Global Transitions Proceedings 2 (2021) 35–41

Contents lists available at ScienceDirect

Global Transitions Proceedings

journal homepage: www.elsevier.com/locate/gltp

Credit card fraud detection using artificial neural network

Asha RB*, Suresh Kumar KR

Department of ISE MSRIT, Bengaluru, Karnataka, India

ARTICLE INFO

Keywords:
Artificial neural network
Credit card
Fraud
k-Nearest Neighbor
Machine learning and support vector machine

ABSTRACT

Frauds in credit card transactions are common today as most of us are using the credit card payment methods more frequently. This is due to the advancement of Technology and increase in online transaction resulting in frauds causing huge financial loss. Therefore, there is need for effective methods to reduce the loss. In addition, fraudsters find ways to steal the credit card information of the user by sending fake SMS and calls, also through masquerading attack, phishing attack and so on. This paper aims in using the multiple algorithms of Machine learning such as support vector machine (SVM), k-nearest neighbor (Knn) and artificial neural network (ANN) in predicting the occurrence of the fraud. Further, we conduct a differentiation of the accomplished supervised machine learning and deep learning techniques to differentiate between fraud and non-fraud transactions.

1. Introduction

In recent years, as there is advancement of technology, most of them are using credit card for buying their needs so the frauds associated with it is also rising gradually. In the present world, almost all the enterprises from small to big industries are using the credit card as mode of payment. Credit card fraud is happening in all organization such as appliances industry, automobile industry, banks and so on. Many of the process like data mining, machine learning algorithmic approaches are applied to identify the fraud in the credit card transactions but did not get considerable result. Hence, there is a need of effective and efficient algorithms to be developed that works significantly. we try to avoid the fraudster using our credit card before the transaction gets approved by using artificial neural network algorithm and compared with few other machine learning algorithms.

1.1. Overview of credit card fraud detection

Fraud is an offensive activity, carried out by an unauthorized person by cheating innocent. Credit card fraud involves stealing the essential credentials from the cardholder and using it unauthorized manner by the fraudsters either by using phone calls or SMS. This fraud in credit card may also happen using some software applications that are under the control of fraudsters.

The credit card fraud detection takes place as: the user or the customer enters the necessary credentials in order to make any transaction using credit card and the transaction should get approved only upon being checked for ay fraud activity. For this to happen, we first pass the transaction details to the verification module where, it is classified under fraud and non-fraud categories. Any transaction that is put under fraud category is rejected. Otherwise, the transaction gets approved.

1.2. Classifications of credit card frauds [13]

1 Application fraud: When a fraudster acquires the control over the application, steals the credentials of customer, and makes a fake account and then the transactions takes place.
2 Electronic or manual card imprints: In this form of fraud, the fraudster skims the information from the magnetic strip which is present on the card then uses the credentials and fraud transactions are carried out
3 Card not present: This is a type of credit card in which physical card is not present during transaction
4 Counterfeit card fraud: the fraud type in which the fraudster will copies all the data from magnetic strip and real card looks like original card works as original card only. This card used for fraud.
5 Lost/stolen card: This type of fraud is due to loosing of the card by the cardholder or by the stealing the card from the cardholder.
6 Card id theft: the type of fraud in which the id of the cardholder is stolen and fraud takes place.
7 Mail non-received card fraud: while issuing the credit card there will be procedure of sending a mail to the recipient, here fraud can occur by defrauding the mail or phishing.
8 Account Takeover: here the fraudster will take the complete control of the account holder and make a fraud.

* Corresponding author.
E-mail addresses: ashabatageri@gmail.com (A. RB), sureshkumar@msrit.edu (S.K. KR).

https://doi.org/10.1016/j.gltp.2021.01.006

Available online 23 January 2021

9 Fake fraud in website: fraudster will introduce a malicious code which does their work in the website
10 Merchant collision: In this fraud type, cardholder details are shared third party or the fraudster by merchants without cardholder authorization.

The fraud in credit card transaction occurs when the stealer uses the other person card without authorization of the respective person by stealing the necessary information like PIN, password and other credentials with or without the physical card. Using fraud detection module involving machine learning and deep learning, we can find out whether the upcoming transaction is fraud and legitimate.

Machine Learning is the treading and most used technology because of its various applications and less time consumption, more accurate in result. Machine learning is a technology that deals with the algorithm, which provides the computer, a capability to study and advance through experience without being explicitly programmed. Machine learning has application in multiple fields. For example, medical, diagnosis, regression etc. Machine learning involves the combination of algorithm and statically models which allow computer to perform the task without hard coding then a model is build through a training data and then it is tested on the trained model.

Deep learning is a part of machine learning techniques that makes use of neural networks. Some of methods that come under deep learning are artificial neural network, Convolution neural network, autoencoders, recurrent neural networks, restricted Boltzmann machine etc. Deep learning makes uses of neural networks, which resembles the human brain in processing the data and making the decisions.

1.3. Problem statement

Now -a-days, most of them are using credit cards for buying the goods which are so much in need but can't afford at the moment. In order to meet the needs credit cards are used and the fraud associated with it is also increasing so there is a need of developing a model that's fit well and predicts at higher accuracy.

1.4. Objectives

- The main objective of the research is to find a fraudulent transactions in credit card transactions
- Comparison between the supervised learning and deep learning and deep learning algorithm outperformed based on accuracy.

1.5. Existing system

The existing systems are carried out by considering machine learning algorithms like Support Vector Machine, Naïve Bayes, k-Nearest Neighbor and so on and some of them used random dataset. Very few have used artificial neural network for credit card fraud detection.

1.6. Proposed system

The Proposed system uses the Artificial Neural Network to find the fraud in the credit card transactions. Performance is measured and accuracy is calculated based on prediction. And also classification algorithms such as Support vector machine and k-Nearest Neighbor are used to build a credit card fraud detection model. We compare all the three algorithms used in the experiment and made a decision that artificial neural networks predicts well than system developed using support vector machine and k-nearest neighbor algorithms. The dataset used in the experiment consist of 31 attributes out of which 30 attributes consist of information related to name, age, account information and so on and last attribute give the outcome of the transaction in either 0 or 1.

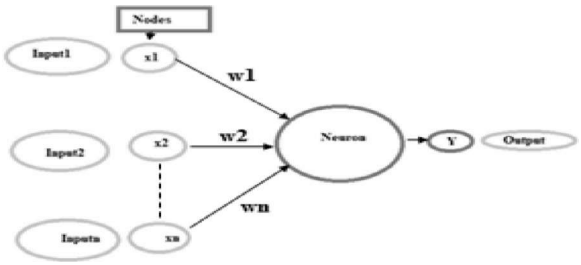

Fig. 1. Architecture of Artificial neural network.

ANN is biologically inspired by human brain. The neurons are interconnected in the human brain like the same nodes are interconnected in artificial neural network. Fig. 1 depicts the structure of ANN with input, output and hidden layers. Inputs are x1, x2…Xn and output is y. w1…wn are the weights associated with inputs x1…xn respectively. There are 15 hidden layers used in this neural network. The activation function used in our credit card fraud detection model is RELU.

2. Related work

Some of the related study made by various researchers is described in this section.

Altab Althar Taha and Sareef Jameel Malbery described that upgradation in e-commerce and communication technology have made credit card usage more popular way of payment and the fraud associated with transactions is also increasing. They have used the optimized light gradient boosting machine, where Bayesian based hyper-parameter optimization are combined to tune with parameter of light gradient boosting machine (LightGBM). In this approach they used two set of real world public dataset consisting of fraudulent and non-fraudulent transactions. Based on the comparison with other techniques, their proposed system outperformed in terms of accuracy. The proposed system produces the accuracy of 98.40%, area under receiver operating characteristics curve (AUC) of 92.88%, Precision of 97.34% and F1-score of 56.95% [1].

S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid and H. Zeineddine research describes that the credit card fraud cause huge financial loss. Most of the researchers have been working on this to provide an innovative ways to eradicate this loss and most of the available methods are costly, time consuming and labor incentive task. The authors have found out that the imbalanced classification of dataset is the main reason for the inaccurate results after many experimental studies. These imbalance classifications consist of un-balanced dataset, which caused the model to predict inaccurate and causes the financial loss. Therefore, they have found that LR, C5.0 decision tree algorithm, SVM and ANN are best algorithm based on accuracy, AUCPR and sensitivity. They have used the balanced dataset in order to train these models [2].

C. Jiang, J. Song, G. Liu, L. Zheng and W. Luan proposed a novel process with multiple stages. First they collect the transactions made by card holder, then based on the behavioural patterns transactions are aggregated, next the dataset is classified, further the model is trained and finally the model is tested. If any abnormal behaviour arises then a feedback is provided to system about the abnormal behaviour through feedback mechanism [3].

Ishan Sohony, Rameshwar Pratap, and Ullas Nambiar proposed an ensemble learning approach for credit card fraud detection as the ratio of fraud to normal transaction is bit appropriate. They observed that Random forest is best suited to provide a higher accuracy and neural networks for detecting the fraud instances. They also experimented with the large real world credit card transactions. Ensemble learning is combinations of Random forest and neural networks [4].

Phuong Hanh Tran, Kim Phuc Tran, Truong Thu Huong, Cédric Heuchenne, Phuong HienTran, and Thi Minh Huong Le research describes that in past few years' credit card fraud increased gradually. Many methods performed using machine-learning algorithms to detect the fraud transactions and block them. They introduced two new data driven approaches, which uses the optimal anomaly technique for fraud transaction in credit card transaction. The two ways are kernel parameter selection and T^2 control chart [5].

Imane Sadgali, Nawal Sael, and Faouzia Benabbou work describes that now-a-days banking transactions like online transactions, credit card transactions and mobile transactions etc. are gaining popularity because all the people prefer digital and paperless transactions. Millions of transactions carried out; all of them subjected to a type of fraud. Many of the researchers have analyzed, designed and developed the model for detecting the fraud using machine learning. They presented a comparison among the entire machine-learning algorithm to select which model is best for fraud identification in card transaction [6].

Debachudamani Prusti and Santhnu Kumar Rath designed an application with applied machine learning approaches such as Decision tree (DT), k-nearest algorithm (kNN), Extreme learning machine (ELM), Multilayer perceptron (MLP) and support vector machine (SVM) to detect the accuracy in fraud identification. They proposed a model by hybridizing the DT, SVM and kNN techniques. They used two web-based protocols such as simple object access protocol (SOAP) and Representational state transfer (REST) for efficient exchange of data across multiple heterogeneous platforms. They compared five machine learning algorithm results based on accuracy metric. SVM performed better than other algorithms by 81.63% but the hybrid system proposed by them had higher accuracy of 82.58% [7].

Mohamad Zamini and Golamali montazar proposed unsupervised credit card Fraud detection system using autoencoders based clustering. They used three hidden layers and k means is used for clustering and tested on the European dataset which performed well compared to other existing system [8].

Akila and Srinivasulu reddy proposed a misrepresentation location framework with non-overlapped risk based bagging ensemble (NRBE) model to deal with the unevenness dataset and to keep away from the noisiness contained in the transactions. They bagging model breaks all the irregularity in dataset and non-vital nature. The sacking model is reached out by pack of creation and danger based base student. The bag creation eradicates the problem faced by imbalanced data and Naïve bayes destroys the issue brought about by noisiness created in the transactions. The proposed model has been beated with 5% in BCR and BER, half of recall and 2x or 2.5x decreased expense to fraud detection by utilizing the NBRE. It was distinguished that NRBE model best suits for fraud detection and it is most appropriate for business dynamic technique [9].

M. S. Kumar, V. Soundarya, S. Kavitha, E. S. Keerthika and E. Aswini worked on project that made a model for detection of fraud in credit card transactions using Random forest techniques. The random forest algorithm (RFA) is a supervised machine learning techniques which uses the decision tree for classification of the credit card transactions and further performance is calculated using confusion matrix. The proposed system gives an accuracy of 90% [10].

Z. Li, G. Liu, S. Wang, S. Xuan and C. Jiang proposed fraud detection system via Kernel based supervised hashing (KSH). This KSH model based on approximate nearest neighbor idea. It is best suited for large dataset with high dimension data. It is for the first time KSH is used for prediction, which performs better than other existing systems [11].

Pawan Kumar and fahad Iqbal made a survey on all techniques used to detect the MasterCard fraud detection using machine-learning algorithms and evaluate the performance with the metrics. Tons of analysis carried out over this domain. They say that there is a need to use more efficient system that performs well at every situation [12].

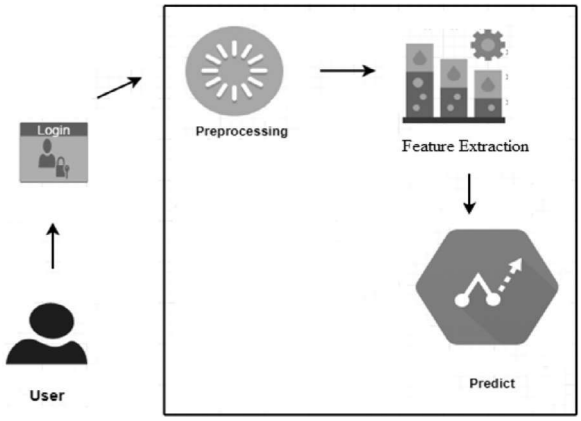

Fig. 2. Complete architecture of proposed system.

3. Architectural design

The system architecture will provide us a complete design of proposed system. It explains about the implemented system architecture.

The design of proposed system is shown in Fig. 2. The architecture depicts that first, user needs to register and then login into the system by entering the credentials like user name, password, and email-id and phone number. Than it displays the algorithm selection windows where user need to select the algorithm. After selecting the algorithm, data pre-processing takes place, which involves data cleansing, and normalization of dataset before fed into the actual model then splitting of data. Now the model is trained using Machine Learning algorithm and finally, tested and predicted the end results whether transaction is fraud or not.

4. Methods

This section explains about the implementation, which includes the algorithm used for implementation of proposed system.

In this paper, Implementations starts from loading the dataset. Than data pre-processing carried out that includes data cleansing and normalizing the data. Dataset is splitted into two dataset as train data and test data and model is trained and tested. Finally, system predicts whether transaction is fraud or non-fraud.

4.1. Programming language used

In the implementation of proposed system, we used python as programming language. Python is beginner's language, which provides various applications. In recent years, python had set the new trend because it is easy to use, interpreted, object-oriented, high-level, scripting language. Python is one of the best languages for the implementing machine learning. It provides rich packages and libraries that used in machine learning.

4.2. Packages and libraries used

Some of the python library and packages used in proposed system are as follows:

- Numpy

Numpy is a python library. Abbreviation of Numpy is numerical python library. Numpy package is used for multidimensional arrays and linear algebraic operations.

- Pandas

Pandas is a python library. Pandas is used for data analysis and data manipulation tool. It is used to read the dataset and load the dataset. It is fast, flexible when working with data.

- Scikitlearn

A python package which is suitable for statistical model and machine learning models. A best suited python package for machine learning modeling.

- Keras

Keras is advanced stage of neural network application programming interface (API). It is able of run on top of tensor flow. Keras is mainly used while implementing deep learning algorithms such as CNN, RNN because its user friendly, modularity, and easy to extensibility. It runs on both CPU and GPU. In the experiment of finding the fraud or non fraud credit card transaction we had used Keras along with backend running tensor flow. This Keras along with tenor flow backend makes excellent choice for training neural network architecture.

- MySQL

MySQL is database which is used for storage purpose. In the experiment of fraud identification in card transaction we had used MySQL for storing the user details namely user name, password, email-id and phone number. While entering into application, user needs to register by providing the credential. These credentials are stored in database. Thereafter, user needs to login by giving username and password. The application will validate the login and registered information than user is moved to next window.

- Tkinter

Tkinter is python library which is used for Graphical User interface (GUI). It can be used on both Unix and Windows platform. We can create it by importing Tkinter module then GUI is created and one or more widgets are added finally, called in loop.

4.3. Classification techniques

List of algorithm used in implementation of our experiment are:

1 Support Vector Machine
2 K-nearest algorithm
3 Artificial Neural Network

- **Support Vector Machine:**

Pseudocode:

- Importing the necessary packages

Example: import pandas as pd

- def SVM

Step 1: START
Step 2**:** Reading the dataset. pd.read.csv (file name) # reads the dataset file
Step 3: Data cleaning and preprocessing of data

- Resampling the data as normal and fraud class i.e. normal = 0 and fraud =1 under
- Under sampling of data is done
- Data is scaled (if any null value then eliminated) and normalized.
- Dataset is splitted into two set as train data and test data using split () on training data is used to split the data.

Step 4: Training the data using the SVM algorithm

- SVM classifier is called as classifier.predict () # which predicts whether transaction fraud or nonfraud.

Step 5: Calculating the fraud transactions and valid transactions, then calculating the recall, precision and accuracy and stored in the respective locations
Step 6: STOP

- **K-Nearest Neighbor**

Pseudocode:
Step 1: START
Step 2: Loading of dataset pd.read.csv (csv file) # reads the csv file and loads
Step 3: Cleaning and normalization of data

- Normal = 0
- Fraud = 1 # resampling
- Data is scaled and normalized
- Train_test_split() # splitting of dataset into train and test data

Step 4: Train the model then fit the trained model

- Trained the data using Knn classifier
- KNeighborsClassifier() # knn classifier which does classification of transactions

Step 5: Calculating the number of fraud, valid transactions and recall, precision and accuracy calculated.
Step 6**:** STOP

- **Artificial Neural Network (ANN):**

Pseudocode:
The ANN algorithm has two parts: Training part and testing part.
Training part:
Def ANN:
Step 1: START
Step2: Loading and observing the dataset

- pd.read.csv(.csv) # reads the dataset
- resampling of data
- StandardScaler() #scaling and normalization of data

Step 3: Data pre-processing

- Train_test_split() #Splitting of data

Step 4: Training the model

- Dense() #Adding data to activation function

Step 5Analyzing the model

- Prediction of fraud is made and this trained data is stored .it can used to test (training the model takes longer time so it is stored)

Step 6: STOP
Testing part:
Def ANN
It is carried out similar way only difference is that the stored trained model is used to test the data and classify it.

5. Results and discussion

5.1. Dataset

The proposed system makes use of the dataset downloaded from this website: www.kaggle.com . Dataset used is the transactions made by customer in a European bank in the year 2013–14. It consist of 31 columns, in which 30 columns are the features and the one class is the target class which decides about whether the transaction is fraud or non-valid.

5.2. Evaluation measure

The end result is evaluated based on the confusion matrix and precision, recall and accuracy is calculated. It contains two classes: actual class and predicted class. The confusion metrics depends on these features:

True Positive: in which both the values positive that is 1.

True Negative: it is case where both values are negative that is 0.

False Positive: this is the case where true class is 0 and non-true class is 1.

False Negative: It is the case when actual class is 1 and non-true class is 0.

- Precision defined as follows:

  Precision = true positive / Actual result
  Precision = true positive/(true positive + false positive)

- Recall defined as follows:

  Recall = true positive / predicted result
  Recall = true positive/(true positive + false negative)

- Accuracy defined as:

  Accuracy = (true positive + true negative)/ total

5.3. Result

The Table 1 shows the results of the used algorithms on the performance metrics such as accuracy, precision and recall.

Table 1
Represents the accuracy, recall and precision.

Algorithms	Accuracy	Precision	Recall
SVM	0.9349	0.9743	0.8976
KNN	0.9982	0.7142	0.0393
ANN	0.9992	0.8115	0.7619

The Figs. 3–5 show the screenshot of results obtained by SVN, K-NN AND ANN algorithm. It consists of count of fraud and non-fraud instances, precision, recall and accuracy.

The Fig. 6 shows the accuracy performance measure of SVM, KNN and ANN algorithms,This show that credit card fraud detection using artificial neural networks predicts at higher accuracy then Support vector machine and k-nearest neighbor algorithms for fraud detection in credit card transactions.

6. Conclusion

In this research, we have proposed a method to detect the fraud in credit card transactions that is based on deep learning. We first compare it with machine learning algorithms such as k-Nearest Neighbor, Support vector machine etc. Finally we have used the neural network, even though tough to train the model which would fit fine to model for detecting a fraud in credit card Transactions. In our model, by using an artificial neural network (ANN) which gives accuracy approximately equal to 100% is best suited for credit card fraud detection .It gives accuracy more than that of the unsupervised learning algorithms. In this

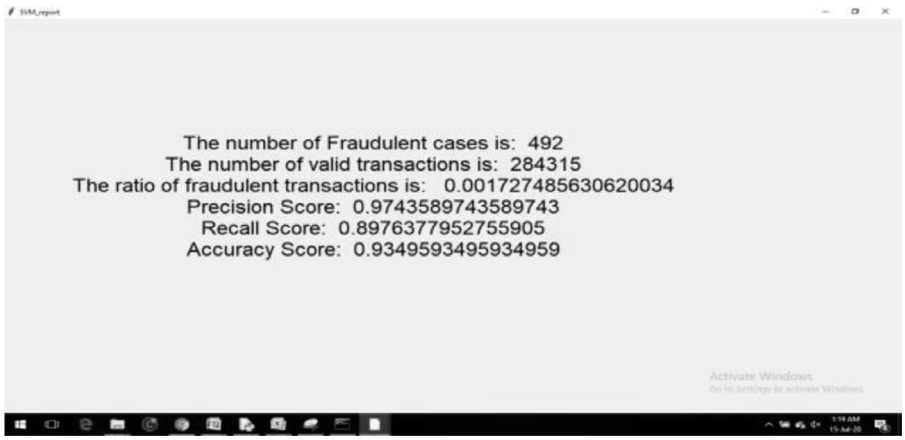

Fig. 3. shows the precision, accuracy and recall of the SVM classifier.

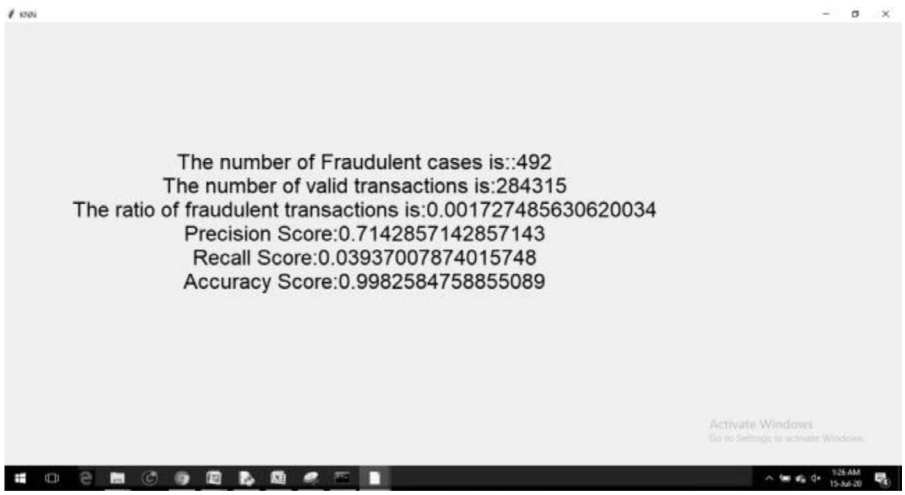

Fig. 4. shows the result of Performance metrics using K-NN classifier.

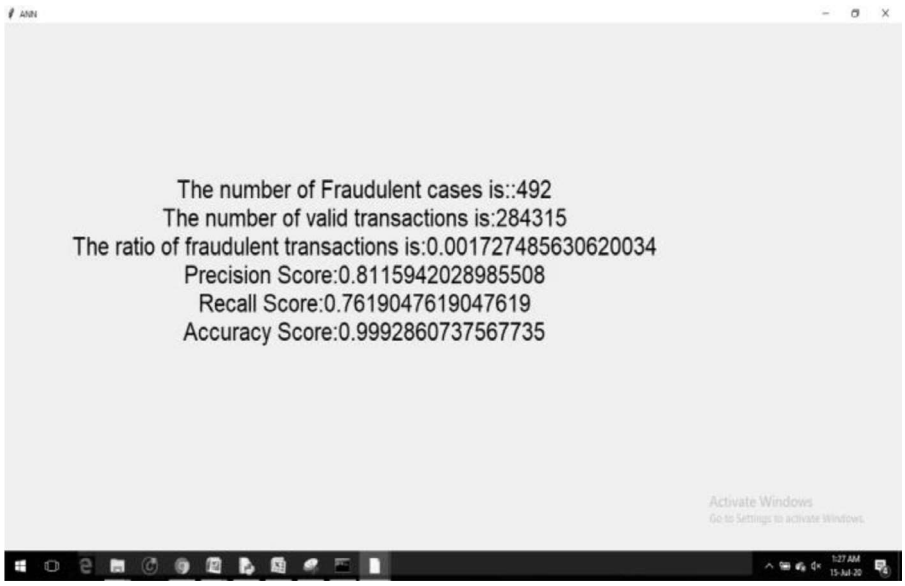

Fig. 5. shows the results using ANN.

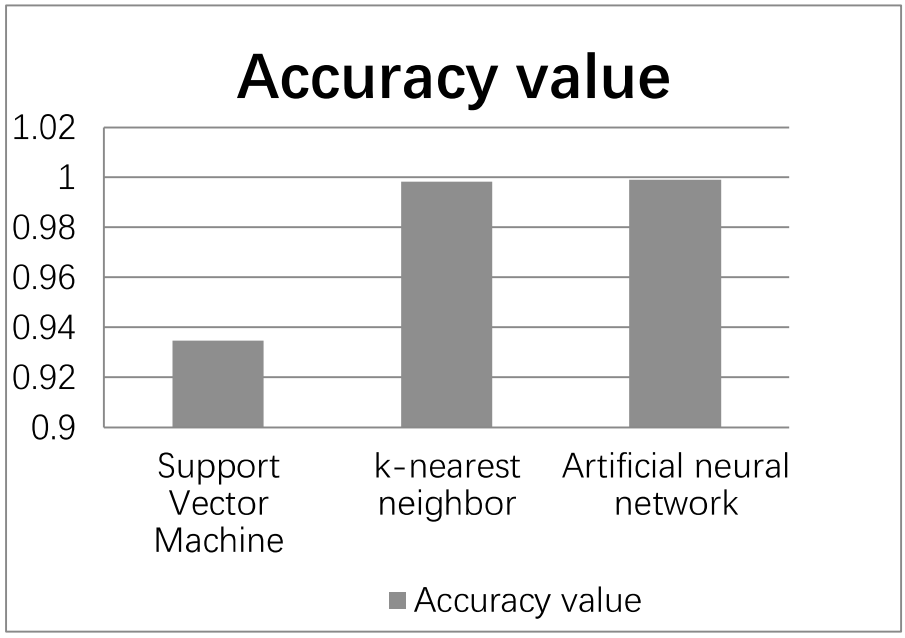

Fig. 6. represent the plot of accuracy obtained using SVM, KNN and ANN.

research work, data pre-processing, normalization and under-sampling carried out to overcome the problems faced by using an imbalanced dataset.

References

[1] A.A. Taha, S.J. Malebary, An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine, IEEE Access 8 (2020) 25579–25587, doi:10.1109/ACCESS.2020.2971354.

[2] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. Hacid, H. Zeineddine, An experimental study with imbalanced classification approaches for credit card fraud detection, IEEE Access 7 (2019) 93010–93022, doi:10.1109/ACCESS.2019.2927266.

[3] C. Jiang, J. Song, G. Liu, L. Zheng, W. Luan, Credit card fraud detection: a novel approach using aggregation strategy and feedback mechanism, IEEE Internet Things J. 5 (5) (Oct. 2018) 3637–3647, doi:10.1109/JIOT.2018.2816007.

[4] I. Sohony, R. Pratap, U. Nambiar, Ensemble learning for credit card fraud detection, in: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data Association for Computing Machinery, New York, NY, USA, 2018, pp. 289–294, doi:10.1145/3152494.3156815.

[5] P.H. Tran, K.P. Tran, T.T. Huong, C. Heuchenne, P. HienTran, T.M.H. Le, Real time data-driven approaches for credit card fraud detection, in: Proceedings of the 2018 International Conference on E-Business and Application. Association for Computing Machinery, New York, NY, USA, 2018, pp. 6–9, doi:10.1145/3194188.3194196.

[6] I. Sadgali, N. Sael, F. Benabbou, Fraud detection in credit card transaction using neural networks, in: Proceedings of the 4th International Conference on Smart City Applications (SCA '19). Association for Computing Machinery, New York, NY, USA, 2019, pp. 1–4, doi:10.1145/3368756.3369082. Article 95.

[7] D. Prusti, S.K. Rath, Web service based credit card fraud detection by applying machine learning techniques, in: Proceedings of the TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 492–497, doi:10.1109/TENCON.2019.8929372.

[8] M. Zamini, G. Montazer, Credit card fraud detection using autoencoders based clustering, in: Proceedings of the 9th International Symposium on Telecommunications (IST), Tehran, Iran, 2018, pp. 486–491, doi:10.1109/ISTEL.2018.8661129.

[9] S. Akila, U.S. Reddy, Credit card fraud detection using non-overlapped risk based bagging ensemble (NRBE), in: Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, 2017, pp. 1–4, doi:10.1109/ICCIC.2017.8524418.

[10] M.S. Kumar, V. Soundarya, S. Kavitha, E.S. Keerthika, E. Aswini, Credit card fraud detection using random forest algorithm, in: Proceedings of the 3rd International Conference on Computing and Communications Technologies (ICCCT), Chennai, India, 2019, pp. 149–153, doi:10.1109/ICCCT2.2019.8824930.

[11] Z. Li, G. Liu, S. Wang, S. Xuan, C. Jiang, Credit card fraud detection via kernel-based supervised hashing, in: Proceedings of the IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, 2018, pp. 1249–1254, doi:10.1109/SmartWorld.2018.00217.

[12] P. Kumar, F. Iqbal, Credit card fraud identification using machine learning approaches, in: Proceedings of the 1st International Conference on Innovations in Information and Communication Technology (ICIICT), CHENNAI, India, 2019, pp. 1–4, doi:10.1109/ICIICT1.2019.8741490.

[13] Y. Jain, N. Tiwari, S. Dubey, S. Jain, A comparative analysis of various credit card fraud detection techniques, Int. J. Recent Technol. Eng. 7 (2019) 402–407.