International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

CREDIT CARD FRAUD DETECTION USING MACHINE LEARNING TECHNIQUES

[1]Vineet Pathak , [2]Tariq Siddiqui , [3]Jeetendra singh yadav

1 M.TECH Scholar, Bhabha University, Bhopal
2 Asst. Professor CSE DEPARTMENT, Bhabha University, Bhopal
3 HOD CSE DEPARTMENT, Bhabha University, Bhopal

Email : Vineetpathak48@gmail.com, tariq.qazi007@gmail.com , jeetendra2201@gmail.com

Abstract- In this paper we deploy machine learning application on cloud on Google cloud Engine (By Google) for classification techniques which takes dataset of credit card transactions containing 284,807 transactions records and these techniques has capability to process the numerical and categorical datasets. In these we can take decision tree and random forest classifier and the classification result of these model is then compared the performance measure of these models.

Index Terms- fraud Random forest algorithm, decision tree algorithm, fraud detection cloud computing, private cloud, machine learning application

I. INTRODUCTION

The digital payments market is soaring as the world shifts towards online and card-based payment methods at a faster rate. With such a shift comes the growing issue of cyber security and fraud, which is more common than ever. According to a recent report, credit card fraud within the next 5 years will cause global losses of about $43 billion. Another study revealed that as many as 80% of the US credit cards currently in use have been compromised.

Enhancing credit card fraud detection is a priority for all banks and financial organizations. Thanks to machine learning (ML), credit card fraud detection is becoming easier and more efficient.* ML-based fraud detection solutions can track patterns and prevent abnormal transactions.

Machine learning models can recognize unusual credit card transactions and fraud.

II. CREDIT CARD FRAUD

Here we are focusing on Credit Card Fraud, as the world grows the internet use in every sector, shopping is also one of the favourite subject over internet that made possible by e-commerce as the use of online shopping the intruders got new way for their malicious works that is known as the frauds and those frauds intended by credit card or bank account details [4]. The cheats of Credit Card can be named as Credit card blackmail can be supported, where the affirmed customer themselves estimates a portion to another record which is compelled by

a hoodlum, or unapproved, where the record holder doesn't offer endorsement to the portion to proceed and the trade is finished by an untouchable or by getting to Master card nuances or record nuances.
Credit Card fakes has been ordered into two sorts:

- Offline extortion is submitted via way of means of using a taken real card at name attention or a few different spot.
- On-line fraud is in which card holder is not present and committed via shopping, internet, phone, web.

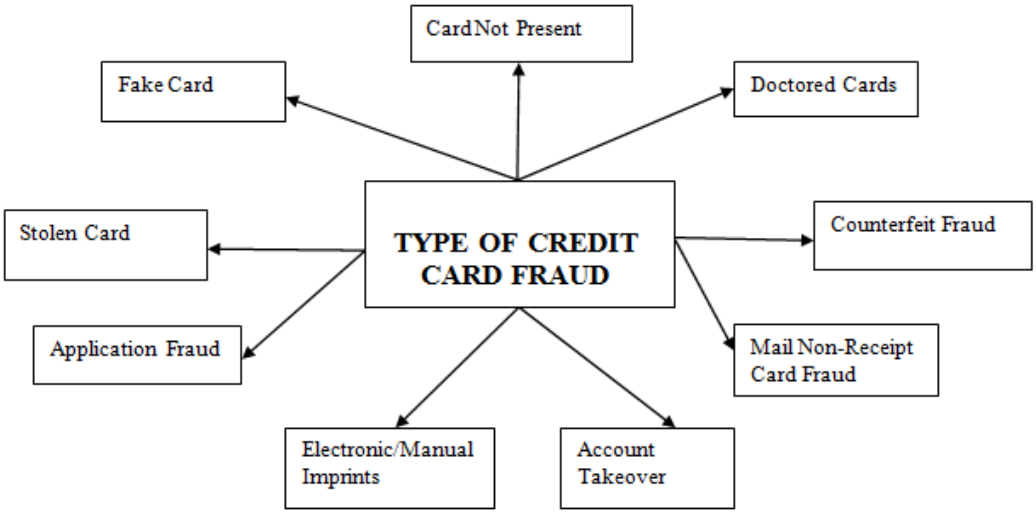

Figure 1.1 Type of Credit Card Fraud

III. DATA SET

Context: By recognizing the fraudulent credit card transactions, the credit card companies should make their customers not to bear any charge for the article that they are not intended to purchase.

Content: In September 2013, the Credit card datasets is made that contains transactions made by European cardholders. The transactions in the dataset are happened in 2 days, in which 492 fake transactions from 284,807 total exchanges. The dataset is profoundly lopsided, the positive class (cheats) represent 0.172% of all exchanges. It contains just mathematical information factors which are the consequence of a PCA transformation. Principal segment analysis (PCA) is a procedure utilized for ID of fewer uncorrelated factors known as head segments from a bigger arrangement of information. Shockingly, because of secrecy issues, the first features and more foundation data about the information can't be given.

Highlights V1, V2……V28 are the crucial elements were given with PCA, the lone highlights that have now no longer been modified with PCA are ‘Time’ and ‘Amount’. Highlight ‘Time’ incorporates the seconds exceeded among each change and the principle change withinside the dataset. The element ‘Amount’ is the

change Amount, this element may be applied for example dependant price sensitive learning. Highlight 'Class' is the response variable and it takes esteem 1 if there have to get up an incidence of misrepresentation and zero in case.

IV. LITERATURE REVIEW

In [13] Along sides the incredible expansion in Mastercard exchanges, charge card extortion has become progressively widespread as of late. In Modern day the misrepresentation is one of the significant reasons for extraordinary monetary misfortunes, not just for shippers, singular customers are likewise influenced. Three techniques to recognize extortion are introduced. First, the group modeling is used to group legitimate and fake exchanges using information, clusterization of locales of boundary esteem. Besides, Gaussian combination model is utilized to display the likelihood thickness of Master card client's previous conduct so that the likelihood of Current behavior can be determined to identify anomalies from past behavior, and ultimately Bayesian organizations are used to represent a particular client's ideas and measurements of various blackmail situations. The fundamental assignment is to investigate various perspectives on the same issue and see what can be gained from the use of each extraordinary strategy.
There is various fraud detection technologies is present through which they can identify the credit card frauds. In [3] they discussed about various problem challenges and issues they are facing in detecting frauds. Such as application frauds on which there is identity frauds is useful for North American nation to unravel the matter of master card fraud. The add [3], Application fraud when the customer apply for the credit card and there is data mismatch problems like multiple applications are submitted by one user with only one set of user details is known as duplication fraud[8]. This paper investigates ongoing writing on all the previously mentioned themes as they identify with distributed computing and looks at various techniques which propose to utilize AI to either take into account more powerful asset the board, better energy productivity, or higher security. Also, the proposed strategies are contrasted with each other to show their specific qualities and shortcomings, to permit further work to expand upon the ends came to and to propose persistently improving techniques. AI to either take into account more powerful asset the board, better energy productivity, or higher security. Also, the proposed strategies are contrasted with each other to show their specific qualities and shortcomings, to permit further work to expand upon the ends came to and to propose persistently improving techniques.

According to [10] They used hidden mark off model for detecting the credit card frauds and its very effective also on detecting attacks [9]. In these there is a profile analyzer who continuously scans or analyze the group of transactions sequence and detect if there is any fraud sequence is coming based on the cardholder's past behavior. In this paper, we have proposed a utilization of HMM in Master card extortion identification. The various strides in credit card exchange handling are addressed as the fundamental stochastic interaction of a HMM. We have utilized the scopes of exchange sum as the perception images, while the kinds of thing have been viewed as states of the HMM. We have proposed a technique for finding the spending profile of cardholders, just as utilization of this information in choosing the worth of perception images and starting evaluation of the model boundaries. It has likewise been clarified how the HMM can identify whether an approaching exchange is fake or not. Test results show the exhibition and viability of our framework and show the convenience of learning the spending profile of the cardholders. Near examinations uncover that the Accuracy of the framework is near 80% over a wide variety in the info information. The framework is too versatile for taking care of huge volumes of exchanges.

In [16] Currency extortion is an ever evolving threat with great results in the money business. Information mining had played a fundamental role in locating counterfeit credit cards on online exchanges. First, the profiles of typical and fraudulent practices are constantly changing, and the collections of credit card extortion information are also exceptionally skewed. s) used. This paper researches the presentation of innocent bayes, neighbor of K-closets and calculated relapse on profound slant for the information of card is chargeable. Dataset of Visa exchanges is sourced from European cardholders containing 284,807 exchanges. A mixture method of under- testing and over sampling is done on the slanted information. The three procedures are applied on the crude and preprocessed information. The project is carried out in Python. The exhibition of the strategies is assessed dependent on exactness, affectability, particularity, accuracy, Matthew's relationship coefficients and adjusted characterization rate. The outcomes shows of ideal precision for the neighbor of k- closet, guileless bayes and calculated relapse classifiers. The similar outcomes show that k-closest neighbor performs better compared to guileless bayes and calculated relapse methods.

In [14] The objective of information examination is to portray covered up examples and use them to help educated choices in an assortment regarding circumstances. Charge card misrepresentation is raising altogether with the headway of the modernized innovation and become an obvious objective for deceitful. Visa extortion is an extreme issue in the monetary assistance and costs billions of a dollar consistently. The plan of misrepresentation recognition calculation is a difficult errand with the absence of genuine exchange dataset in light of privacy and the profoundly imbalanced freely accessible datasets. In this paper, we apply distinctive directed AI calculations to identify Visa false exchange utilizing a genuine world dataset. Besides, we utilize these calculations to carry out a super classifier utilizing troupe learning techniques. We distinguish the main factors that may prompt higher exactness in Mastercard deceitful exchange discovery. Also, we analyze and talk about the presentation of different regulated AI calculations exist in writing against the super classifier that we carried out in this paper.

In [17] Advancement of correspondence advances and internet business has made the Mastercard as the most well-known strategy of installment for both on the web and customary buys. So, security in this framework is profoundly expected to forestall misrepresentation exchanges. Extortion exchanges in charge card information exchange are expanding every year. Toward this path, specialists are too attempting the novel methods to identify and forestall such cheats. In any case, there is consistently a need of certain procedures that ought to accurately and effectively distinguish these cheats. This paper proposes a plan for identifying fakes in Mastercard information which utilizes a Neural Network (NN) based unaided learning method. Proposed technique outflanks the current approaches of Auto Encoder (AE), Local Outlier Factor (LOF), Seclusion Forest (IF) and K-Means grouping.

In [21] At the point when an application sent in the cloud faces evolving responsibility, the administrations of the application need increasing or down accordingly. The administrations run on Virtual Machines (VM) or holder examples. Application Providers (APs) settle on how the applications are scaled through VM provisioning and through the arrangement of the administrations on those VMs. Different drivers guide this dynamic. Application execution and cost are two such

drivers. In this paper, we answer the topic of how APs can meet the exhibition limitations of their applications while limiting the expense of the running VMs. A VM provisioning issue is formed which hopes to meet mean reaction time imperatives and limit the expense, where VM-types having distinctive expense rates are utilized. The proposed arrangement depends on hereditary calculation what's more, bottleneck strength esteem. For the

contextual investigation, a choice creator is carried out for a web application. The proposed arrangement is looked at against a thorough pursuit, a basic hereditary calculation and an arbitrary inquiry. It is shown that our answer is capable meet reaction time imperatives with close ideal minimization of cost. The arrangement additionally results in better cost than irregular hunt and the plain hereditary calculation arrangement to the detriment of marginally longer runtime. A contextual analysis of a bicycle courses web application model utilizing the chief has been introduced in this paper. The execution model from the application is determined first and taken care of into the chief. In view of the client responsibilities, the chief chooses an organization setup, comprising of errand to-VM assignments, to meet mean execution imperatives and to limit cost.

In [22] In request to analyze malignant movement that happens in an organization or a framework, interruption recognition framework is utilized. Interruption Detection is programming or a gadget that checks a framework or an organization for an incredulous movement. Due to the developing availability between PCs, interruption identification gets indispensable to perform network security. Different machine learning procedures and factual philosophies have been used to assemble various kinds of Intrusion Detection Systems to ensure the organizations. Execution of an Intrusion Location is principally relies upon exactness. Exactness for Interruption location should be improved to lessen bogus alerts also, to build the location rate. To improve the execution, various strategies have been utilized in later works. Dissecting immense organization traffic information is the primary work of interruption identification framework. An efficient characterization system is needed to conquer this issue. This issue is adopted in proposed strategy. AI strategies like Support Vector Machine (SVM) and Naïve Bayes are applied. These procedures are notable to tackle the characterization issues. For assessment of interruption location framework, NSL– KDD information disclosure Dataset is taken. The results show that SVM works better compared to Naïve Bayes. To perform near examination, compelling characterization strategies like Support Vector Machine and Credulous Bayes are taken, their exactness and misclassification rate get determined. Execution of an Intrusion Location is principally relies upon exactness. Exactness for Interruption location should be improved to lessen bogus alerts also, to build the location rate. To improve the execution, various strategies have been utilized in later works. Dissecting immense organization traffic information is the primary work of interruption identification framework. An efficient characterization system is needed to conquer this issue.

Problem Using Cloud

The cloud depends on the Internet Protocol (IP), so for an application to be thought of, it should utilize IP as its correspondence component. While there are many protocols that can be run over IP, the use of Transport Control Protocol (TCP) is preferred [9]. The security issue has assumed the main part in impeding Cloud figuring. Irrefutably, putting your information, running your thing at another person's hard disk utilizing another person's CPU has all the earmarks of being overwhelming to various. Striking security issues like data adversity, phishing, botnet (running remotely on an arrangement of machines) present authentic perils to affiliation's data and programming because in cloud every time we interface with the virtual machine another IP address machine will assigned [2].

Problem in Credit Card Data Classification

The technique which is effective on detection of frauds have faces such type of problems for achieving the result which will be best [2].

Data Imbalance: The nature of dataset of the transactions of credit card is imbalanced

Importance Different misclassification: The different importance of errors in misclassification is there.

Data Overlapping: It is very hard to maintaining a low false negative and false positive rate. Because there are many numbers of transaction in the dataset which considered as frauds it may normal and not to be normal.

V. PROPOSED WORK

We propose private cloud preferred over others because cloud service is provided by third party providers so for security reason private cloud give better security than others because the connection between user and virtual machine is secured by ssh. And on private cloud we can easily scale the storage and processing power at any time whenever application required. The secure, high-availability Web application is up and running. When the application needs to be updated, the virtual machine images can be updated, copied across the development chain, and the entire infrastructure can be redeployed.

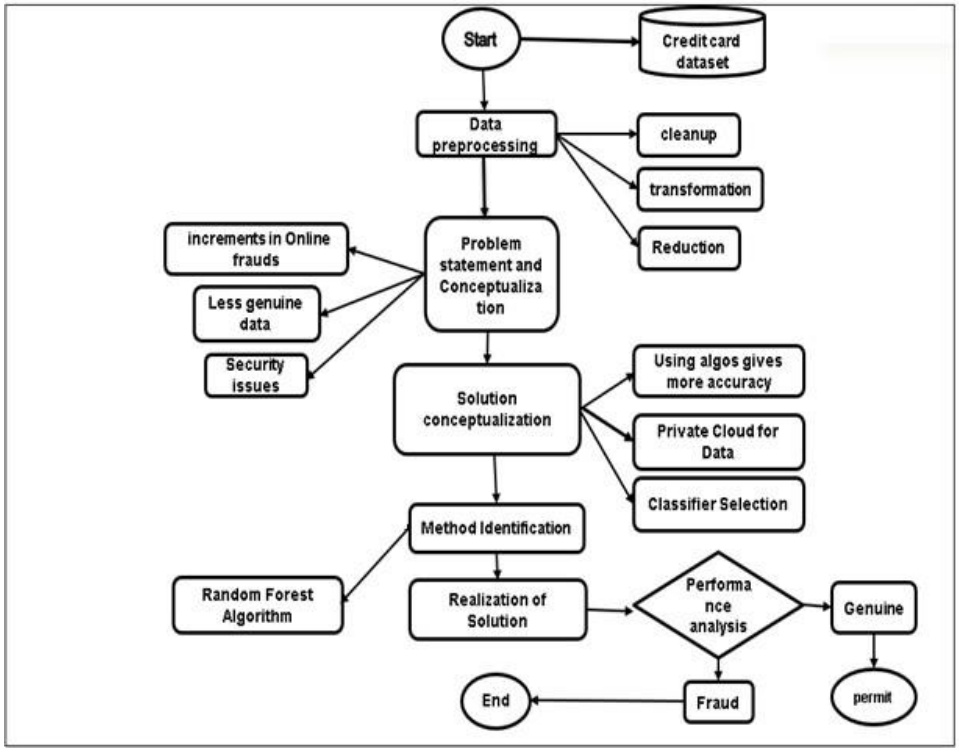

Figure 4.1 Proposed Flow Diagram

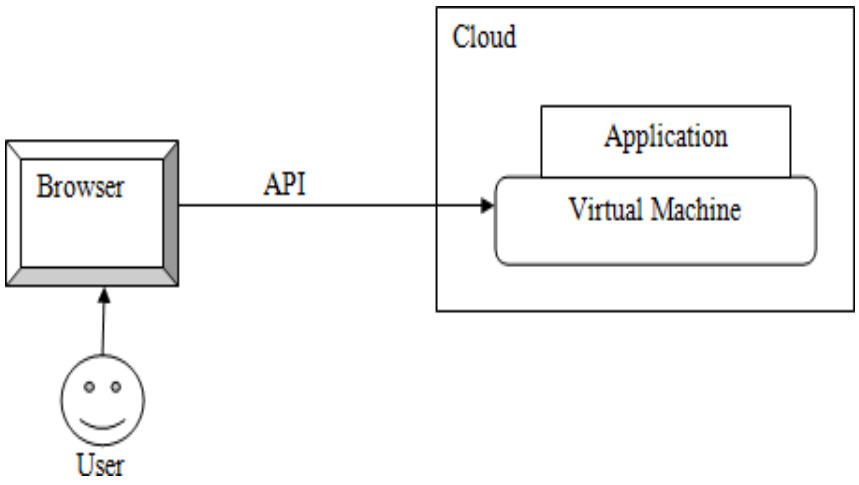

Figure 4.2 Proposed Block Diagram

Step-1. First user will access the browser and login into the cloud service website.

Step-2. After Successful login a secure ssh connection has been developed between user browser and virtual machine on cloud through API call.

Step-3. We can deploy a machine learning application on virtual machine over cloud.

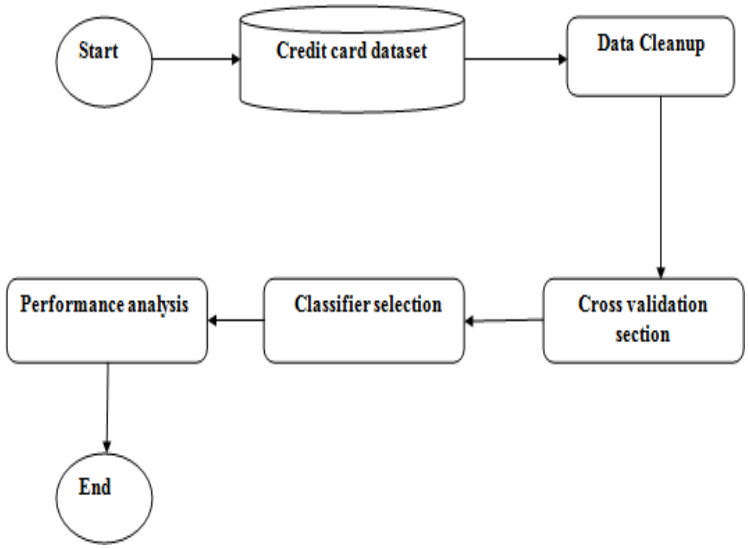

Figure 4.3 Proposed Model

Data Cleanup

The information which we get contains un-important, boisterous or inadequate data which we influence the models for anticipating the results. In this way, before these we can pre-measure the information and eliminate these pointless information from the datasets. In these the information preprocessing steps comprises an information joining and information change [7]. In information mix we can changes over the information types dependent on the models and in these we can giving an appropriate properties names which will helps in highlight choice of the models and the specialist eliminates the missing qualities records.

Information change is utilized to lessen the overhead the changing the information types at the run time and, additionally eliminate un-important qualities which won't valuable in the preparation of the model.

Data Classification

At the point when the information pre-handling is finished, we can get the prepared information which is utilized to prepare the order or indicator models [8]. There are a few calculations are there like neural organizations, administered and unsurprised learning calculations which will be accustomed to preparing and framing a model and subsequent to preparing is finished the models order the records into the different names.

The both processes are required before we gave dataset for training through machine learning. Because dataset may contain null values and default values, that can create problem during training and when we test data there may be possibilities of errors.

ALGORITHM

Random Forest Algorithm

First, Random Forest formula may be a supervised classification formula. We are able to see from its name, this is to make a forest by away and make it random. There's an immediate relationship between the quantities of trees within the forest and therefore the results it will get: the larger the quantity of trees, a lot of correct the result [1]. However, one issue to notice is that making the forest isn't constant as constructing the choice with data gain or gain index approach.

The one mentioned repeatedly by the author is that it is often used for each classification and regression tasks. Over fitting is one essential drawback that will build the results worse, except for Random Forest formula, if there square measure enough trees within the forest. The third advantage is that the classifier of Random Forest will handle missing values, and therefore the last advantage is that the Random Forest classifiers are often sculptures for categorical values.

We assume that the user is aware of regarding the development of single classification trees. Random Forests grows several classification trees. To classifying a brand-new object from Associates input vector, place the input vector down every of the trees within the forest. Every tree provides a classification, and that we say the tree "votes" for that category. The forest picks the order having the chief votes (over every one of the trees inside the forest).

Each tree is big as follows:

for S = 0, .., W do
Hi ← Bootstrap sample from H Ti ← Contruct tree using Hi for node = 1, .., No.Nodes do
nodei ← choose random subset m of all features. end for
end for
X ← take the majority vote for all trees

Decision Tree Algorithm

Algorithm creates a tree structured based on the various attributes and the leaf nodes denotes the classification result. The algorithm selects the best attributes which will help in the decision tree making and after creating a tree is will classify the records. There are several decision tree based algorithm [1]. This algorithmic rule may auspiciously handle the missing attributes and helps in pruning the trees once creation, wherever within the size of call tree is reduced by removing branches/leaves that offer little or no power to classify instances. The trees are made during a top- down algorithmic divide-and-conquer manner. During call tree, when the choices are created, leaf nodes have 2 values. The primary worth demonstrates the whole variety of instances reaching the leaf and second worth indicates the amount of misclassified instances. Within the case of missing values, fragmentary values are exhibited at the leaves. This tree by default uses 10-fold cross validation that states that ninetieth of the info is employed for coaching and 100 percent for testing. As ninetieth isn't too removed from 100 percent, it provides affair estimate of the worth. These cross- validation folds are often switched at the "Test options". The accuracy achieved with a choice tree a lot of much more than different data processing techniques that clearly states that call tree is associate degree economical technique.

Each tree is big as follows:

Input:

Data partition, T, which is a set of training tuple and their related class labels; Attribute_list;
Attribute_selection_method, to determine the splitting criterion that "best" partition.

Output: A decision tree.

Step 1- CREATING A ROOT NODE
1. Create a root node M
2. If tuples in T are all of the similar class, A then

3. Return M as a leaf node label with the class A;
4. If attribute list is empty then
5. Return M as a leaf node label with the majority class in T

Step 2- ATTRIBUTE SELECTION
6. Apply attribute_selection_method(T, attribute_list) to discover the “best “ splitting _criterion attribute;
7. Label node M with splitting _criterion;
8. Update the attribute_list

Step 3- SPLIT THE TREE
9. for each outcome j of splitting_criterion
//partition the tuples and produce subtrees for each partition 10.Based on splitting_criterion attribute Split the tree into two part
12. attach a leaf labeled with the majority class T in node M:
13. else attach the node returned by Generate_decision_tree(Tj:attribute_list)to node M:
end for
14. return M,

For experiment we can take an virtual machine on cloud preferred over others because cloud service is provided by third party providers (Google cloud platform (GCP)) so for security reason private cloud give better security than others because the connection between user and virtual machine is secured by ssh. And on private cloud we can easily scale the storage and processing power at any time whenever application required.
After clicking on start button, out virtual machine gets started and to with these we click on ssh which is nothing but a channel created between browser and virtual machine terminal that is secure. ssh gives a secure login so we can securely access our resources.
Presently we will carry out the Random Forest Algorithm tree utilizing Python. For this, we will utilize the equivalent dataset. By utilizing the equivalent dataset, we can contrast the Random Forest classifier and other arrangement models, for example, Decision tree Classifier, KNN, SVM, Logistic Regression, and so forth.

After that we can connect the python and jupyter notebook (python IDE for Machine Learning) through web browser using a secure ip and port number. when we go through the URL it can ask for secure token or password using which we can access other notebooks so after giving it the secure token we can connect with the jupyter notebook.

Deploying ML Classification over Cloud Machine

For test we will loaded the dataset into the Python the dataset of credit card facts set includes 31 variables over 284807 observations.
Mix and split the information into preparing information and test information. Reorder the accompanying code into the following code cell and pick Run.

The preparation information 70% of clients is utilized during the model preparing. We use inclination based advancement to iteratively refine the model boundaries. Inclination based enhancement is an approach to discover model boundary esteems that limit the model mistake, utilizing the angle of the model misfortune work.
The test information staying 30% of clients is utilized to assess the exhibition of the model and measure how well the prepared model sums up to inconspicuous information.

MODELS TRANING

Using Random Forest Model

Random forest model is an ensemble of classification (or regression) trees. The data mining merges from adaptability, convenience, that make decision tree algorithm popular, as far as deal with different types of information attributes and interpretability. However, a single tree model may be unreliable and too sensitive to prepare clear information point. When the individual instances are scattered and the random forest uses two randomness critical points between individual trees to become different, clustering can work well: first, each tree is based on an isolated test of preparing information seeds; in addition, each center Only consider a randomly selected set of information, namely individual trees in the building. Therefore, random forest combines packaging ideas, in which a single model in a group is created by learning and replacing prepared information, and a random subspace method, in which each tree is constructed in a set by a subset of random attributes.

1. On the off chance that the amount of cases inside the instructing set is N, test N cases aimlessly - anyway with substitution, from the primary information. This example will be the instructing set for developing the tree.

2. On the off chance that there are M information factors, assortment m.

Using Decision Tree Model

A decision tree is a technique for extracting facts that consistently express autonomous attributes and reliable quality, and for constructing trees, is IF-THEN connections and each test must be successful if any standard is to be established. The decision tree usually isolates the complicated problem into several individual problems and solves the sub-problems more than once. from findings to possible outcome.

Accuracy of Models

The results of Random forest algorithm and decision tree algorithm over the given dataset is shown in the tabular form. And both the model accuracy has little difference but when on the huge real time data, it will secure thousands of cases from the fraud.

Table-1 Accuracy of Models

	Model	Accuracy	Precision	Recall	F1 Score
0	Decision tree	0.999333	0.835821	0.761905	0.797153
1	Random Forest (n=100)	0.999520	0.941667	0.768707	0.846442

Comparison of Performance

The performance of both algorithms is shown by the column chart, this makes easy to understand the difference in accuracy in Random Forest Algorithm and Decision Tree Algorithm. So the performance of Random Forest Algorithm is more effective rather than Decision Tree.

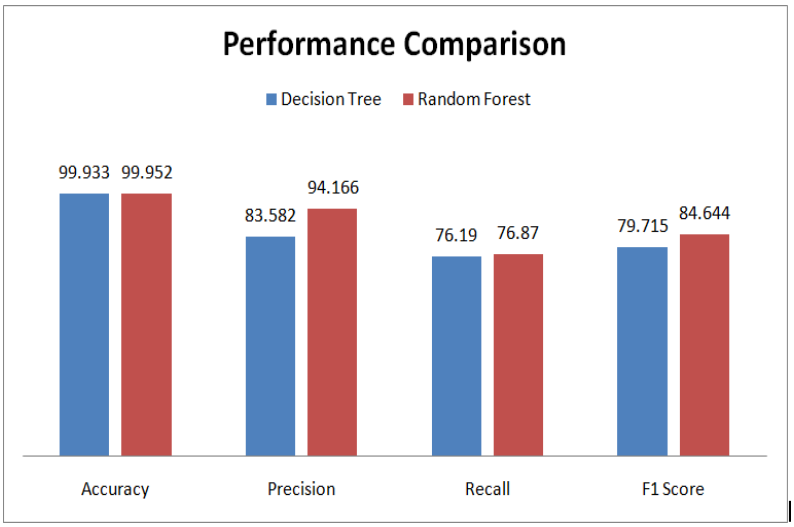

Figure 4.9 Performances of the Models

CONCLUSION

Cloud computing is a very flexible paradigm for delivering computational power. It means many things to many people. For some it means being able to set up a new start-up company knowing that initial resources will be inexpensive but a sudden increase in demand from users won't make the company a victim of its own success, as has happened in some cases in the past where servers have been unable to cope with demand, and the company loses clients as they become unhappy with poor response times.

In this, we can take a virtual machine on cloud preferred over others because cloud service is provided by third party providers (Google cloud platform (GCP)) so for security reason private cloud give better security than others because the connection between user and virtual machine is secured by ssh. And on private cloud we can easily scale the storage and processing power at any time whenever application required. The credit card usage become more famous among every one's life for daily needs, with that frauds related to plastic card also widespread. The credit card imbalanced dataset is classified in imbalance form for the Random forest and Decision tree model, and we analyze the performance of each classification with models respectively. By this analysis we clearly says that Random Forest is better than Decision tree performance.

REFERENCES

[01] Samidha Khatri ; Aishwarya Arora ; Arun Prakash Agrawal ; “Supervised Machine Learning Algorithms for Credit Card Fraud Detection” in IEEE 2020.

[02] Ambika Gupta; Pragati Goswami; Nishi Chaudhary; Rashi Bansal "Deploying an Application using Google Cloud Platform" in 2020, IEEE

[03] S. Mittal and S. Tyagi, "Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection", 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2019.

[04] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi and G. Bontempi,” Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy,” in IEEE Transactions on Neural Networks and Learning Systems,vol. 29, Aug. 2018.

[05] John O. Awoyemi, Adebayo O. Adetunmbi, Samuel A. Oluwadare, " Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis" in 2017 IEEE.

[06] Bolton, R. J., and Hand, D. J. (2002). Statistical fraud detection: a review. Statistical Science, 17(3), 235-249

[07] P Chan, W Fan, A Prodromidis& S Stolfo. 1999. Distributed data mining in credit card fraud detection, IEEE Intelligent Systems, 14(6): 67–74.

[08] Manuel Parra-Royon, Jose M. Benitez" Delivering Data Mining Services in Cloud Computing " in IEEE 2019

[09] Srivastava, A., Kundu, A., Sural, S., and Majumdar, A. (2008). Credit card fraud detection using hidden markov model. IEEE Transactions on Dependable and Secure Computing, 5(1), 37-48.

[10] Ghosh, S. and Reilly D. L. 1994. Credit card fraud detection with a neural network. In Proceedings of the 27th Hawaii International Conference on system Science.

[11] Ju, W. H., and Vardi, Y. (2001). A hybrid high-order markov chain model for computer intrusion detection. Journal of Computational and Graphical Statistics, 10(2), 277-295.

[12] Chen, R.-C., Luo, S.-T., Liang, X. and Lee, V. C. S. Personalized approach based on SVM and ANN for detecting credit card fraud. In Proceedings of the IEEE 2005.

[13] V.Dheepa and Dr. R.Dhanapal, et al. “Analysis of Credit Card Fraud Detection Methods”, “IJRTE”, Vol 2, No. 3, November 2009

[14] S. Dhankhad, E. Mohammed and B. Far,” Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study,” 2018 IEEE International Conference on Information Reuse and Integration

(IRI), Salt Lake City, UT, 2018.

[15] K. Chaudhary, J. Yadav, and B. Mallick, "A review of Fraud Detection Techniques: Credit Card," Int. J. Comput. Appl., vol. 45, no. 1, pp. 975-8887, 2012.

[16] J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, ''Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, 2017

[17] Arun Kumar Rai; Rajendra Kumar Dwivedi; "Fraud Detection in Credit Card Data Using Unsupervised Machine Learning Based Scheme", IEEE 2020.

[18] O. S. Yee, S. Sagadevan, N. Hashimah, and A. Hassain, "Credit Card Fraud Detection Using Machine Learning As Data Mining Technique," vol.10, no. 1,

[19] N.Sivakumar, Dr.R.Balasubramanian "Credit Card Fraud Detection: Incidents, Challenges And Solutions" in IJARCSA.

[20] Joe Fiala; "A Survey of Machine Learning Applications to Cloud Computing" http://www.cse.wustl.edu/~jain/cse570-15.

[21] Yasir Shoaib; Olivia Das; "Cloud VM provisioning using analytical performance models".IEEE 2019.

[22] Anish Halimaa A; Dr. K.Sundarakantham; "Machine Learning Based Intrusion Detection System". ICOEI 2019.