Credit Card Fraud Detection Using AI (Python)

Arshee Naz[1], Praveen Kumar[2], Dr. Ashad Ullah Qureshi[3]

Junior Research Fellow,National Institute of Technology, Kurukshetra, Haryana[1]

Website Administrator, Shri Vishwakarma Skill University, Dudhola[2]

Technical Officer, Indian Institute of Information Technology, Sonepat, Haryana[3]

arshee_jrf@nitkkr.ac.in[1], praveen.kumar@svsu.ac.in[2], aqureshi@iiitsonepat.ac.in[3]

Abstract: *Credit card fraud poses a formidable obstacle in the financial sector, resulting in considerable monetary damages for both individuals and institutions. The necessity for efficient fraud detection systems has become paramount due to the rising number of online transactions and the advancement of fraudulent methods. Machine learning methods have demonstrated potential in tackling this issue by utilizing past transaction data to detect fraudulent behavior. This study provides a thorough examination and evaluation of different machine learning techniques used in the detection of credit card fraud.*

The aim of this study is to assess and compare the efficacy of several machine learning algorithms in identifying fraudulent credit card transactions. The dataset utilized for experimentation is acquired from a prominent financial institution, including of both authentic and deceitful transactions. The dataset has been preprocessed to address missing values, outliers, and feature scaling.

A variety of machine learning algorithms, such as "logistic regression, decision trees, random forests, support vector machines (SVM), and artificial neural networks (ANN)," are utilized and trained on the preprocessed information. The evaluation of each method is conducted using criteria like as "accuracy, precision, recall, and F1-score. In addition, several evaluation methods, such as k-fold cross-validation", are used to assure the reliability of the findings.

The empirical findings suggest that machine learning algorithms has the capability to accurately identify fraudulent credit card transactions. The algorithms demonstrate varying performance across different parameters, with certain algorithms displaying higher accuracy but worse precision or recall. The "Support Vector Machine (SVM)" algorithm gets the maximum accuracy rate of 98%, while the "Artificial Neural Network (ANN)" model displays the optimal balance between precision and recall..

Keywords: Credit card fraud

REFERENCES

[1]. Abdulhamid, O., Osho, O., & Shuaib, M. (2018). Evaluation of classification algorithms for phishing url detection. I-Manager's Journal on Computer Science, 6(3), 34 Retrieved from http://search.ebscohost.com/login.aspx?direct=true&AuthType=shib&db=edb&AN= 135838191&site=eds-live

[2]. Akila, R., & Bhuvaneswari, M. (2018). Credit card fraud recognition using data mining techniques. International Journal of Advanced Research in Computer Science; Udaipur, 9(Special Issue 1), 86–87.

[3]. http://dx.doi.org.contentproxy.phoenix.edu/10.26483/ijarcs.v9i0.5618 Albrecht, W. S., Albrecht, C. O., Albrecht, C. C., & Zimbelman, M. F. (2016). Fraud Examination, 5e. Boston, MA: Cengage Learning.

[4]. Azim, M. A., & Bhuiyan, M. H. (2018). Text to emotion extraction using supervised machine learning techniques. Telkomnika, 16(3), 1394-1401. doi:http://franklin.captechu.edu:2123/10.12928/TELKOMNIKA.v16i3.8387

[5]. Badal-Valero, E., Alvarez-Jareño, J. A., & Pavía, J. M. (2018). Combining Benford's Law and machine learning to detect money laundering. An actual Spanish court case. Forensic Science International, 282, 24–34. https://doi.org/10.1016/j.forsciint.2017.11.008

[6]. Banerjee, A.V., Chassang, S., Snowberg, E. (2017). Decision theoretic approaches to experiment design and external validity. Handbook of Economic Field Experiments 1, p 141-174 https://doi.org/10.1016/bs.hefe.2016.08.005

[7]. Banerjee, P., Dehnbostel, F. O., & Preissner, R. (2018). Prediction is a balancing act: 109 importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. Frontiers in Chemistry, 6. Retrieved from https://doaj.org

[8]. Bang, S., & Kalavadekar, P. N. (2014). Determining K-most demanding products using data mining technique. International Journal of Computer Science and Network Security (IJCSNS), 14(6), 18. Retrieved from https://franklin.captechu.edu:2074/docview/1786151177?accountid=44888

[9]. Bangira, T., Alfieri, S. M., Menenti, M., & Adriaan, v. N. (2019). Comparing thresholding with machine learning classifiers for mapping complex water. Remote Sensing, 11(11) doi:http://franklin.captechu.edu:2123/10.3390/rs11111351

[10]. Boeren, E. (2018). The methodological underdog: A review of quantitative research in the key adult education journals. Adult Education Quarterly, 68(1), 63–79. https://doi.org/10.1177/0741713617739347

[11]. Boughorbel, S., Jarray, F., & El-Anbari, M. (2017). Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS ONE, 12(6), e0177678.

[12]. Canovas-Garcia, F., & Alonso-Sarria, F. (2015). Optimal combination of classification algorithm and feature ranking methods for object-based classification of submeter resolution Z/I-Imaging DMC imagery. Remote Sensing, (4), 4651. https://doi.org/10.3390/rs704404651