IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ANALYSIS ON THE APPLICATION OF DATA SCIENCE IN FRAUD DETECTION AND PREVENTION

Sugavasi Nomitha, Student, Computer Science and Technology, Narayanamma Institute of Technology and Science (For Women), Attapur, Hyderabad

Raaga Pravija Gaddam, Student, Computer Science and Technology, College - Narayanamma Institute of Technology and Science (For Women), Bachupally, Hyderabad

Abstract.

The spectacular surge in the proportion of credit card transactions, web based purchases, has led to a surge in fraudulent activities recently. For any business establishment, credit card security is a major concern. In this respect, credit card fraud is hard to identify. Thus it became imperative to implement effectual fraud detection systems for all credit card issuing banks to mitigate their losses. Betrayed transactions with real transactions in actuality are often dispersed and simple methods of matching are not enough to detect them accurately. The paper proposes an algorithm based on Machine Learning credit card fraud detection to solve the issue of a fraudulent transaction. This framework nominally increases the probability of card fraud by exponential activity. The results show that the accuracy of Random Forest, Support Vector Machine and KNN classifiers achieves respectively 94.84%, 89.46%. Random Forest could even predict new fraud cases very quickly. Keywords:Credit Card Fraud, Machine Learning algorithm, Fraud detection.

1.Introduction :

Using credit cards for a variety of purchases is quickly becoming a necessity. With this fantastic and quick method of performing transactions comes a large amount of risk. Based on rapid internet funding growth, digital transfers and rapid banking industry expansion, credit card card usages are becoming more common in daily living[1].A credit card can be made in two forms.When it is displayed physically, the first time you use this card for a charge, cancellation or transfer. If a card is not available, e.g. for online transactions or payments (some information is required, such as the CVV number, cardholder name, PIN, security query, etc.); Accurate, quick and effective methods for the detection of credit cards have become a hot issue in recent investigation. Currently, the Bayes

network algorithm is the most commonly used data mining algorithm for credit card fraud[2]. In credit card data leak prevention, there seems to be no genuine solution for ensuring the safety of the card if it is safely put in the pocket of the owner, or when it is used by a third party. The investigating organisation has been interested in thefrauds by credit card, and a number of approaches have been recommended to detect theft.

2.LiteratureSurvey :

For multiple technologies, ml algorithms have been used to secure credit card payments.Fraudulent credit card data set tested the impactof the naives and regression models Fraudulent credit card dataset was assessed for the effectiveness of naive bays, k-near neighbour and logistic regression. A large part of the progress that has been made with cyber retailers is with regard to marketing and other tactics that have the effect of attracting clients to a data center. In order to solve the data problem of information asymmetry and aim to evolutionize embedding capacity by two important algorithm techniques, a data capture system for credit cards using whal optimization and SMOTE (synthetic minority optimization technique) has been implemented. In many ways, the existing solution is unfeasible. Analysis of the transaction model is nothing short of an invasion of privacy and more about a system full of risks and misjudgments.

3. Proposed Approach:

This study has a number of steps to resolve the problem. Beginning with the data set to be used, the data set must be pre-processed.

The test and train data must be set before the application of the machine learning classifier. The classifier can be incorporated after configuration of the data. The last step was to determine the algorithms for machine learning. The steps are as described The following explains Figure.1 in more detail. In subchapters each of the steps is explained.

A.Dataset

The information comes from Kaggle, which gathers data on the history of prior transactions. The data is stable and the classifications are good. The dataset includes 30 transaction records features and 1 label. Class Label: 1- Fraud, 0- Standard.

B. Preprocessing

As discussed earlier in this thread, our dataset consists of 30 features and even the evaluation of those lowest effective dose in the dataset. This method searches key features and selects all 30 features for training the machine using Machine learning models.Then split the data into test and train.Train the machine with 80% training data .This training data contains feature and label.Remain 20% split-up was used to test the machine learning model.

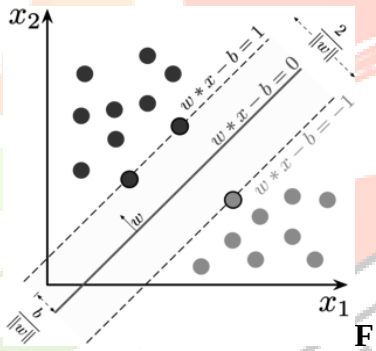
C.Model Selection

Machine Learning is known a part of Artificial Intelligence, involving the progress of approaches and practices for learning the computer. The analysis of the supervised learning algorithms in this study is performed. Here the supervised classifying algorithm is used to relate the model assessment [3]

I. Support Vector Machine

Support vector machines (SVM) are a series of supervised classification and regression learning methods. Support vectors are the nearest data points to the surface of decision (or hyperplane). Optimal hyperplane is derived from the lowest independent enhancements function class. Set hyperplanes H to the following in figure 2.

w•xi +b +1 for yi = +1 for w•xi +b = -1 for yi = -1The planes
H1 and H2 are:
H1: xi w• +b = +1W•xi +b =-1 H2:
The support tips are the points of planes H1 and H2.



F igure.2 support vector machine

The nearest neighbouring algorithms belong to the "simplest" machines supervised and have been well-studied over the previous decade in the area of information processing [4]. Although neural networks are gaining wide acceptance of computer vision and machine learning, the cross-section between computer vision, pattern classification and biometrics is one of the most frequently used neighbouring models. KNNis a supervised learning algorithm that simply stores labelled workouts that show the training phase [5]. Thatis why kNN is also known as a lazy algorithm in learning in figure 3.

k-Nearest Neighbor
Classify (**X**, **Y**, *x*) // **X**: training data, **Y**: class labels of **X**, *x*: unknown sample
for $i = 1$ **to** m **do**
  Compute distance $d(\mathbf{X}_i, x)$
end for
Compute set I containing indices for the k smallest distances $d(\mathbf{X}_i, x)$.
return majority label for $\{\mathbf{Y}_i$ where $i \in I\}$

Figure.3 KNN Classifier

4. Experiments And Result Analysis

In Python IDE, this work was carried out with python. We have chosen to divide our primary data based on the exact results of various data proportions: 80percent of total for learning and 20percent of the total for the test phase in figure 4-6.

This study will examine the exact comparison of three classification algorithms. Upon completion of the classification, results of a comparison of precision show that Random Forest algorithm achieved high accuracy 94%. Besides, the results view will be displayed.In this study, to analyze the estimation accuracy,confusion matrix and ROC curve was used.

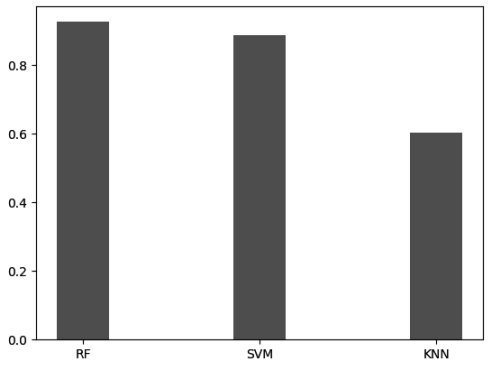


Figure.4 Accuracy Comparison

Algorithm	Accuracy	Precision	Recall	F-1 Score
KNN	**0.60**	**0.61**	**0.61**	**0.60**
SVM	**0.89**	**0.89**	**0.90**	**0.89**
Random Forest	**0.94**	0.93	0.93	0.93

Figure.5 ROC Curve

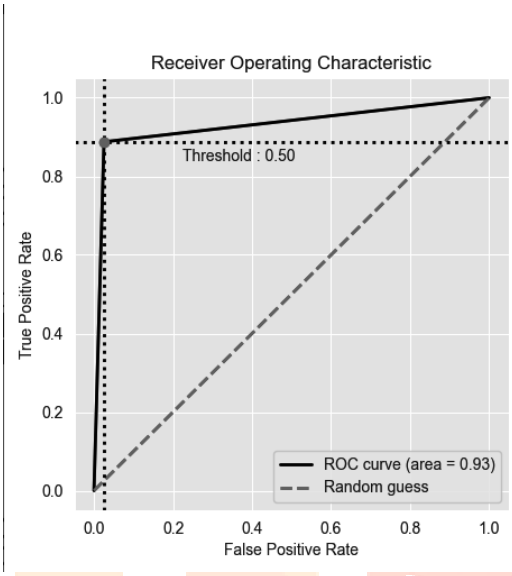


Figure.6 Model Evaluation

5. Conclusion

Fraud detection by credit card is a serious problem. Companies are therefore making significant investmentsin new algorithms to help identify and avoid suspicious charges. This article is intended to pinpoint business transactions financial fraud using machine algorithms. The results have shown that Random Forest is more effective than SVM and KNN using various methods such as the matrix of accuracy, recall, reliability, true positive rate and false positive rates.Our future research involves the construction of this model as a prototype. The prototype includes a highlight on how the exit solution is restricted and how this e-business framework has enough card holder authentication

References

[1] A. O. Adewumi and A. A. Akinyelu, A survey of machine-learning and nature-inspired based credit card fraud detection techniques, Int. J. Syst. Assur. Eng. Manag., vol. **8**, pp. 937–953, Nov. 2017

[2] B. Lebichot, Y.-A. Le Borgne, L. He-Guelton, F. Oblé, and G. Bontempi, Deep-learning domain adaptation techniques for credit cards fraud detection, in *Proc. INNS Big Data Deep Learn. Conference*, Genoa, Italy, 2019, pp. 78–88.

[3] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, Credit card fraud detection using machine learning techniques: A comparative analysis, Proc. IEEE Int. Conf. Comput. Netw. Informatics, ICCNI 2017, vol. 2017-Janua, pp. 1–9, 2017.

[4] Haldorai, A. Ramu, and S. Murugan, Social Aware Cognitive Radio Networks, Social Network Analytics for Contemporary Business Organizations, pp. 188–202. doi:10.4018/978-1-5225- 5097-6.ch010

[5] R. Arulmurugan and H. Anandakumar, Region-based seed point cell segmentation and detection for biomedical image analysis, International Journal of Biomedical Engineering and Technology, vol. **27**, no. 4, p. 273, 2018.

[6] Argyriou, E.N.,Symvonis, A. 2012, Detecting periodicity in serial data through visualization. Advances in Visual Computing, vol. 7432, pp. 295-304.

[7] Bresfelean, Vasile Paul, Mihaela Bresfelean, Nicolae Ghisoiu, and Calin-Adrian Comes. 2007. "Data Mining Clustering Techniques in Academia." In ICEIS (2), pp. 407-410.

[8] Bresfelean, V. P., Bresfelean, M., Ghisoiu, N., & Comes, C. A. 2008. Determining students' academic failure profile founded on data mining methods. In Information Technology Interfaces, IEEE, pp. 317-322

[9] Cofan S.M., Ivan, L., Dogaru V., Cios A., Savin M. 2014. Analiza Informațiilor. Manual, ed. Ministerului Afacerilor Interne, ISBN 978- 973-745-129-3.

[10] Burge, P., Shawe-Taylor, J. 2001, An Unsupervised Neural, Network Approach to Profiling the Behaviour of Mobile Phone, Users for Use in Fraud Detection. Journal of Parallel and Distributed Computing 61: 915–925

[11] Cox, K., Eick, S., Wills, G. 1997. Visual Data Mining: Recognizing Telephone Calling Fraud. Data Mining and Knowledge Discovery 1: 225–231.

[12] Spann , D. D. 2013, Fraud Analytics: Strategies and Methods for Detection and Prevention, ISBN-13: 978-1118230688.

[13] Westphal, C., 2009, Data Mining For Intelligence, Fraud, & Criminal Detection Advanced. Analytics & Information Sharing Technologies.

[14]CRC Press Taylor & Francis Group. 978-1-4200-6723-1.

[15] Young, M.R., 2014, Financial Fraud Prevention and Detection. Governance and Effective Practices, John Wiley & Sons, Inc., Hoboken, New Jersey, ISBN 9781118617632.