Advances and Applications in Mathematical Sciences
Volume 21, Issue 10, August 2022, Pages 5801-5810

CREDIT CARD FRAUD DETECTION USING MULTIPLE MODELS: A COMPARATIVE ANALYSIS

TANYA AGRAWAL[1], SAI KALKI JAJULA[1], ANIL KUMAR MISHRA[1] and AKSHAT AGRAWAL[1]

[1]Department of Computer Science
Amity University, Gurugram, India
E-mail: tanyaagarwal.cool@gmail.com
saikalkij@gmail.com
akmishra2@ggn.amity.edu
akshatag20@gmail.com

Abstract

Financial fraud is a growing menace to the financial industry with far-reaching implications. As a result, financial institutions must be vigilant in detecting and preventing fraudulent activity. As the credit card has become a widely used payment mechanism, fraud involving credit card payment technology is fast increasing. As a result, it is the financial institution's responsibility to adopt a default strategy for preventing fraudulent activity. Although most of the work in this field has been done using standard mathematical and machine learning approaches. This research paper recounts the comparative analysis of several models, including LSTM, GRU, CNN, and machine learning models. The credit card detection function's goal is to build a machine learning model for current credit card payments that includes both fraud and non-fraud, and then use that model to assess whether a new incoming transaction is fraudulent or not. In-depth learning algorithms are actually a form of machine learning method which uses numerous line-based processing units to extract and convert features. The features found in one layer are used to process the next layer. Deep Learning algorithms acquire central concepts from both crude and focused input in this way. The findings provided by these models are generally positive, although we get to know that LSTM underperformed than rest of the models while the SVM and CNN model gave unexpected results.

1. Introduction

Fraudulent use of a credit card is a crime. It is wreaking havoc on financial institutions and individuals alike. Credit cards are called "excellent

2020 Mathematics Subject Classification: 68M12.

Keywords: Fraud detection, Deep Learning, Neural Networks, Machine Learning.

Received January 20, 2022; Accepted February 25, 2022

intents for fraud" because attackers may make a lot of money in a short amount of time with little risk, and most frauds are discovered after a few days [1]. Fraudsters require sensitive numbers in order to commit offline credit card fraud.

There are two types of purchases made with a credit card. There is a physical card and a virtual card.

When making a card payment or a purchase physically, the cardholder hands over his or her card to the seller. The attacker must steal a credit card in order to make fraudulent payments on this type of purchase. You only need to know the card's important details to make a virtual purchase for example expiration date of the card, card number and the security code. Generally, these kinds of purchases are made over the phone or online. The legitimate cardholder is usually completely unaware when their card details has already been seen or stolen [2]. As the credit card has become a widely used payment mechanism, fraud involving credit card payment technology is fast increasing. In this sector, in-depth learning algorithms like LSTM‟s and repeated neural networks have recently been demonstrated to be promising. As a result, this study will seek to provide a thorough examination and comparison of solutions to the problem of inequality. Also, point out their flaws to assist researchers in focusing their emphasis on real-world difficulties.

2. Data

The dataset used in this investigation came from a datacenter run by a commercial bank. Over the course of eight months, this database collected about 80 million anonymous credit card transactions. There is a wealth of transaction and account data available, including the amount of money traded, the type of vendor used, and account opening date. The output variable is set to '1' for fraudulent transactions. The output variable is set to '0' for legal transactions.

3. Data Preprocessing

The main focus of data preprocessing is lining up the original/initial business data with the new business model, remove qualities that aren't

relevant to the data mining purpose, and it delivers the data which is simplified, accurate and clean so as to improve the quality and efficiency [3]. Data preparation contains steps like: Data cleaning, Data integration, Data transformation, and reduction of data. The purpose of Data cleaning is smoothing the noisy data, filling up the null values and clearing data to make it worth. The purpose of Data integration is to combine the data from variety of resources into a single location. The purpose of Data conversion is to transform the data into an excavation-friendly format and the purpose of Data compression is to compile the dataset using compressed data, which is way smaller than the original one but yet maintains the integrity.

4. Feature Engineering

The use of knowledge acquisition aspects is found to improve credit card fraud detection algorithms' predictive accuracy [4]. We discovered and developed a few essential and industry-standard predictors during our investigation. These are the predictors which are invented and integrated into the data. Monthly transaction frequency which provide information about spending nature of the account holder, use of dummy data/information for the missing values, use of the dummy variables for indicating purchase some merchants, such as petrol stations and restaurants, are frequently used by fraudsters to test a stolen card before making a larger payment [5], account history characteristics such as the number of transactions from the account throughout the dataset's 8-month period and use of a dummy factor which indicates whether a transaction authorization amount which is made at a particular merchant is larger than 10 percent of the standard deviation of the mean of the merchant's non-fraudulent transactions.

5. Under Sampling

Because the dataset has an elegance imbalance, genuine transactions were under-sampled at the account stage when the additional functions were introduced, as previously mentioned [6]. As a result, we had to make sure that each account we sampled in the training set had all of its transactions. Data was segregated into two databases, one fraudulent and the other non-fraudulent. A sampling ratio of 1.10 (fraudulent to non-fraudulent) was used to detect credit card fraud. Finally, to represent categorical data, we used one-hotencoding.

6. Methods

LSTM network, GRU model, CNN model, and machine learning models are employed in this study report.

6.1 Overview of Recurrent Neural Network. The output of the previous step is provided as an input to the current step of the Recurrent Neural Network (RNN) [7]. When we need to predict the next coming word in a sentence, we will need the previous word, so we need to remember the previous word. As a result, RNNs have been developed that use hidden layers to solve problems [8].

6.2 Lstmnetworks. Long Short Term Memory networks ("LSTMs") are a type of RNN that can learn long-term dependencies. LSTMs are specially designed to prevent long-term dependence problems [9]. We don't have to make a long effort to learn the knowledge. It's like a second nature for them! LSTM"s uses a combination of "gates" that regulate how data enter and exit the network in sequence. A typical LSTM has three gates: a forget gate, an input gate, and an outputgate.

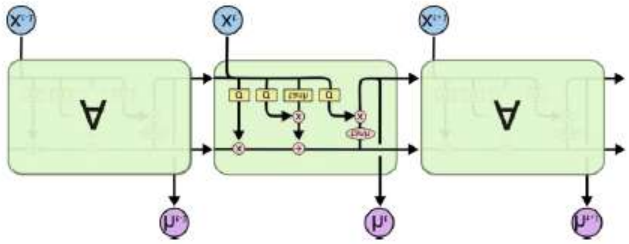

Figure 3. An LSTM's repeating module has four levels that interact [12].

6.3 GRU. GRUs has helped to improve recurrent neural networks. The distinction between GRU and RNN lies in the operation and gates connected with every GRU unit as well as performance given by GRU is better than RNN.GRU requires minimal memory and performs faster than RNN as well as LSTM since it employs fewer training parameters. It varies from LSTM in that it just has three gates and it does not keep track of the state of the internal cell [10]. The GRU's gates are broken down as Update Gate which specifies what information from the past must be passed on to the future and the Reset Gate is the second gate, and it regulates how much of the previous information is to be fully wiped.

To solve the classic RNN's vanishing gradient problem, GRU actually uses update gate and the reset gate. What information is conveyed to the output is determined by two vectors. In essence, there are two vectors. They can be taught to filter out information that isn't relevant to the prediction in addition to remembering information from the distant past [11].

6.4 CNN (Convolutional Neural Network). It is a deep learning model which takes two inputs in convolution operation. The two inputs are image matrix and a filter or a kernel, multiplies them together to get another matrix called "Featuremap" [13].

There are mainly 6 layers in complete CNN model making. Convolution Layer which contains the elements like: Input image, feature/filter, feature map, strides and padding. Pooling Layer which reduces the number of parameters or the spatial volume from the feature map so that the model does not become computationally expensive. In this study we have make use of Max Pooling layer. Batch Normalization enables substantially higher learning rates, allowing networks to train at a faster rate. Flatten Layer is used when we got a multidimensional input and we want to make it linear to pass it onto a denselayer. Dropout refers to dropping out or ignoring units or certain set of neurons during the training phase so as to prevent overfitting. Fully Connected Layer is mainly the dense layer which is used for creating the neural network containing neurons, activation function etc. It also contains the output layer.

The CNN structure used here is the VGGNet structure which consists of 16 convolutional layers and widely used for its uniform structure. Here, we used only 3 convolutional layers according to our dataset and it performed very well.

6.5 Machine Learning Approach.

7.5.1 SVM (Support VectorMachine). SVM is a supervised learning algorithm which can be used for both predicting the continuous values and the categorical variables. Here, non-linear SVM model is used.

There are 4 important terms used in support vector model. Support Vectors are the data points which are located near the boundary lines. These points are known as support vectors because these 2 points are supporting

the whole algorithm. Boundary line creates the margin or we can say that these lines are drawn to the side of the support vectors. Hyperplane is a line that lays between the boundary lines. It separates the data points and predicts the categorical values. Kernel is used to map the data of lower dimension to higher dimension. The default kernel is the "rbf" (radial basis function) kernel which is used in this study. It is a non-linear kernel and is used when the boundaries are hypothesized to be curve shaped.

7.5.2 NaïveBayes. This technique is a classifier technique which is based on bayes theorem and it assumes that independent variables are actually independent of each other [14]. It's naive since it makes the unrealistic assumption that the probabilities of different traits are unrelated to one another. Credit card fraud detection uses a probability-based classifier that is the naive bayes. The probabilities of the target classes and the probabilities of the test data are calculated in the probability-based classification approach. The target set is the test set that is closest to the probability class.

7.5.3 K-Nearest Neighbor (KNN). To categorize a new case, the K-NN algorithm first compares it with the cases which are previously classified so as to see which ones are most likely the existing ones. It is also known as a "lazy learner" algorithm since it does not instantly learn from the training set, but rather stores the information and executes an action on it when it is time to classify. The classifier returns the categorization for the input point based on the majority of these elements. The parameter k was tuned for optimal performance for $k = 1, 3, 5, 7, 9, 11, 13, 15$ and $k = 3$. Classifier uses $k = 3$ as its input. Weight function is valued as "uniform" so that each neighborhood's points are equally weighted. The distance metrics used here is Euclidean distance with $p = 2$, here "p" is the parameter for power where $p = 2$ is used for Euclidean distance.

7. Results

According to all evaluation metrics in final experience, the support vector machine strategy outperformed the rest. It had the maximum specificity and accuracy as 96.96% when kernel was set as "rbf" as we are dealing with the non-linear SVM model. Coming to KNN classifier which also performed better after SVM and it gives 95.95% accuracy when neighbors were 3. In

naïve bayes classifier when used Gaussian classifier it gave an accuracy of 94.94% which actually performs very well when we deal with continuous dataset. There was a no noticeable difference in performance between the Gaussian and Bernoulli naïve bayes as Bernoulli also depicted 94.94% accuracy in this case. CNN in this case also gave an upright accuracy i.e. 96.96% with test loss as 0.11 which is very similar to SVM‟s accuracy. As VGGNet does not contain batch normalization and dropout layer but here we have involved its use with the convolutional layers to increase its performance i.e. the overall speed, learning rate and removing redundant neurons to avoid over-fitting. Here, three convolutional layers, batch normalization layers, dropout layers, a flatten layer and a dense layer with 128 units are used. Optimizer used is "Adam" with binary cross entropy as a loss function and overall performance by this neural network was quite good. GRU gave an accuracy of 94.94% and the test loss as 0.16 which is less than the performance shown by CNN model but quite well than the performance of LSTM as it is less complex and require less parameters. Here, LSTM underperformed these two neural networks and gave an accuracy of 93.93% with the test loss as 0.17. In first experience we observe that SVM and CNN performed well again but Naïve Bayes has underperformed in this case. LSTM has also carried out well but loss and AUPRC is higher in 1st experience. GRU underperformed among neural networks.

Table 2. Performance comparison of different methods (1st experience).

Method	Accuracy	Loss	AUPRC	Performance
SVM	96.96%	0.11	0.91	Highest Accuracy
KNN	90.54%	0.31	0.85	Highest Loss
Naïve Bayes (Gaussian)	87.91%	0.21	0.88	Low accuracy as comparison
Naïve Bayes (Bernoulli)	87.91%	0.21	0.88	Low accuracy as comparison
CNN	94.93%	0.18	0.89	Higher accuracy

GRU	93.66%	0.20	0.82	High Accuracy
LSTM	94.21%	0.21	0.84	High accuracy

Table 3. Performance comparison of different methods (2nd experience).

Method	Accuracy	Loss	AUPRC	Performance
SVM	96.96%	0.09	0.92	Highest Accuracy
KNN	95.95%	0.11	0.89	Higher Accuracy
Naïve Bayes (Gaussian)	94.94%	0.15	0.85	High Accuracy
Naïve Bayes (Bernoulli)	94.94%	0.15	0.85	High Accuracy
CNN	96.96%	0.11	0.91	Highest Accuracy
GRU	94.94%	0.16	0.88	High Accuracy
LSTM	93.93%	0.17	0.87	Low accuracy as comparison

8. Conclusion

In machine learning models, this study compares the performance of Support vector machine, Naive Bayes, and K-nearest neighbor models in binary classification of imbalanced credit card fraud data, as well as the performance of CNN, GRU, and LSTM networks in deep learning neural networks. The following is a summary of the paper's contribution:

1. Three classifiers (SVM, Naive Bayes, and KNN) based on different machine learning techniques are they are trained on credit card transaction data, and their performance is evaluated and compared using relevant metrics.

2. Three popular neural networks (CNN, GRU, LSTM) are trained and made an analysis on how these three performed and which outperformed the other based on the different parameters, neural network layers, optimizers, learning rate, epochs used.

3. Classifier performance varies depending on the evaluation metric. When the results of the experiment are compared, the SVM in machine learning and CNN in neural networks show substantial performance. Future study could look into meta-classifiers and metal earning techniques for dealing with severely skewed credit card fraud data. Other sampling method's effects can also be examined.

References

[1] Maes Sam, et al., Credit card fraud detection using Bayesian and neural networks, Proceedings of the 1st international naiso congress on neuro fuzzy technologies 7 (2002).

[2] Adewumi, O. Aderemi and Andronicus A. Akinyelu, A survey of machine-learning and nature-inspired based credit card fraud detection techniques, International Journal of System Assurance Engineering and Management 8(2) (2017), 937-953.

[3] J. Bolton Richard and David J. Hand, Unsupervised profiling methods for fraud detection, Credit scoring and credit control VII (2001), 235-255.

[4] Rahman, Rizwana, Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative survey, (2021).

[5] Patidar Raghavendra and Lokesh Sharma, Credit card fraud detection using neural network, International Journal of Soft Computing and Engineering (IJSCE) 1 (2011), 32-38.

[6] Pawan Verma and Prateek Agrawal, Study and Detection of Fake News: P2C2 Based Machine Learning Approach, 4th International Conference on Data Management, Analytics and Innovation (ICDMAI), New Delhi, Springer (2020), 261-278.

[7] K. Goel, C. Gupta, R. Rawal, P. Agrawal and V. Madaan, FaD-CODS Fake News Detection on COVID-19 Using Description Logics and Semantic Reasoning. International Journal of Information Technology and Web Engineering (IJITWE) 16(3) (2021), 1-20.

[8] P. K. Verma, P. Agrawal, V. Madaan and C. Gupta, UCred: fusion of machine learning and deep learning methods for user credibility on social media, Social Network Analysis and Mining 12(1) (2022), 1-10.

[9] A. Shankhdhar, P. K. Verma, P. Agrawal, V. Madaan and C. Gupta, Quality analysis for reliable complex multiclass neuroscience signal classification via electroencephalography, International Journal of Quality and Reliability Management, (2022).

[10] C. Gupta, D. Gaur, P. Agrawal and D. Virmani, HuDA_COVID Human disposition analysis during COVID-19 using machine learning. International Journal of E-Health and Medical Communications (IJEHMC) 13(2) (2021), 1-15.

[11] Chaudhary Khyati, Jyoti Yadav and Bhawna Mallick, A review of fraud detection techniques: Credit card, International Journal of Computer Applications 45(1) (2012), 39-44.

[12] Ghosh Sushmito and Douglas L. Reilly, Credit card fraud detection with a neural-network, System Sciences, Proceedings of the Twenty-Seventh Hawaii International Conference on 3 (1994).

[13] Abhishek Sharma, Prateek Agrawal, Vishu Madaan and Shubham Goyal, Prediction on Diabetes Patient's Hospital Readmission Rates, In International Conference on Advanced Informatics for Computing Research (2019), 1-5.

[14] Sherstinsky Alex, Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network, Physica D: Nonlinear Phenomena 404 (2020), 132306.

[15] Prateek Agrawal, Vishu Madaan, Aditya Roy, Ranjna Kumari, Harshal Deore, "FOCOMO: Forecasting and monitoring the worldwide spread of COVID-19 using machine learning methods, Journal of Interdisciplinary Mathematics, 1-25. DOI:10.1080/09720502.2021.1885812.

[16] Vishu Madaan, Aditya Roy, Charu Gupta, Prateek Agrawal, Anand Sharma, Christian Bologa and Radu Prodan, XCOVNet: Chest X-ray Image Classification for COVID-19 Early Detection Using Convolutional Neural Networks, New Generation Computing, pp. 1-15. DOI: 10.1007/s00354-021 00121-7.

[17] Charu Gupta, Prateek Agrawal, Rohan Ahuja, Kunal Vats, Chirag Pahuja and Tanuj Ahuja, Pragmatic Analysis of Classification Techniques based on Hyperparameter Tuning for Sentiment Analysis, In International Semantic Intelligence Conference (2021), 453-459.

[18] Prateek Agrawal, Anatoliy Zabrovskiy, Adithyan Ilagovan, Christian Timmerer and Radu Prodan, FastTTPS: Fast approach for video transcoding time prediction and scheduling for HTTP adaptive streaming videos, Cluster Computing (CLUS), 1-17, https://doi.org/10.1007/s10586-020-03207-x.

[19] Prateek Agrawal, Deepak Chaudhary, Vishu Madaan, Anatoliy Zabrovskiy, Radu Prodan, Dragi Kimovski and Christian Timmerer, Automated bank cheque verification using image processing and deep learning methods, Multimedia tools and applications (MTAP) 80(1) 5319-5350. https://doi.org/10.1007/s11042-020-09818-1

[20] Kaur Rupinder, Vishu Madaan and Prateek Agrawal, Diagnosis of Arthritis Using K-Nearest Neighbor Approach, In International Conference on Advanced Informatics for Computing Research (2019), 160-171.

[21] Abiodun Oludare Isaac, et al., State-of-the-art in artificial neural network applications: A survey, Heliyon 4(11) (2018), e00938.

[22] Pisner A. Derek and David M. Schnyer, Support vector machine, Machine learning. Academic Press (2020), 101-121.