Credit Card Fraud Detection using Deep Learning and Machine Learning Algorithms

Saba Zahid[1], Hafiz Muhammad Usman Hafeez[2], Muhammad Javaid Iqbal*[3], Ali Asif[4], Sobia Yaqoob[5], Fozia Mehboob[6]

Received: 17 Sep,2023; Accepted: 29 Fab, 2024; Published: 30 March, 2024

Abstract: Use of credit card is very common these days. And the number of frauds related to credit cards are also increasing. With the increase in the usage of internet, many organizations have shifted their work from offline to online. Same is the case with financial department. On one side, this thing has increased the ease of people but on the other hand, number of frauds have been tremendously increased. On one side, people are doing shopping without cash, paying bills without standing in long queues, doing booking online and on the other side fake accounts, scamming, credit card frauds have been increased resulting in huge amount of loss to financial system every year. Fraud is a criminal activity done by un authorized person. Credit card frauds are very common these days. There are many types of credit card frauds. Sometime they do fake calls or messages and sometimes they steal customer's online information. Many techniques using machine learning models have been implemented in order to stop these types of frauds. But fraudsters are sometimes by pass theses traditional protective systems and make successful transaction. Traditional machine learning models are not capable enough to detect frauds using sequence of data. For this purpose, neural networks are recently used. In this paper, six machine learning algorithms are applied. Among them Random Forest and Extra trees classifier are best. And in case of neural networks, long short-term memory LSTM is best. Obtained results outperform the existed work that have been previously done in this field.

Keywords: Machine learning, Credit card, Fraud detection

1- Introduction

As the world is progressing, people are enjoying the privileges of new technologies. They try to prefer use of credit cards as compared to cash. Instead of taking cash everywhere they are fond of carrying their credit cards with them. Moreover, as now a day, everywhere there is facility that either you can pay with cash or with your card. And they like to avail the second opportunity. Four

[1]*Department of Computer Science, Lahore University, Lahore Pakistan*
[2]*University of Engineering & Technology Lahore*
[3] *Faculty of Computer Science and Information Technology, LEADS University, Lahore 54000, Pakistan*
[4]*Department of Computer Science, COMSATS University Islamababd, Sahival Pakistan,*
[5]*Department of Computer Science, University of Okara, Pakistan,* sobia.yaqoob@uo.edu.pk
[6]*Department of Computer Science & Applications, Mälardalen University, Sweden*
fozia.mehboob@mdu.se
[*]*Corresponding Author javaid.always@gmai.com*

to five years back in 2019, Covid-19 pandemic hit the world. There were lockdowns everywhere. So, people preferred online shopping methods at that time using their cards. In the era of digitized world, at on side where credit cards have made our life easier than on the other side, shadows of threats and frauds are always around us. Financial departments are facing a lot of challenges regarding these frauds and threats. Detecting fraudulent transaction accurately is one of them. Second one is to ensure that when authorized owner of card makes any transaction then system should not mark it as fraud transaction as it will lead to damaging the bank repute Infront of customer by making false alarm.

Any transaction done by unauthorized person comes under credit card fraud. In the era of modern technologies, credit card frauds are very common. Owing to rapid growing online banking system, it was estimated that 44% people from United nation do online transactions. Now as the use of credit card is increasing, the number of frauds related to the credit card are also increasing. Online and offline both types of frauds are linked with credit card. Criminals are using many techniques and websites to steal your private credit card information. Sometimes they steal your card information while you are doing online transaction and sometimes card is lost. Similarly, there are many other types of frauds. Skimming and phishing are among of them. Some fraudsters steal information by sending fake emails and Messages and by doing fake calls. Some steal credit card itself and some steal sensitive information of credit card while doing online transactions.

Huge amount of people prefers online banking system. Increased number of frauds related to credit cards are resulting in huge amount of financial loss to banks as well as customers. Many proposed models are used for this purpose and any mistake in the detecting models leads in the huge financial loss. And if we talk about fraudsters, they are also enjoying privileges of new technologies by changing their way of doing crime every day. They are now efficient enough that sometimes they bypass the conventional detecting methods and make successful fraudulent transactions. Many machine learning algorithms have been implemented in order to stop these types of frauds such as Logistic Regression, Random Forest, Decision Tree, Support Vector Machine etc. but still there is need to improve this system. Recently, neural networks are the focus of attention for many researchers.

Some important machine learning algorithms in this context are:

- Logistic Regression
- Random Forest
- Extra Trees Classifier
- XGB (Extreme Gradient Boosting)
- LGBM (Light Gradient Boosting Machine)
- CatBoost (Categorical Boosting)
- Support Vector Machine (SVM)
- Decision Tree etc.

And if we talk about deep learning then following neural networks are mostly used.

- Artificial Neural Network (ANN)
- Convolutional Neural Network (CNN)
- Recurrent Neural Network (RNN)

2- Literature Review

Tremendous amount of work has been done in order to stop or avoid these types of frauds. Some have used machine learning algorithms while others focus on sampling techniques to address issue of class imbalance. Tanouz et al. used open source kaggle dataset in this paper. With the help of undersampling, the size of data set is reduced considerably. Then the method of outlier detection and removal is used. This will result in the further reduction of data set.70% dataset is used for training and remaining 30% is used for testing. After analysis, it is determined that random forest is best among all of them on the basis of evaluating metrics. in this case accuracy of 96.774% and still there are false negatives (FN) and false positives (FP). In the future work, the authors aim to obtain much higher accuracy as they are unable to obtain 100% accuracy and reduce the value of false negatives[1]. Noor and Suliman used two methodologies in order to recognize the fraudulent transactions. In the first method, Kaggle data set is used without any resampling technique and nine different machine learning algorithms are applied. CatBoost, XGBoost, RF are the best three algorithms. In the second method, as the data set is highly imbalanced so 19 resampling techniques are used.11 undersampling techniques and 6 oversampling techniques along with 2 under and oversampling techniques at once are used. Now best three algorithms of first method are applied. And after performing analysis and with the help of evaluation metrics, AllKNN-CatBoost is determined as best one[2].

Alarfaj et al. not only apply top best machine learning algorithms but also apply deep learning algorithms. Data set from kaggle has been used in this paper. First of all, best six machine learning algorithms Decision Tree, k-nearest neighbours (KNN), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression, XG Boost are applied. Accuracy and f1-score of these conventional machine learning algorithms are good such as 99.95% and 85.71 as in KNN case but result in many false positives (FP) and false negative (FN). In order to improve these factors, two deep learning algorithms i.e CNN and baseline are also applied. CNN model is applied with different layer architectures 11-20 and different epoch sizes are applied on balanced as well as imbalanced data. The efficiency of this algorithm is compared with baseline 5-layer architecture. On comparison it is concluded that performance of CNN with multiple layers and baseline model are best with accuracy of 99.72% [3]. The research conducted by Joy Iong-Zong Chen and Kong-Long Lai proposed deep convolution neural network (DCNN) technique in this paper. A random data set containing five million transactions is used. Data normalization and data validation steps have been performed before data splitting.99% efficiency has been obtained and after comparison among optimizers, Adam optimizer comes as best one. Three encoding and decoding layers are used and 500-1000 iteration have been performed. On comparison with Support Vector Machine

(SVM), Logistic Regression (LR) and Random Forest (RF), it is concluded that DCNN is better in terms of execution, prediction time and training[4].

Asha and Suresh proposed artificial neural network technique. ANN used in this paper comprises of input and output nodes. There are fifteen hidden layers and RELU is used as activation function. Data set from kaggle has been used. SVM, KNN, ANN are implemented and their results are compared and discussed. In the end the authors concluded that artificial neural network (ANN) gives almost maximum accuracy when compared to k-nearest neighbours (KNN) and Support Vector Machine (SVM)[5]. In this paper, Zorion et al. discussed previously implemented techniques. First of all, convolution neural network (CNN) is discussed. Adam optimizer with binary cross entropy loss function is used. Batch size of 32-128 and epochs 10-50 are used. Then autoencoder is discussed. MSE loss function with Adam optimizer is used. This model is trained with 32 batch size and 100 epochs. With the help of thresholding, MSE scores are converted to binary predictions. Then Long Short-Term Memory (LSTM) is discussed. RELU activation function is used with 32 units in dense layer. Kaggle Data set consists of 100,000 transactions, seven features and one target column. After detailed analysis, authors stated that LSTM has highest accuracy of about 99.92% and autoencoder has the lowest accuracy of about 32.15%[6].

Table 1: State-of-the-art survey

Paper	Year	Model	Evaluation Measure
[1]	2021	Random Forest	Accuracy
[1]	2021	Logistic Regression	Accuracy
[2]	2022	AllKNN-CatBoost	Accuracy, Recall, F1-Score
[3]	2022	Decision Tree	Accuracy
[3]	2022	KNN	Accuracy
[3]	2022	Logistic Regression	Accuracy
[3]	2022	SVM	Accuracy
[3]	2022	Random Forest	Accuracy
[3]	2022	XG Boost	Accuracy
[3]	2022	CNN	Accuracy
[4]	2021	DCNN	Accuracy
[5]	2021	SVM	Accuracy, Precision, Recall
[5]	2021	KNN	Accuracy, Precision, Recall
[5]	2021	ANN	Accuracy, Precision, Recall

In this paper, Ibtissam et al. proposed Long Short-Term Memory (LSTM) technique for fraud detection. Banksim is used as software tool and data set is downloaded from kaggle. Proposed model LSTM has 9 input neuron, one hidden layer comprising of 15 neuron and one output neuron. LSTM memory size of 15 with epoch number 100 is used. Cross entropy is used as a loss function and Adam optimizer is used as optimizer. SoftMax is used as an activation function. In the end, authors concluded that deviation from the genuine value will be smaller if the values of Root Mean

Square Error (RMSE), Mean Square Error (MSE) and Mean Absolute Error (MAE) are smaller[7]. Garg et al. used Auto ML model for comparison of different machine learning algorithms and extra trees classifier comes as best one among them with accuracy of 99.9%[8]. Bandar Alshawi used Generative Adversarial Networks (GANs) technique. Six machine learning algorithms are applied and their comparison is performed. XGBoost in on the top with accuracy of 98%[9]. Mangathayaru et al. used Support Vector Machine (SVM), Random Forest (RF), gradient boosting and convolutional neural network (CNN). They concluded that convolutional neural network (CNN) is effective and efficient one[10].

Ileberi et al. used synthetic minority over-sampling technique (SMOTE). Six machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), extra tree (ET), extreme gradient boosting (XGBoost), Logistic Regression (LR) and Decision Tree (DT) are. Then one additional technique, adaptive boosting (AdaBoost) is paired with all these algorithms to increase the efficiency and accuracy. In the end, ET-AdaBoost comes as best one with accuracy rate of 99.99%[11]. Yiying Wang has used SMOTE technique and its combination with different machine learning algorithms. He highlighted that by using features from 14-30, model performance has incredibly increased with accuracy rate of 98.57% and with AUC-ROC value of 0.9604[12]. Esenogho et al. proposed neural network LSTM method with Adaboost technique along with SMOTE-ENN sampling method[13]. Nguyen et al. proposed bi-directional LSTM approach. Which uses both previous and future knowledge for detection of fraud[14]. Carcillo et al. proposed combination of supervised and unsupervised techniques for efficient detection of fault. They found the cluster based approach, more promising among all[15]. Tzu-Hsuan Lin and Jehn-Ruey Jiang proposed auto encoder approach along with mathematical probabilistic Random Forest method. Together, they performed quite well[16]. Malik et al. proposed seven hybrid machine learning models in the paper and concluded that Light Gradient Boosting machine along with Adaboost is best among all[17].

Ileberi et al. used feature selection technique for their model using genetic algorithm. Then they applied machine learning models and observed significant impact of feature selection on results[18]. Lebichot et al. proposed fifteen transfer learning techniques in the paper. Ensemble of semi supervised and unsupervised algorithms comes as the best method[19]. Dastidar et al. proposed Neural Aggregate Generator (NAG) technique for the feature extraction and compared their results with LSTM and CNN[20]. Balawi and Aljohani discussed machine learning algorithms along with artificial neural networks. They compared results of ANN and CNN in detail[21]. Khalid et al. proposed ensemble technique which included Random Forest, Support Vector Machine (SVM), boosting, K-Nearest Neighbor(KNN) and bagging classifiers along with SMOTE technique[22]. Azim Mim et al. proposed an ensemble technique that is soft voting and applied this learning approach to imbalanced data sets. Results of this ensemble technique was better than individual classifiers[23]. Ayesha and Adil had discussed various machine learning algorithms using credit card fraud detection data set. They presented comparative analysis of multiple machine learning models in detail[24]. Bandari and Nayudu used multiple algorithms like

CatBoost, XGBoost, Logistic Regression and Light Gradient Boosting Machine. They used k-folds method as well as maximum voting ensemble technique to evaluate their performances[25].

3- Methodology

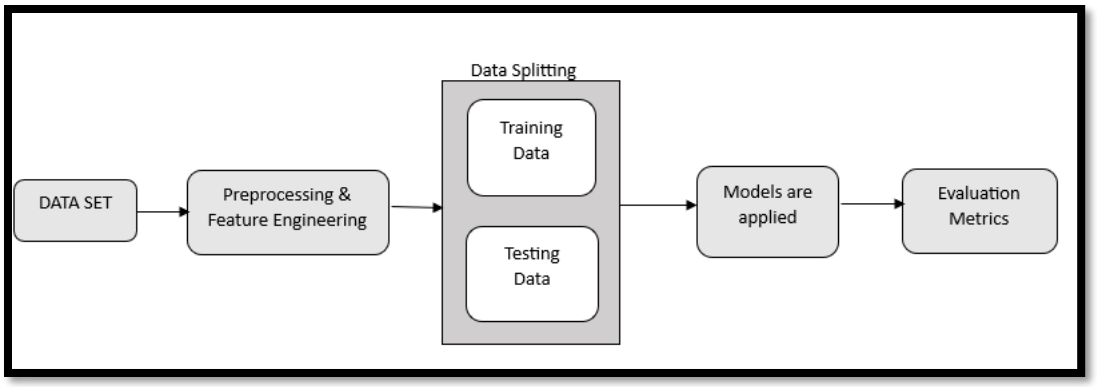

Fig. 1. Shows flowchart of methodology

3.1 Data Acquisition

Data set is downloaded from kaggle website[26]. I have used new data set which have been uploaded in 2023 instead of 2013, which is highly imbalanced. This data is highly balanced as the number of instances of fraud cases are equal to number of non-fraud cases. So, no over and under sampling techniques are required. Data set consists of 31 columns and have five hundred thousand plus entries.

3.2 Software Tools

In order to perform this research work, google Collaboratory platform have been used. Programming language Python is used.

3.3 Pre-Processing & Feature Engineering

For pre-processing and feature engineering, various Python libraries have been used. Exploratory Data Analysis (EDA) is performed to visualize and understand the data. Data cleaning and feature scaling is performed in this stage.

3.4 Train-Test Split

Python library is used to split the data in the testing and training parts. 70% is reserved for training and 30 % for testing. K-folds technique is also used to enhance the model performance.

3.5 Machine Learning Model

Following machine learning and deep learning models are used to train and test the data set.

Logistic Regression. In machine learning there are three types of learning. Supervised learning, unsupervised learning and reinforcement. Logistic regression is a machine learning technique that comes under the supervised learning. It is classifier that is used for binary classification. It tends to predict output by analyzing the dataset and as well as the relationship between independent and dependent variables. Firstly, it predicts that what will be the probability that a particular instance is belonging to a particular class then after this step it will classify the instance on the basis of threshold which is normally 0.5 in case of binary classification. Sigmoid function is used in logistic regression which squashes the output between 0 and 1.

Random Forest. It is basically ensemble technique where it uses many decision trees to predict the outcome. Instead of one tree, it uses many decision trees to make prediction which will be more accurate than that one which is predicted using single decision tree. In case of classification problems, it predicts that particular class from which the current instance belongs. As it uses many decision trees, so each makes its own prediction about the current sample. Then on the basis of majority voting, correct class is predicted. For example, in our fraud detection classification problem, it will classify whether the current transaction is fraudulent or not based on mode of results of decision trees. While in case of regression problem, it takes the average of prediction made by decision trees.

Extra Trees Classifier. Extra trees classifier is also an ensemble learning technique. It is very much similar to random forest method as it also uses multiple decision trees. It is actually extension of random forest algorithm. The random selection of both the features and split point makes this classifier unique. Hence, it increases the diversity and tries to reduce overfitting. Firstly, it constructs multiple decision trees and then predicts the output on the basis of majority voting among predictions made by these trees. In case of regression, it takes the average of these predictions. Bootstrap Aggregating (Bagging) technique is used in this algorithm. In bootstrap Aggregating, bootstrap sample is used to train each tree. Bootstrap sampling is the sampling with replacement method.

Extreme Gradient Boosting. It can easily handle complex and diverse types of data sets. It is one of the efficient tools and is useful for both classification as well as regression tasks. It is basically ensemble technique which combines decision trees and apply the concept of gradient boosting.in other words, it adds the decision trees in sequence in such a way that each new one corrects the errors related to previous one. It is best in handling tabular or structured data and hence is popular in finance departments, industries as well as in academia. Its regularization methods reduce the chances of model overfitting. Computational speed and accuracy offered by it makes it one of the best and efficient machine learning models.

Light Gradient Boosting Machine. Light Gradient Boosting Machine also known as LightGBM is an open-source machine learning model. Basic technique that is used in LightGBM is that it uses histograms where bins of feature values are recorded. It is useful for both classification as well as regression problems. It is popular among other machine learning models

due to many reasons such as its training speed is fast. It is efficient one and can easily handle huge data sets. It can easily handle categorical features. It can easily interact with other programming languages like Python etc. Its efficiency, training speed, best usage of memory and scalability make it popular in both academia as well as industry sector.

Cat Boost. Categorical boosting also known as CatBoost, was first introduced in 2017 by an IT company. It is one of the efficient machine learning models. It basically follows gradient boosting technique and is very useful while dealing with categorical type of data. It also uses oblivious trees technique and also reduces chances of model overfitting. It is written in C++ but can easily interact with other programming languages. It is popular in industry and academia due to many reasons such as it addresses the challenges that we encounter while dealing with categorical features. It is efficient in handling categorical features hence require less pre-processing.

Long Short-Term Memory. Long Short-Term Memory (LSTM) is the form of RNN. Recurrent neural network (RNN) is a type of neural network that have feedback loops. As traditional machine learning model are not able to detect sequence or patterns in data sets. In case of credit card fraud detection, there is a sequence related to transactions. Usually, customers have any type of pattern related to their normal transactions. So, whenever any deviation from this pattern occurs, model should detect this abnormal behavior. Only RNN can do this. But RNN struggle a lot in case of long-term dependencies problems. LSTM performs very well and is one the efficient machine learning model in case of time sequence or sequential data.

3.6 Evaluation Metrics

Following evaluation metrics are used in order to evaluate model performance.

- Accuracy (includes both training & validation(testing)accuracies)
- Precision
- F1-Score
- Recall
- Confusion Matrix

4- Results

In this paper, various machine learning models are applied on credit card fraud detection data sets. In order to evaluate the performance of these algorithms, various evaluation metrics have been calculated. Results of these performance metrics are shown below.

Table 2. Shows results of Classification Report

Models	Accuracy	Precision	Recall	F1-Score
Random Forest	1.00	1.00	1.00	1.00
Extra Trees Classifier	1.00	1.00	1.00	1.00
XGBoost	1.00	1.00	1.00	1.00

CatBoost	1.00	1.00	1.00	1.00
LGBM	1.00	1.00	1.00	1.00
Logistic Regression	0.97	0.97	0.97	0.97
LSTM	0.9997	0.9998	0.9996	0.9997

4.1. Long Short-Term Memory (LSTM)

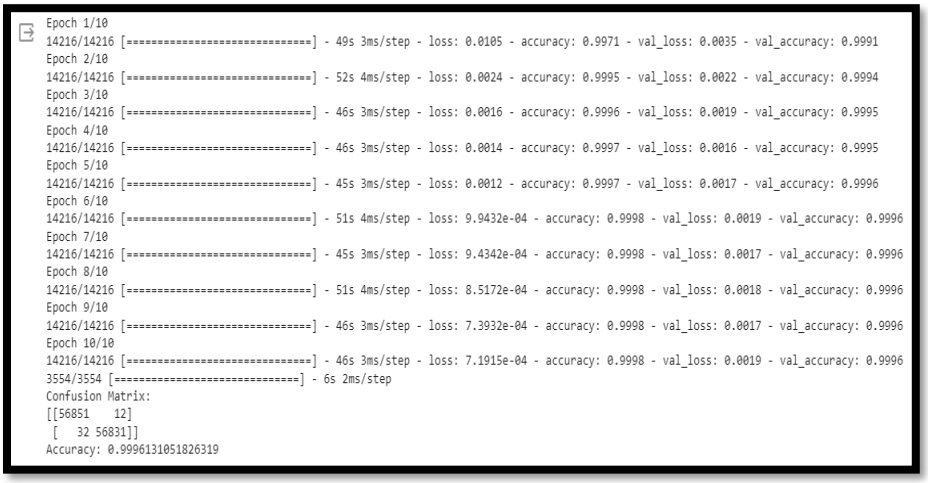

Fig. 2. Shows results of LSTM using 10 epochs

Figure 2 shows training and validation accuracy using LSTM. Batch size is 32 and result of 10 epochs is shown in figure. At the end, confusion matrix is also shown along with accuracy. As training and validation accuracy are close to each other so it indicates model will perform well on unseen data also.

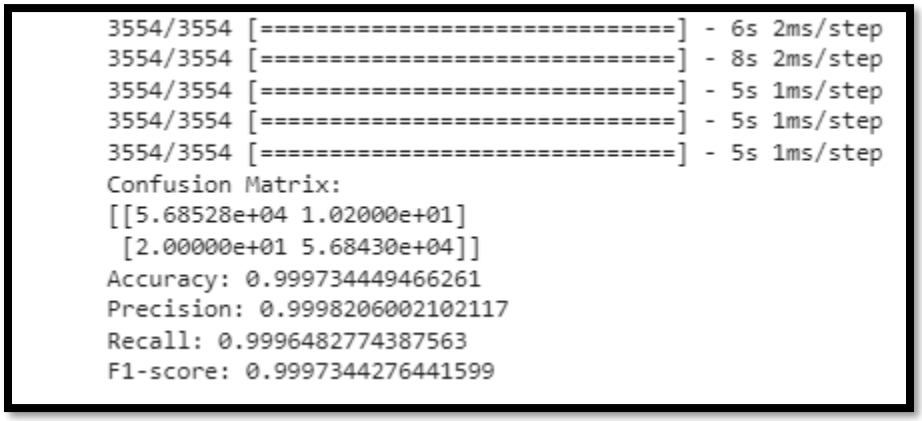

Fig. 3. Shows results of LSTM using 5 k-folds

Figure 3 shows results of Long Short-Term Memory using k folds cross validation technique. Here 5 k folds have been used. Then confusion matrix is calculated. Moreover, average Accuracy, Precision, Recall, F1-score are also calculated.

Discussion

Obtained results by applying machine learning and deep learning models show that Random Forest & Extra Trees Classifier are best among all. XGBoost is also good option if we compare its training time with Random Forest and Extra Trees Classifier. Moreover, if we talk about Long short-term memory then it is also best among all neural networks for the analysis of sequential type data.

On comparison with existing work that have already been done in this field, it is obvious that our results outperform those results. Here is the brief comparison:

Table 3. comparison of results with other papers

Paper	Year	Model	Accuracy	Precision	Recall	F1-score
[1]	2021	Random Forest	0.9677	-	-	-
[1]	2021	Logistic Regression	0.9516	-	-	-
[2]	2022	AllKNN-CatBoost	0.999	-	0.9591	0.8740
[3]	2022	Decision Tree	0.9993	-	-	-
[3]	2022	KNN	0.9995	-	-	-
[3]	2022	Logistic Regression	0.9991	-	-	-
[3]	2022	SVM	0.9993	-	-	-
[3]	2022	Random Forest	0.9992	-	-	-
[3]	2022	XG Boost	0.9994	-	-	-
[3]	2022	CNN	0.99	-	-	-
[4]	2021	DCNN	0.99	-	-	-
[5]	2021	SVM	0.9349	0.9743	0.8976	-
[5]	2021	KNN	0.9982	0.7142	0.0393	-
[5]	2021	ANN	0.9992	0.8115	0.7619	-
Ours	2024	Random Forest	1.00	1.00	1.00	1.00
Ours	2024	Extra Trees Classifier	1.00	1.00	1.00	1.00
Ours	20204	XGBoost	1.00	1.00	1.00	1.00
Ours	2024	CatBoost	1.00	1.00	1.00	1.00
Ours	2024	LGBM	1.00	1.00	1.00	1.00
Ours	2024	Logistic Regression	0.97	0.97	0.97	0.97
Ours	2024	LSTM	0.9997	0.9998	0.9996	0.9997

5- Conclusion

Credit card fraud detection is one of the biggest challenges faced by financial departments. It not only affects the banking system but also the customers of banks. In the past, various machine learning models have been applied in order to avoid these types of frauds. But still, 100% protection against this fraud is not possible. Now a days, neural networks are the center of attention of many researchers. In this paper, first of all, preprocessing of data is performed. Then data set is divided into testing and training parts. 70% is reserved for training purpose and 30% for testing of

proposed model. Moreover, k folds technique is also used for cross validation. Multiple evaluation metrics have been calculated in order to evaluate the performances of various machine learning algorithms. Six machine learning and one deep learning algorithms are applied. Their results outperform the earlier obtained results. Among them, Random Forest and Extra trees classifier perform best by giving highest validation accuracy.

Traditional machine learning models usually perform very well as described above but they lack to identify any type of patterns in data set. In the case of credit card fraud detection, data set includes spending patterns of person. So, machine should learn the spending patterns of the owner and if any type of deviation occurs then it should generate alarm for both the financial department and card holder. For this type of pattern recognition, neural networks are used. And among neural networks, long short-term memory LSTM performs well. In the future, if data set will be used from original bank instead of kaggle website, then results would be more realistic. And predicted model in that case, would be more effective in the detection of these types of fraud. But right now, no bank is willing to share such a sensitive information regarding their department and customers.

Acknowledgements. The work is based on MSc course work submitted to the Faculty of Computer Science & Information Technology, Superior University Gold Campus, Lahore.

References:

[1] Tanouz D, Subramanian RR, Eswar D, Reddy GP, Kumar AR, Praneeth CV (2021) Credit card fraud detection using machine learning. IEEE, pp 967–972

[2] Alfaiz NS, Fati SM (2022) Enhanced Credit Card Fraud Detection Model Using Machine Learning. Electronics 11:662

[3] Alarfaj FK, Malik I, Khan HU, Almusallam N, Ramzan M, Ahmed M (2022) Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms. IEEE Access 10:39700–39715

[4] Chen JI-Z, Lai K-L (2021) Deep Convolution Neural Network Model for Credit-Card Fraud Detection and Alert. JAICN 3:101–112

[5] Rb A, Kr SK (2021) Credit card fraud detection using artificial neural network. Global Transitions Proceedings 2:35–41

[6] Zorion PK, Sachan L, Chhabra R, Pandey V, Fatima DrH (2023) Credit Card Financial Fraud Detection Using Deep Learning. SSRN Journal. https://doi.org/10.2139/ssrn.4629093

[7] Faculty of Sciences IPSS, University Mohammed V, Rabat, Morocco, Benchaji I, Douzi S, Ouahidi BE (2021) Credit Card Fraud Detection Model Based on LSTM Recurrent Neural Networks. JAIT 12:113–118

[8] Garg V, Chaudhary S, Mishra A (2021) ANALYSING AUTO ML MODEL FOR CREDIT CARD FRAUD DETECTION. ijircst. https://doi.org/10.21276/ijircst.2021.9.3.5

[9] Alshawi B (2023) Utilizing GANs for Credit Card Fraud Detection: A Comparison of Supervised Learning Algorithms. Eng Technol Appl Sci Res 13:12264–12270

[10] Mangathayaru DN, Kumar NR, Kumar DGR (2023) FRAUDULENT TRANSACTION DETECTION BY MACHINE AND DEEP LEARNING ALGORITHMS. 30:

[11] Ileberi E, Sun Y, Wang Z (2021) Performance Evaluation of Machine Learning Methods for Credit Card Fraud Detection Using SMOTE and AdaBoost. IEEE Access 9:165286–165294

[12] Wang Y (2023) Fraud detection based on FS-SMOTE model for credit card. HSET 70:316–323

[13] Esenogho E, Mienye ID, Swart TG, Aruleba K, Obaido G (2022) A Neural Network Ensemble With Feature Engineering for Improved Credit Card Fraud Detection. IEEE Access 10:16400–16407

[14] Nguyen VB, Dastidar KG The Importance of Future Information in Credit Card Fraud Detection.

[15] Carcillo F, Le Borgne Y-A, Caelen O, Kessaci Y, Oblé F, Bontempi G (2021) Combining unsupervised and supervised learning in credit card fraud detection. Information Sciences 557:317–331

[16] Lin T-H, Jiang J-R (2021) Credit Card Fraud Detection with Autoencoder and Probabilistic Random Forest. Mathematics 9:2683

[17] Malik EF, Khaw KW, Belaton B, Wong WP, Chew X (2022) Credit Card Fraud Detection Using a New Hybrid Machine Learning Architecture. Mathematics 10:1480

[18] Ileberi E, Sun Y, Wang Z (2022) A machine learning based credit card fraud detection using the GA algorithm for feature selection. J Big Data 9:24

[19] Lebichot B, Verhelst T, Le Borgne Y-A, He-Guelton L, Oble F, Bontempi G (2021) Transfer Learning Strategies for Credit Card Fraud Detection. IEEE Access 9:114754–114766

[20] Ghosh Dastidar K, Jurgovsky J, Siblini W, Granitzer M (2022) NAG: neural feature aggregation framework for credit card fraud detection. Knowl Inf Syst 64:831–858

[21] Al Balawi S, Aljohani N (2023) Credit-card Fraud Detection System using Neural Networks. IAJIT. https://doi.org/10.34028/iajit/20/2/10

[22] Khalid AR, Owoh N, Uthmani O, Ashawa M, Osamor J, Adejoh J (2024) Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach. BDCC 8:6

[23] Azim Mim M, Majadi N, Mazumder P (2024) A soft voting ensemble learning approach for credit card fraud detection. Heliyon 10:e25466

[24] Aslam A, Hussain A (2024) A Performance Analysis of Machine Learning Techniques for Credit Card Fraud Detection. JAI 6:1–21

[25] Bandari M, Nayudu GSH (2023) FRAUD TRANSACTIONS DETECTION USING MACHINE LEARNING. 13:

[26] Kaggle: Your Machine Learning and Data Science Community. https://www.kaggle.com/. Accessed 24 Feb 2024

[27] Iqbal, Muhammad Javaid, Usama Ijaz Bajwa, Ghulam Gilanie, Muhammad Aksam Iftikhar, And Muhammad Waqas Anwar. "Automatic Brain Tumor Segmentation From Magnetic Resonance Images Using Superpixel-Based Approach." Multimedia Tools And Applications 81, No. 27 (2022): 38409-38427.

[28] Iqbal, Muhammad Javaid, Muhammad Waseem Iqbal, Muhammad Anwar, Muhammad Murad Khan, Abd Jabar Nazimi, And Mohammad Nazir Ahmad. "Brain Tumor Segmentation In Multimodal MRI Using U-Net Layered Structure." Comput Mater Contin 74, No. 3 (2022): 5267-5281.

[29] NOOR, FATIMA, MUHAMMAD JAVAID IQBAL, SOBIA YAQOOB, SHAHID MEHMOOD, ARFAN JAFFAR, And INAM UL HAQ. "Depression Detection In Social Media Using Bagging Classifier." DEPRESSION 42, No. 01-2023 (2023).

[30] Chaudhry, Nadeem Jabbar, M. Bilal Khan, M. Javaid Iqbal, And Siddiqui Muhammad Yasir. "Modeling & Evaluating The Performance Of Convolutional Neural Networks For Classifying Steel Surface Defects." Journal Of Artificial Intelligence (2579-0021) 4, No. 4 (2022).

[31] Rehman, Laiba, Muhammad Javaid Iqbal, Saba Ramzan, Sobia Yaqoob, Inam Ul Haq, Arfan Jaffar, And Sharjeel Nawaz. "Long-Lived Bugs Prediction Using Machine Learning Approaches."

[32] Iqbal, Muhammad Javaid, Muhammad Usman Nasir, Muhammad Umer, Muhammad Waseem Iqbal, Arfan Jaffar, And Ali Asif. "Blindness Detection Using Machine Learning Approaches."