***WIRE FAME DESIGN***

***CREDIT CARD FRAUD Detection***

ABSTRACT:

***credit cards are quite useful for day to day life. the main aim of this project is to detect fraud accurately. with the increase in fraud rates, researchers have started using different machine learning methods to detect and analyze frauds in online transactions. 'fraud' in credit card transactions is unauthorized and unwanted usage of an account by someone other than the owner of that account. fraud detection involves monitoring the activities of users in whole in order to estimate, perceive or avoid objectionable behavior, which consist of fraud, intrusion, and defaulting. the problem itself is more challenging with respect to data science since the number of valid transactions far outnumber fraudulent ones. also, the transaction patterns often change their statistical properties over the course of time. however, the massive stream of payment requests is quickly scanned by automatic tools that determine which transactions to authorize. also, messages are generated to confirm with the owner about the transactions. machine learning algorithms are employed to analyze all the authorized transactions and report the suspicious ones. these reports are investigated by professionals who contact the cardholders to confirm if the transaction was genuine or fraudulent. this project also designs and develops a novel fraud detection method for streaming transaction data, with an objective, to analyze the past transaction details of the customers and extract the behavioral patterns. to name a few techniques which we are going to implement are isolation forest algorithm, random forest algorithm, logistic regression, confusion matrix and sliding-window method. keywords- credit card fraud, isolation forest, local outlier factor, random forest, confusion matrix***

1. ***introduction***

***credit card generally refers to a card that is assigned to the customer (cardholder), usually allowing them to purchase goods and services within the credit limit or withdraw cash in advance. fraud in credit card transactions is the unbidden use of someone’s account without the owner being aware of it. credit card frauds are easy targets. fraudsters always try to make every fraudulent transaction legitimate, which makes fraud detection a very challenging and difficult task to detect.***

***there are many machine learning techniques to overcome this problem. our project aims to prevent such fraudulent practices by analyzing and studying these fraud transactions to avoid similar situations in the upcoming transactions.***

1. ***literature survey:***

***the 1st we referred to is credit card fraud detection using machine learning algorithms1 . by vaishnavi nath dornadulaa, geetha s, published by ijcsmc in the year 2019. we observed that the matthews correlation coefficient was the better parameter to deal with the imbalance dataset. by applying the smote, we tried balancing the dataset, where we found that the classifiers were performing better than before. the other way of handling an imbalanced dataset is to use one-class classifiers like one-class svm. we finally observed that logistic regression, decision tree and random forest are the algorithms that gave better results***

***the 2nd paper referred to is machine learning for credit card fraud detection system2 by lakshmi s, selvani deepthi kavila, published by research india publications in the year 2018, sensitivity, specificity, accuracy and error rate are used to evaluate the performance for the proposed system. the accuracy for logistic regression, decision tree and random forest classifier are 90.0, 94.3, and 95.5 respectively. by comparing all the methods, we found that the random forest classifier is better than the logistic regression and decision tree.***

***the 3rd paper referred is credit card fraud detection using machine learning and data science3 by s p maniraj, aditya saini, swarna deep sarkar shadab ahmed,published by ijert, in 2019.in this paper, the algorithm does reach over 99.6% accuracy, its precision remains only at 28% when a tenth of the data set is taken into consideration. however, when the entire dataset is fed into the algorithm, the precision rises to 33%. this project allows for multiple algorithms to be integrated together as modules and their results can be combined to increase the accuracy of the final result. this model can further be improved with the addition of more algorithms into it.***

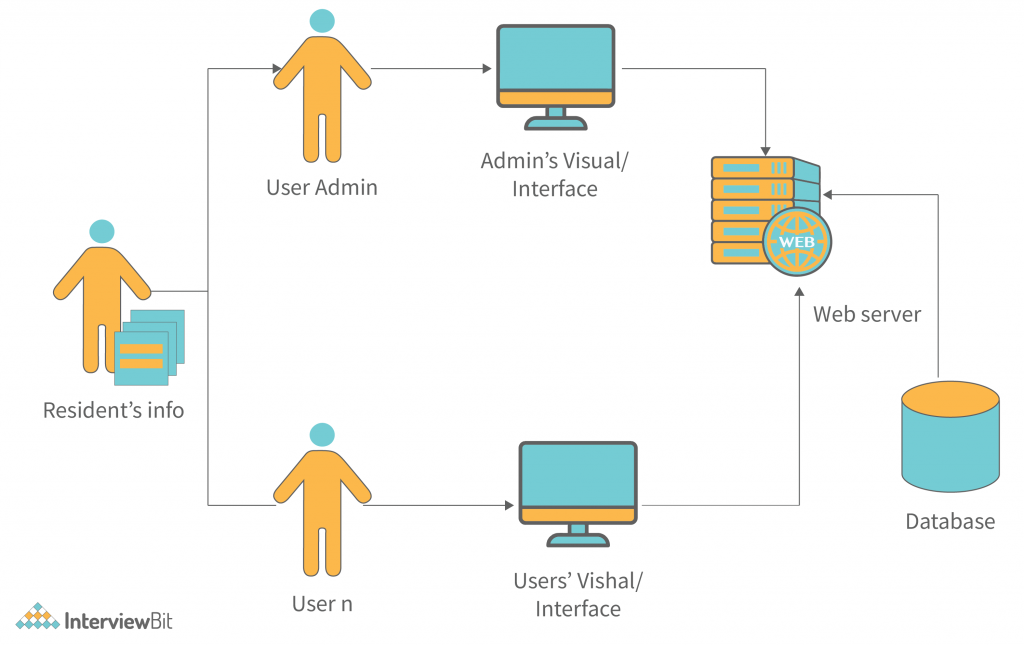
***the 4th paper is credit card fraud detection using machine learning: a survey4 by dr. yvan lucas, dr. johannes jurgovsky, arxiv, 2020.different approaches for tackling each of these challenges are highlighted in this survey and for each of these approaches at minimum one research work is described in full detail. the goal is to provide the reader with useful information on all the different research subjects introduced here.***

1. ***methodology:***

***we have gathered data from the kaggle website. the dataset is trained and tested using the following techniques: logistic regression, random forests with decision trees, xgboost, isolation forest and confusion matrix . if our algorithm is applied into bank credit card fraud detection systems, the probability of fraud transactions can be predicted soon after credit card transaction occurs. thereafter a series of anti-fraud strategies can be adopted to prevent banks from great losses and reduce risks.***

* 1. ***system architecture:***

***this is a block diagram for our system. it is the actual representation of the algorithm which we implemented. the first step is to read the data set and then it is sent for sampling. training and testing of the data set is done.***

******

***after the feature selection, the data will be sent to the algorithm which is the combination of the logistic regression, decision tree, random forest & mcc. the resultant data is stored in test sample data. the prediction of outcome is done based on test sample data & the result of the combined algorithm. later the performance & accuracy results are plotted. it has a feature which validates the results if the transaction is legitimate then the transaction is said to be true or else it is false. in case of false transactions the bank is made aware of it. 3.2.***

***requirements:***

***the implementation requirements are given in this section.***

***3.2.1 software***

***1. operating system : windows 8/10.***

***2. ide tool : visual studios and jupyter notebook***

***3. coding language : python 3.6 & up***

***4. apis : numpy, pandas, matplotlib,imblearn***

***5. framework : streamlit***

***3.2.2 hardware***

***1. processor : pentium i3 or higher.***

***2. ram : 4 gb or higher.***

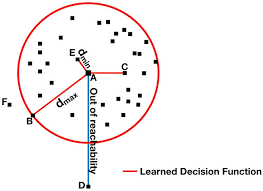
***3. hard disk drive : 20 gb (free).***

***3.3 algorithms:***

***a] logisticregression logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable.***

***accuracy on training data : 0.9440914866581956 accuracy score on test data : 0.9543147208121827***

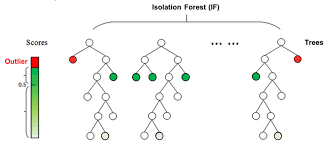
***b] local outlier factor in anomaly detection, the local outlier factor is an algorithm proposed by markus m. breunig, hans-peter kriegel, raymond t. ng and jörg sander in 2000 for finding anomalous data points by measuring the local deviation of a given data point with respect to its neighbors. following is the result of applying local outlier factor on our dataset c]isolation forest algorithm isolation forest is an anomaly detection algorithm. it detects anomalies using isolation, rather than modeling the normal points. following is the result after implementing the isolation forest algorithm on our dataset.***

******

***c]isolation forest algorithm***

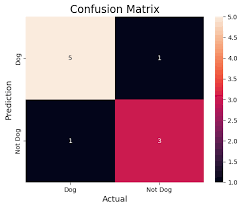
***isolation forest is an anomaly detection algorithm. it detects anomalies using isolation, rather than modeling the normal points.***

***following is the result after implementing the isolation forest algorithm on our dataset***

******

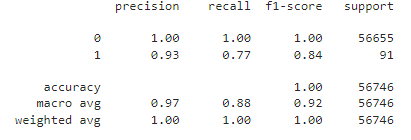
***d]confusion matrix***

***in the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one following is our confusion matrix***

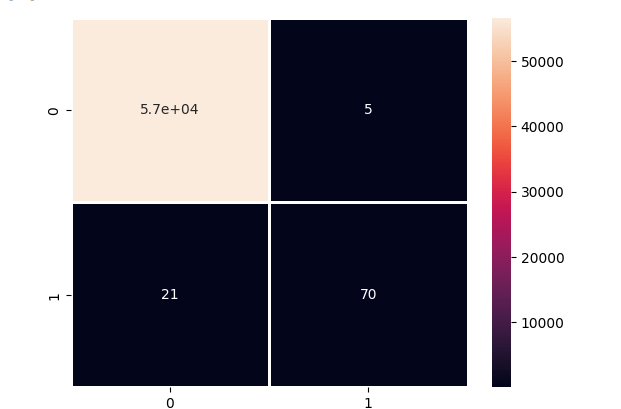
******

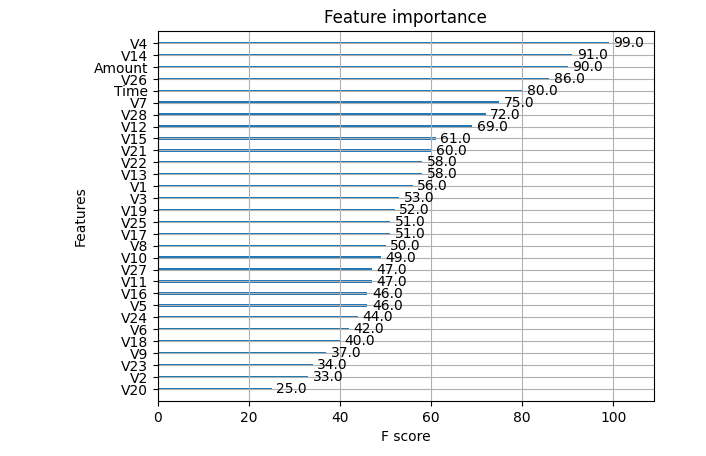
***e]xgboost:***

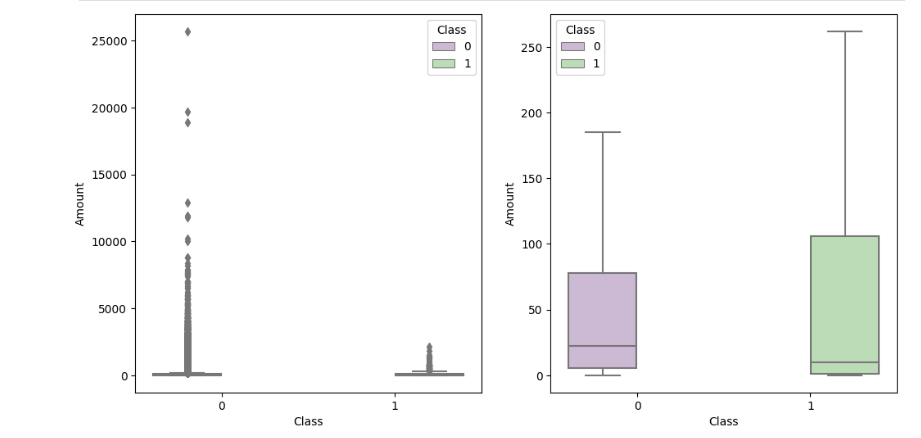
***xgboost is a decision-tree-based ensemble machine learning algorithm that uses a gradient boosting framework.***

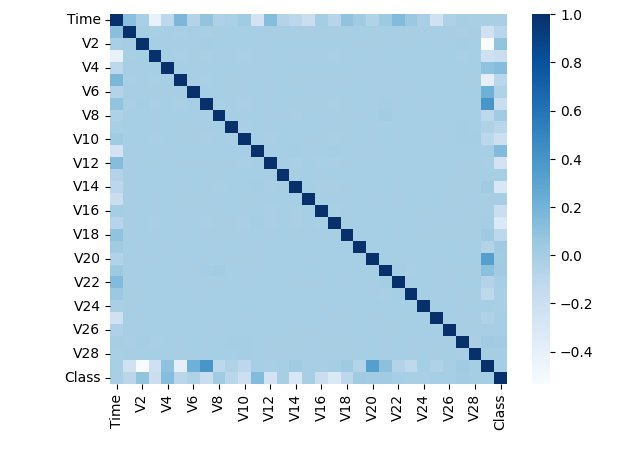
******

***4. Results***

******

******

******

******

***conclusion:***

***credit card fraud is an act of criminal dishonesty. this paper. this paper has various machine learning algorithms. all these techniques are tested based on accuracy and precision. we have selected supervised learning technique random forest to classify the alert as fraudulent or authorized. this classifier will be trained using feedback and delayed supervised samples. next it will aggregate each probability to detect alerts. further we proposed a learning to rank approach where alerts will be ranked based on priority. the suggested method will be able to solve the class imbalance and concept drift problem. future work will include applying semi-supervised learning methods for classification of alerts in fraud detection systems.***