussion of agenda items: One researcher faculty explains the rules and regulations as**

**well of the project as well as talked about how to create a good project report**

**Summary of agreed action plan: I understood the researcher faculty’s explanations and**

**tried to follow his structure to improve my MSC thesis.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 2 Date/Time: 12th October at 11 AM**

**Agenda for meeting: Discussion about how to collect the data online from the**

**website and which we will use for data collection.**

**Discussion of agenda items: Discussion about Kaggle, and Github websites as well as**

**how to collect from these websites. Sir explains practically how to collect the data.**

**Summary of agreed action plan: I understand the explanation perfectly of the data**

**collection process and will be applied to my research**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 3 Date/Time: 20th October at 12 PM**

**Agenda for meeting: Discussion about how to use Anaconda, Python, and Jupyter to**

**show my project deployment.**

**Discussion of agenda items: Practical evaluation of the Jupyter platform like how to**

**open, install, and how to use the Anaconda application, etc.**

**Summary of agreed action plan: I agreed with my supervisor and after that installed**

**the application on my pc and started to learn how to do deployment as my supervisor told me.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 4 Date/Time: 28th October at 12 PM**

**Agenda for meeting: Discussion about the methodology section as I have confusion**

**about what I can include in this chapter.**

**Discussion of agenda items: discussed the structure of the methodology, my supervisor**

**told me that see some papers that explain how to write and what can I include in the methodology.**

**Summary of agreed action plan: I understood the supervisor’s explanation and follow**

**his guidelines as I read some papers try to follow.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 5 Date/Time: 31st October at 12 PM**

**Agenda for meeting: Discussion about how to write a literature review in the report**

**Discussion of agenda items: My supervisor explains the literature review. Additionally,**

**he saw some literature reviews online from google scholar and explain how to write.**

**Summary of agreed action plan: I agreed with my supervisor and try to learn from the**

**paper how to write the literature review. Tried to implement it in my research.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 6 Date/Time: 2nd November at 11 AM**

**Agenda for meeting: discussion about how to write a final interim report. Like about**

**sections included in the interim report.**

**Discussion of agenda items: Discuss on follow project handbook as it is on aula and my**

**supervisor explains all the sections that will be included in the interim report such as the introduction, literature review, proposed methodology, and project planning.**

**Summary of agreed action plan: I understood the all sections clearly and try to**

**implement them in my interim report.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 7 Date/Time: 13th November at 12 PM**

**Agenda for meeting: Discussion about the feedback of the interim report with the supervisor**

**Discussion of agenda items: Discussed the need for changes in literature reviews,**

**introduction, as well as some points, to be included in the methodology chapter.**

**Summary of agreed action plan: I agreed with the changes and follow the instructions**

**of the supervisor as well as the moderator.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 8 Date/Time: 18th November at 10 AM**

**Agenda for meeting: Discussed methodology deployment section.**

**Discussion of agenda items: I am confused about what I need to include in the**

**methodology deployment section such as whether I include some screenshots of project deployment and add some equations or not.**

**Summary of agreed action plan: My supervisor explained about the methodology**

**deployment chapter. He told me that screenshots are required if it is going through Python coding and need to add an equation if the project development has some equation or not. I follow the instructions of the supervisor and implemented in my project report.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 9 Date/Time: 25th November at 12 PM**

**Agenda for meeting: Discussed results analysis and visualization and discuss results chapter**

**Discussion of agenda items: I am a little confused about these two sections. Therefore,**

**discussed what I wanted to write and what is the difference between the results analysis and visualization section as well as discuss results chapter.**

**Summary of agreed action plan: I understood the difference between the results**

**analysis and visualization section as well as discuss results chapter. Additionally, I learned that both are quite different, and tried to follow the supervisor’s instructions.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 10 Date/Time: 29th November at 12 PM**

**Agenda for meeting: Discussion about how to write a final report. Like what we can**

**include in our research report.**

**Discussion of agenda items: Discuss on follow project handbook as it is on aula and the**

**supervisor explains all the sections that will be included in the report such as abstract, project deployment, methodology, result discussion, conclusion, critical reflection, etc.**

**He explained the formation of the thesis report.**

**Summary of agreed action plan: I understood all sections clearly and try to**

**implement them in my research report. I also follow the project handbook as there are guidelines to complete the final report.**

**Notes:**

**MSc PROJECT (COMP11024)**

**PROJECT PROCESS DOCUMENTATION TEMPLATE**

**Student: Bhoomi Kaushal Sharma Supervisor: Dr. Raja Ujjan**

**Meeting Number: 11 Date/Time: 9th December at 12 PM**

**Agenda for meeting: Discussion about the critical reflection chapter.**

**Discussion of agenda items: In the previous meeting the supervisor explained all the**

**sections although I have confused about what I can write in the critical reflection section. So, the supervisor explained about critical reflection section.**

**Summary of agreed action plan: I understood the critical reflection section and**

**implement in my project’s final report.**

**Notes:**

University of the West of Scotland

School of Computing, Engineering and Physical Sciences

MSc Project Specification

Student name: Bhoomi Kaushal Sharma

Banner ID: B00733638

Email: B00733638@studentmail.uws.ac.uk

MSc Programme/stream: MSC IT

MSc Programme Leader: Dr. Costas Iliopoulos

Project Title:

|  |
| --- |
| A Python-based Logistic regression model to improve credit card fraud detection using a machine learning algorithm. |

Research Question to be answered:

|  |
| --- |
| How Python-based Logistic regression algorithm works to detect fraudulent credit card based on different features? |

Outline (overview) and overall aim of project:

|  |
| --- |
| * Nowadays, most of transactions are done with online payments, and in doing so person’s credit card details are stored online though it is not safe due to the hacking system. As a result, credit card holder’s details are hacked, and fraudulent transactions are done by another person without the permission of credit card owners. It is a very ethical issue in this era, so it is necessary to mitigate this issue. * It is essential to find out about anomalies in the system for some companies’ transactions with the use of credit cards. In today’s digitalization world, most credit card frauds doing online, and therefore it will be imperative to solve this issue and get safe as well as a trustable transaction for every person. * The major aim of this project is to build a credit card fraud detection model, which tells if the transactions made by a credit card are legitimate or fraudulent. Therefore, it is basically based on transaction and their labels as fraud or non-fraud detection processes if the new transaction made by the customer is fraud or not. Additionally, this research project gives two outputs such as which one is a true transaction, and which one is fraud. The model detects approximately 95% of accurate results and improves the quality of the result. * According to the existing literature reviews logistic regression model provides an accurate 95% result to show what is a fraud transaction and what is a real transaction. On top of that, the Logistic regression model very fast detects legitimate and fraudulent credit cards rather than other models. Moreover, this model is improving the visualization as well as the normalization process easy than other models or other algorithms. |

Objectives (list of tasks to be undertaken to achieve overall aim of the project and to answer the research question posed):

|  |
| --- |
| This research will be based on a python machine-learning algorithm by Logistic regression model. It will depend on the data sets provided that will be provided as secondary data sets. This research will investigate or recognize normal and fraudulent transactions.   1. This research will first collect the data sets from Kaggle and import them into Anaconda API in the Jupyter platform. Additionally, to perform datasets will use the CSV file, which is 143 MB data size. Furthermore, this project must employ Python as a computer language and it is a scripting language that is simple to use, interpreted, object-oriented programming, and a high-level language. 2. After this process will need to install libraries as it’s a member of Python libraries and the most common jobs are multi-dimensional groups. Add to this, it will use for information collecting and modification. 3. Secondly, after library installation, will need to perform explanatory data analysis on the dataset. This process cleans the data and makes it ready to train the model. It will help to balance the data from the imbalanced dataset. 4. Then, will separate the normal and fraud transaction and analyze and visualize the fraud and normal data. And then this data will be divided into portions such as the training phase and testing phase. After this, data will be split into features and labels. 5. Thirdly, need to create a Logistic regression model using a machine learning algorithm. After fitting data into the model will perform a model evaluation to check the model’s accuracy. In order to perform this process first model test data on training data and then test data on testing data. Both training and testing data will give model accuracy and model efficiency. 6. Lastly, after evaluating this process will analyze and visualize data using tables and charts. These will include some results such as showing normal and fraudulent transactions, how many transactions will be fraudulent or legitimate, how many transactions will be done etc. To perform these tasks will use Matplotlib tools as it is easy to use and show perfect tables and graphs. 7. Additionally, this research will recognize the fraud credit card detection, therefore, will add Python-based algorithm coding in this project to show how to implement the method step by step. |

Relationship of proposed project to MSc programme/stream:

|  |
| --- |
| This project is mainly based on the machine learning algorithm using Python as how it works to detect fraudulent credit cards. Additionally, credit card fraud is mostly related to the subject of Mobile Technology as learned new mobile technology growing day by day and because of this credit card fraud is significantly soaring in this digitalization world. Furthermore, Mobile technology also learned how to solve this issue using new technology and mitigate this problem in the future.  However, from the research design and methodology subject learned how to do research step by step and how to build the specification of the project. Moreover, it explains how to write aims, objectives, abstract, literature review, proposed method, and data analysis. On top of that this project is mostly related to research methods and it will be very helpful to develop research skills as well as give knowledge about how to write the dissertation. |

Indicative reading list and resources:

|  |
| --- |
| **References:**  Alfaiz, N.S. and Fati, S.M. (2022). Enhanced Credit Card Fraud Detection Model Using Machine Learning. *Electronics*, 11(4), p.662. doi:[10.3390/electronics11040662](file:///C:\Users\bhoom\Downloads\10.3390\electronics11040662).  Darwish, M. and belal (2022). APPLICABILITY OF SMART CONTRACTS FOR REAL ESTATE: A LITERATURE REVIEW. *Future Computing and Informatics Journal*, [online] 7(1). doi:[10.54623/fue.fcij.7.1.2](file:///C:\Users\bhoom\Downloads\10.54623\fue.fcij.7.1.2).  Fang, Y., Zhang, Y. and Huang, C. (2019). Credit Card Fraud Detection Based on Machine Learning. *Computers, Materials & Continua*, 61(1), pp.185–195. doi:[10.32604/cmc.2019.06144](file:///C:\Users\bhoom\Downloads\10.32604\cmc.2019.06144).  Garg, V., Chaudhary, S. and Mishra, A. (2021). ANALYSING AUTO ML MODEL FOR CREDIT CARD FRAUD DETECTION. *International Journal of Innovative Research in Computer Science & Technology*, 9(3). doi:[10.21276/ijircst.2021.9.3.5](file:///C:\Users\bhoom\Downloads\10.21276\ijircst.2021.9.3.5).  Ileberi, E., Sun, Y. and Wang, Z. (2022). A machine learning based credit card fraud detection using the GA algorithm for feature selection. *Journal of Big Data*, 9(1). doi:[10.1186/s40537-022-00573-8](file:///C:\Users\bhoom\Downloads\10.1186\s40537-022-00573-8).  Madhu, S. (2022). REVIEW on FRAUD DETECTION in CREDIT CARD TRANSACTIONS USING MACHINE LEARNING TECHNIQUES. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 06(01). doi:[10.55041/ijsrem11522](file:///C:\Users\bhoom\Downloads\10.55041\ijsrem11522).  Marabad, S. (2021). CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING. *ASIAN JOURNAL OF CONVERGENCE IN TECHNOLOGY*, 7(2), pp.121–127. doi:[10.33130/ajct.2021v07i02.023](file:///C:\Users\bhoom\Downloads\10.33130\ajct.2021v07i02.023).  Naik, H. and Kanikar, P. (2019). Credit card Fraud Detection based on Machine Learning Algorithms. *International Journal of Computer Applications*, 182(44), pp.8–12. doi:[10.5120/ijca2019918521](file:///C:\Users\bhoom\Downloads\10.5120\ijca2019918521).  SONI, K.B. (2021). Credit Card Fraud Detection Using Machine Learning Approach. *Applied Information System and Management (AISM)*, 4(2), pp.71–76. doi:[10.15408/aism.v4i2.20570](file:///C:\Users\bhoom\Downloads\10.15408\aism.v4i2.20570).  Warghade, S., Desai, S. and Patil, V. (2020). Credit Card Fraud Detection from Imbalanced Dataset Using Machine Learning Algorithm. *International Journal of Computer Trends and Technology*, 68(3), pp.22–28. doi:[10.14445/22312803/ijctt-v68i3p105](file:///C:\Users\bhoom\Downloads\10.14445\22312803\ijctt-v68i3p105).  Yaqoob, M.A. (2022). Credit Card Fraud Detection Using Hybrid Approach of Machine Learning. *International Journal for Research in Applied Science and Engineering Technology*, 10(6), pp.3154–3159. doi:[10.22214/ijraset.2022.44468](file:///C:\Users\bhoom\Downloads\10.22214\ijraset.2022.44468). |

Marking scheme:

|  |
| --- |
| Introduction 5%  Literature review (that sets the academic/theoretical underpinning) 20%  Methods (includes methods of implementation used for the mentioned protocols) 20%  Data analysis (includes the analysis of results collected through implementation) 15%  Discussion (and linkage to academic/theoretical underpinning) 20%  Conclusions and recommendations 10%  Critical self-evaluation 10% |

Supervisor:

|  |
| --- |
| Dr. Raja Ujjan |

Moderator:

|  |
| --- |
| Dr. Volkan Tunali |

Programme Leader:

|  |
| --- |
| Dr. Daune West |

Date specification submitted:

|  |
| --- |
| 16/09/2022 |

Please complete the ‘ethics’ form below for all projects.

**School of Computing, Engineering and Physical Sciences**

**MSc PROJECT – REQUIREMENT FOR ETHICAL APPROVAL**

**SECTION 1: TO BE COMPLETED BY THE STUDENT**

Does your proposed research involve: research with human subjects (including requirements gathering and product/software testing), access to company documents/records, questionnaires, surveys, focus groups and/or other interview techniques? Does your research entail any process which requires ethical approval? (please enter √ in the appropriate box)

|  |  |  |
| --- | --- | --- |
| YES |  | **You must submit an application for approval to the Ethics Review Manager** |
| NO | √ | You do not need to submit an application to the Ethics Review Manager |

**Name of Student (Print name):**

**Signature:**

**Date:**

**SECTION 2: TO BE COMPLETED BY THE PROJECT SUPERVISOR**

I understand that the above project does not require\* ethical approval.

**Supervisor (print name):** Raja Ujjan

**Signature**: A close-up of a sword

Description automatically generated with low confidence

**Date: 23/09/22**

**IMPORTANT: please note that by signing this form all signatories are confirming that any potential ethical issues have been considered and, where necessary, an application for ethical approval has been/will be made via the Ethical Review Manager software.**

**Any project requiring ethical approval but which has not been given approval will not be accepted for marking.**

**Ethical approval cannot be sought in retrospect.**