**Credit Card Fraud Detection: A Comparative Study of Machine Learning Algorithms on Imbalanced Data**

**ABSTRACT:**

Credit card fraud refers to the physical loss of credit card or loss of sensitive credit card information. Many machine learning algorithms can be used for detection. This research shows several algorithms that can be used for classifying transactions as fraud or genuine one. Credit Card Fraud Detection dataset was used in the research. Because the dataset was highly imbalanced, SMOTE technique was used for oversampling. Further, feature selection was performed and dataset was split into two parts, training data and test data. The algorithms used in the experiment were Logistic Regression, Random Forest, Naive Bayes and Multilayer Perception. Results show that each algorithm can be used for credit card fraud detection with high accuracy. Proposed model can be used for detection of other irregularities.

.

|  |  |
| --- | --- |
| **EXISTING SYSTEM** | **PROPOSED SYSTEM** |
| **EXISTING CONCEPT:-**  Check Fraud occurs when person forges a check or pays for something with check knowing that there is not enough money. Internet sales are fraud where fraudster sale fake items or counterfeit items, or taking payment without delivering the item. There are a couple more, such as charities fraud, identity theft, credit card fraud, debt elimination, Insurance fraud and others. Due to increasing popularity of cashless transactions, one of the most common frauds are credit card frauds. Credit card fraud refers to the situation where fraudster uses credit card for their needs while owner of that credit card is not aware of that. | **PROPOSED CONCEPT:** -  There are two types of credit card frauds. One is theft of physical card, and other one is stealing sensitive information from the card, such as card number, cvv code, type of card and other. By stealing credit card information, a fraudster can broach a large amount of money or make a large amount of purchase before cardholder finds out. Because of that, companies use various machine learning methods to recognize which transactions are fraudulent and which are not. |
| **EXISTING TECHNIQUE:-**   * **SVM** | **PROPOSED TECHNIQUE:**-   * **Multilayer Perception** |
| **TECHNIQUE DEFINITION:-**   * It is a classification method. In this algorithm, we plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. * If we only had two features like Height and Hair length of an individual, we’d first plot these two variables in two dimensional space where each point has two co-ordinates | **TECHNIQUE DEFINITION:-**  The purpose of this paper is to analyze various machinelearning algorithms, such as Logistic Regression (LR), Random Forest (RF), Naïve Bayes (NB) and Multilayer Perceptron (MLP) in order to determine which algorithm is most suitable for credit card fraud detection. |
| **DRAWBACKS:-**   * Less accurate | **ADVANTAGES:-**   * More Accurate |

**MINIMUMSYSTEM REQUIREMENTS**

**HARDWARE REQUIREMENTS**

* PROCESSOR : DUAL CORE 2 DUO.
* RAM : 2GB DD RAM
* HARD DISK : 250 GB

**SOFTWARE REQUIREMENTS**

* FRONT END : PYTHON
* OPERATING SYSTEM : WINDOWS 7
* IDE : Spyder3

**Conclusion:**

Credit card frauds represent a very serious business problem. These frauds can lead to huge losses, both business and personal. Because of that, companies invest more and more money in developing new ideas and ways that will help to detect and prevent frauds. The main goal of this paper was to compare certain machine learning algorithms for detection of fraudulent transactions. Hence, comparison was made and it was established that Random Forest algorithm gives the best results i.e. best classifies whether transactions are fraud or not. This was established using different metrics, such as recall, accuracy and precision. For this kind of problem, it is important to have recall with high value. Feature selection and balancing of the dataset have shown to be extremely important in achieving significant results.

**References:**

[1] Global Facts (2019). Topic: Startups worldwide. [online] Available at: https://www.statista.com/topics/4733/startups-worldwide/ [Accessed 10 Jan. 2019].

[2] Legal Dictionary (2019). Fraud - Definition, Meaning, Types, Examples of fraudulent activity. [online] Available at: https://legaldictionary.net/fraud/ [Accessed 15 Jan. 2019].

[3] European Central Bank (2018). Fifth report on card fraud, September 2018. [online]. Available at: <https://www.ecb.europa.eu/pub/cardfraud/html/ecb.cardfraudreport2018>09.en.html#toc1 [Accessed 21 Jan. 2019].

[4] En.wikipedia.org. (2019). Credit card fraud. [online] Available at: https://en.wikipedia.org/wiki/Credit\_card\_fraud [Accessed 24 Jan. 2019].

[5] A. Mishra, C. Ghorpade, “Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques” 2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) pp. 1-5. IEEE.

[6] S. V. S. S. Lakshmi, S. D. Kavilla “Machine Learning For Credit Card Fraud Detection System”, unpublished

[7] N. Malini, Dr. M. Pushpa, “Analysis on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection“, Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), 2017 Third International Conference on pp. 255- 258. IEEE.

[8] Mrs. C. Navamani, M. Phil, S. Krishnan, “Credit Card Nearest Neighbor Based Outlier Detection Techniques”

[9] J. O. Awoyemi, A. O. Adentumbi, S. A. Oluwadare, “Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis”, Computing Networking and Informatics (ICCNI), 2017 International Conference on pp. 1-9. IEEE.

[10] Z. Kazemi, H. Zarrabi, “Using deep networks for fraud detection in the credit card transactions”, Knowledge-Based Engineering and Innovation (KBEI), 2017 IEEE 4th International Conference on pp. 630-633. IEEE.