Credit Card Fraud Detection using Combination of Machine Learning and Deep learning

Monika Kisan Gadage
Graduate in Computer Science
California State University, Chico
Chico, USA
mgadage@csuchico.edu

***Abstract*—Financial fraud occurs when someone deprives a person of their money, capital or harm financial health. A credit card fraud is said to be a fraud when some other person uses your credit card instead of you without your authorization. Due to rapid spread of fraud risks in digital economy, fraud detecting in credit card stands as a prime issue of modern technology. Analysis of fraud cases is difficult because fraud cases contribute to less than 0.3% of overall transactions. In order to find the best classification technique for fraud detection this paper has conducted a thorough experimentation using various Machine learning(ML) and Deep Learning(DL) techniques. Five ML techniques i.e. Gaussian Naïve Bias(GNB), Logistic Regression(LR), Decision Tree(DT), Random Forest (RF), Linear Regression(LIR) along with DL technique i.e. Neural Network(NN) have been implemented. Two different data sets are used along with different sampling methods to analysis the results. The best combination of these classification approaches is selected based on different performance evaluation criteria. After evaluation of classifier it has been shown that RF classifier with over-sampling technique among ML techniques is the best approach and gives 99.7% f1 score, AUC which is very close to NN with over-sampling technique results in 99.5% AUC and 99% f1 score. This analysis shows that over-sampling works best on any Credit Card dataset and any of the two approaches can be used to get accurate results.**

***Index Terms*—Deep Learning, Machine Learning, Samplimg methods**

I. Introduction

Fraud is defined as criminal deception intended to result in financial or personal gain of an individual performing the fraud. It can take boundless variety of diverse forms. With modern technology it becomes more easier for criminals to commit frauds, while companies are moving rapidly towards internet and money is transacted electronically in an ever growing cashless banking environment. Accurate fraud detection technique remains a key concern of modern banking. Credit Card fraud is the most common type of banking fraud which generally falls into two categories, card present fraud and card not present fraud. As the future is heading towards cashless future it becomes a need of the hour to accurately determine fraud transactions and give best customer support. Customers will no longer visit business with cash in their pockets. As a result, companies will need to update their environment to ensure that they can take all type of payments. In 2018, unauthorised financial fraud losses across payment cards and remote banking totalled £844.8 million in the United Kingdom. According to a 2021 annual report, about 50 of all Americans have experienced a fraudulent charge on their credit or debit cards, and more than one in three credit or debit card holders have experienced fraud multiple times. This amounts to 127 million people in the US that have been victims of credit card theft at least once. Looking at all the statistics it is very important to prioritize for upgrade with an automated fraud detection system. The goal of Credit Card fraud detection (CCF) system is to find a ML and DL algorithm based on existing transactions credit card payment details. The model should accurately distinguish between fraud and non-fraud transactions, and use the information to find future incoming fraud transaction.

II. Methodology

In this paper ML classifiers namely: LIR, LR, DT, GNB, RF are implemented along with DL NN algorithm in python. The various stages of generating and handling the classifiers consist of; assembly of data, pre-processing of data, training the classifiers, testing the processes and investigating the classifiers. Different sampling techniques are used in pre-processing stage to generate balanced data. Three sampling techniques are used Under Sample (US) , Over Sample (OS), SMOTE. Primary the models were implemented on imbalanced datasets. Then consecutively all classifier along with NN were implemented on sampled dataset for comparison. In total 4 dataset*2 (different type of dataset) = 8 dataset are modelled through * 6 classier and algorithms. 48 models were performed to effectively classify fraudulent transactions of credit card.

A. Dataset

This paper provides study on two datasets. The first, is the Credit Card Fraud Detection dataset downloaded from Kaggle. The dataset contains transactions made in September 2013 by European cardholders through their credit cards. The transactions took place during two days, where we have 492 frauds out of 284,807 transactions. To maintain the privacy of customers numerical input variables are provide which are the result of Principle component analysis (PCA) transformation. The features which are not transformed with PCA are Time , Class and Amount. Time column gives the value of time

passed (in seconds) since the transaction data have been collected. Amount column gives the amount that is being transacted, which is used for cost sensitive learning. Class column is the result column. If class is '1' it means that the transaction is a fraud transaction and if it is '0', then the particular transaction is a valid transaction. The data is highly imbalance as the fraud transactions account for 0.17 of the total transactions. The second, is a synthetic dataset generated by a stimulator which consists pre-defined list of customers, business and transactions categories. Python library 'faker' is used which takes in account the number of business and customers one wishes to work with. Depending on the age group and gender provide the dataset is generated. This paper uses dataset containing 1000 customers of all age groups and gender doing transactions with 800 business. The dataset has 1296675 records with 23 features. Some of the important columns in the dataset are, dob(date of birth), category(type of business), amount, gender and isfraud which is the result column. If isfraud is '1' it means the transactions is fraud and if it is '0', then the transaction is valid transaction. The percentage of fraud transactions in dataset is 0.58 which is impressive compared to the first dataset.

B. Classifier

Linear Regression(LIR), The Liner Regression model finds the linear relationship between the independent and dependent variable. It predicts a dependent value based on the independent variable. It fits the best line to predict the dependent value i.e. fraud values based on non-fraud values such that the error between the values is minimum.

Logistic Regression(LR), It is consider as a discriminative model also known as logit model. logistic regression is used to estimate the relationship between a dependent variable and one or more independent variables, it is used to make a prediction about a categorical variable which can be true or false, yes or no, 1 or 0. Certain behaviours or characteristics may have a higher association with fraudulent activities which is taken into consideration.

Decision Tree(DT), A Decision tree is a flowchart-like tree structure, the leaf node represents the class label, the internal nodes represents the test on attribute and is the outcome of the test. A tree is learned by splitting the set into subsets in a attribute value test. This process is done recursively and is completed when subset at the node have same value of the target this is called as recursive partitioning. They give good accuracy. They classify the data by sorting down the tree from root node to leaf node. Breath first or depth first approach is used to classify each node. A best formed tree is the one in which the subgroups do not intersect with one another and are notably distinct.

Gaussian Naïve Bias(GNB), A Naive Bayes classifier is a model that's used for classification task it is based on the Bayes theorem. Bayes theorem states that we can find probability of an event happening given that the a event has occurred beforehand. The event which occurred first is the evidence and the event happening later is hypothesis, provided the features are independent. The difference in events is called naïve. It is mainly used for large dataset and easy to build. In gaussian naïve bias we assume that the predictors take up a continuous value and are not discrete, it is assumed that the values are sampled from a gaussian distribution.

Random Forest(RF), It is one of the best classifier where individual decision trees are combined to form a random forest. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes model's prediction. The reason why RF works so well is because a large number of relatively uncorrelated trees operate as a committee and hence is much better than individual trees. Each tree protect each other from individual errors as some of the trees can be wrong many will predict the right results.

C. Deep Learning Model

Neural Network, It allows computer programs to recognize patterns and get results by reflecting the behaviour like a human brain. It consists of different node layers, containing an input layer, one or more hidden layers and an output layer. Each node connects to another and has an weight and threshold associated with it. If output of any individual node is above the specified threshold value, the node is activated and sends data to next layer of network. Otherwise no data is passed along to the next layer of the network. NN rely on training data to learn and improve its accuracy with time. Once this learning algorithms are fine tunes they are very powerful in classifying and clustering data at high velocity.

D. Data Pre Processing

As mentioned before we will be using two different type of dataset European cardholders dataset and Synthetic dataset. The goal is to check how different models will perform on new data after training the models. New data is not always available while building the model hence python train test split library is used which divides the data in two categories Testing data(25%) and Training data(70%). Testing data is selected randomly from the dataset and saved to check for the accuracy of the prediction. Training data is used to build all the models. European dataset is highly imbalanced with only 0.17 of transactions being fraud. Models do not work best with imbalance data, where the classes are not characterized equally. Two commonly approached to make balanced dataset from imbalance dataset are over-sampling and under-sampling.

Over-Sampling, the dataset is balanced by increasing the size of minority samples. New minority samples are generated by using repetition, bootstrapping or SMOTE. This can result in overfitting for some cases due to over replication of elements from the minority class.

SMOTE, it is an improvement of over-sampling dataset by synthesizing new examples from minority class. It selects samples that are close in feature space, drawing lines between samples in the feature space and drawing new sample at point along that line.

Under-Sampling, It removes samples from the training dataset that belong to the majority class in order to better

balance the class distribution such as reducing skew from a 1:100 to a 1:10. This type of sampling works best with binary classification problems like the one we are dealing with.

III. Results and analysis

The paper starts with training all the model with real life imbalance dataset and studying the results, the accuracy of models over both the dataset is compared. 'Fig. 1" shows how different models showed results on imbalance dataset. The models give different accuracy with European and sythetic dataset, it proves that models can be unpredictive LIR, LR , DT and RF showed almost similar accuracy on both dataset but with GNB it is very off. Mostly the accuracy is high because the data is highly imbalance and most of the samples are non fraud and will be sampled correctly.

In undersampling, 'Fig. 2" the models showed different results apart from for RF and on test data the accuracy is same for LR, DT, RF on both dataset.

In over-sampling 'Fig. 3" we see very clean line and the accuracy is the same for both the dataset over training and test data it shows how oversampling is the best sampling used for fraud detection. The accuracy for RF is almost 0.99 which is the highest amongst all the models.

SMOTE 'Fig. 4" shows very similar results to undersampling, however the accuracy is increased which does takes in account the fact that SMOTE takes considers different factors while sampling. After using different sampling the accuracy of models is decreased because sampling diffided the classed into equal parts. NN 'Fig. 5" shows 99% percentage accuracy with over-sampling which further proves that over-sampling is the best sampling menthod for fraud detection.

As RF and NN showed top results by means of different metrics, accuracy, recall, precision and close results on both dataset. It was precisely classified that RF along with over-sampling shows the best results similar to NN with over-sampling these two models can be combined to give the most accurate results while detection fraud detection.

IV. Conclusion and Future Work

This paper compares the performance of five machine learning classifiers along with Deep learning algorithm namely, LIR, LR, DT, GNB, RF and NN on two different type of dataset to predict credit card fraud transactions. Among 48 different approaches it has been found that over-sampled data on RF classifier shows 99.7% accuracy. LR performs better than DT in Oversampling and DT performs better in Undersampling. GNB does not show promising results over different sampling. RF is very powerful classifier as internally it works while implementing many DT along the way.

NN shows same accuracy as RF on over sampled data. It is worth noting that all this detect fraud when the fraud has already happened because usually a very small percentage of transaction are fraudulent. The solution to this is developing a technology when the fraud is taking place. The ability of NN to learn from the past and achieve a high performance rate but the limitations of high processing time can be reduced with a hybrid approach of using RF algorithm.

References

[1] Maes, S., Tuyls, K., Vanschoenwinkel, B. and Manderick, B., (2002). Credit card fraud detection using Bayesian and neural networks. Proceeding International NAISO Congress on Neuro Fuzzy Technologies.

[2] B. Pushpalatha and C.W. Joseph, "Credit Card Treachery Detection Based on the Transaction by Using Data mining Techniques," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, No. 2, pp. 1785-1793, 2017.

[3] O.J.Awoyemi, O.A. Adetunmbi, andS.A.Oluwadare, Machinelearning techniquesfordetectingfraud:AcomparativeReview,"Int.2017.Conf. Conf. App. Netw. Netw. Computer Science,pp. 1-9, 2017.

[4] "Credit Card Fraud Detection: Top ML Solutions in 2021," Technology partner for innovative companies, 20-May-2021. [Online]. Available: https://spd.group/machine-learning/credit-card- fraud-detection/. [Accessed: 2-Jun-2021].

[5] S. K. Shirgave, C. J. Awati, R. More, and S. S. Patil, "A Review On Credit Card Fraud Detection Using Machine Learning," Int. J. Sci. Technol. Res., vol. 8, no. 10, 2019, Accessed: Nov. 16, 2021. [Online]. Available: www.ijstr.org.

[6] A. Walke, "Comparison of Supervised and Unsupervised Fraud Detection," Commun. Comput. Inf. Sci., vol. 1097 CCIS, pp. 8–14, 2019, doi: 10.1007/978-3-030-36365-92.

[7] A. Gepp, K. Kumar and S. Bhattacharya, "Improving models that detect financial statement fraud: A new framework to guide variable selection," Abstract from Accounting and Finance Association of Australia and New Zealand (AFAANZ) 2016, Gold Coast, Australia , 2016, doi: 10.2/JQUERY.MIN.JS.

[8] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," Decision Support Systems, vol. 50, no. 3, pp. 602–613, Feb. 2011, doi: 10.1016/j.dss.2010.08.008.

[9] M.Zareapoor,s.k.seeja.K.R,andM.AfsharAlam"CreditCardFraud Identification Analysis Techniques: Focused on some design requirements," int.j.comput. Vol 52,no.3,pp.35-42, 2012.

[10] Kavya Gupta,Kirtivardhan Singh,and Gaurav Vikram Singh "Machine Learning based Credit Card Fraud Detection- A Review" int.j.comput. Vol 52,no.3,pp.35-42, 2012.

[11] FAWAZ KHALED ALARFAJ 1, IQRA MALIK2, HIKMAT ULLAH KHAN 3, NAIF ALMUSALLAM1,MUHAMMAD RAMZAN 2, AND MUZAMIL AHMED"Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms," int.j.comput. 10.1109/ACCESS.2022.3166891.

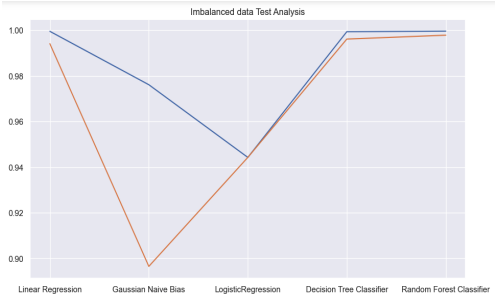
[12] Faroque Ahmed, Rittika Shamsuddin"A Comparative Study of Credit Card Fraud Detection Using the Combination of Machine Learning Techniques with Data Imbalance Solution," 2021 2nd International Conference on Computing and Data Science (CDS).

[13] Pranali Shenvi, Neel Samant, Shubham Kumar and Dr. Vaishali Kulkarni" Credit Card Fraud Detection using Deep Learning," int.j.comput. Vol 019 5th International Conference for Convergence in Technology (I2CT).

[14] Janet B,Joshua Arul Kumar R,and Didugu Phani Sai Ganesh "Credit Card Fraud Detection with Unbalanced Real and Synthetic dataset using Machine Learning models," 2022 IEEE

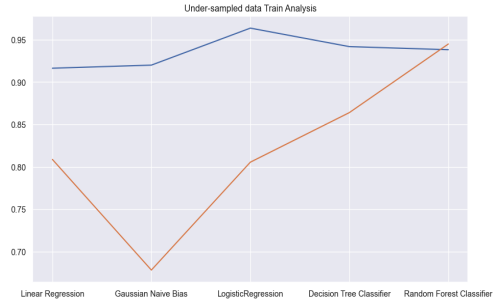
[15] Anu M aria Babu,Dr. Anju Pratap"Credit Card Fraud Detection Using Deep Learning," 2020 IEEE.

(a) Imbalance Training



(b) Imbalance Test

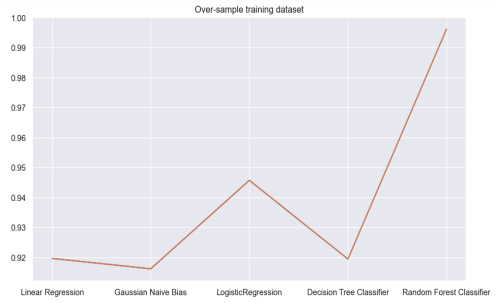
Fig. 1. Use Of imbalance dataset over models



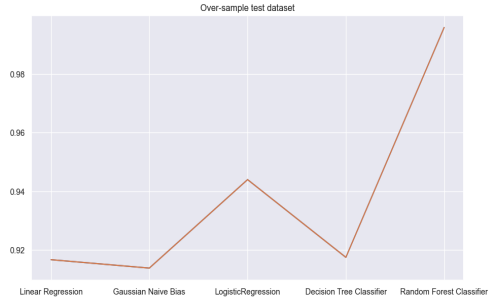
(a) Under Sample Train

(b) Under Sample Test

Fig. 2. Under Sample the data

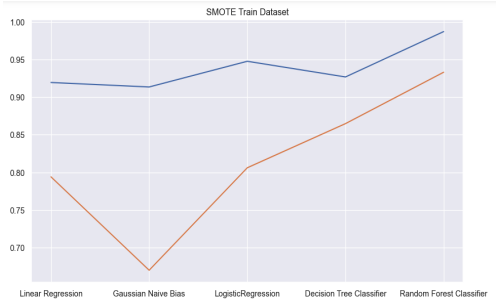


(a) OverSample training



(b) Over-Sample testing

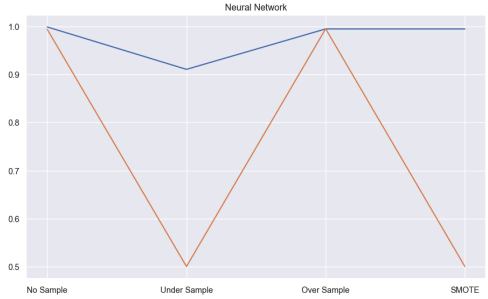
Fig. 3. Over-sample the data



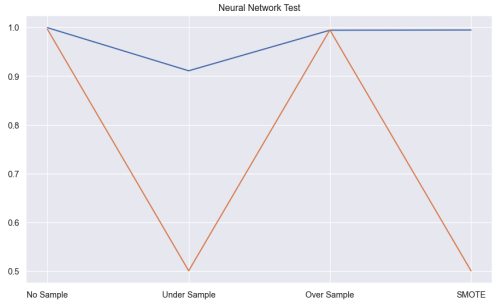
(a) SMOTE train

(b) SMOTE test

Fig. 4. SMOTE sampling on data



(a) Neural Network on training



(b) Neural Network on testing

Fig. 5. Use of Neural Network over different sampling