Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.777235

PREDICTIVE ANALYTICS WITH MACHINE LEARNING FOR FRAUD DETECTION OF ONLINE MARKETING TRANSACTIONS

Tatarao Vana[1*], S Mallikarjun[2], P Pavan Kumar[3], V Rahul[4]

[*1]Assistant Professor, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh

.[*2,3.4] Students, Department of CSE, Raghu Engineering College, Visakhapatnam, Andhra Pradesh.

tatarao.vana@raghuenggcollege.in,
19981a05f2@raghuenggcollege.in, 19981a05d5@raghuenggcollege.in,
19981a05d9@raghuenggcollege.in

ABSTRACT:

Predictive analytics with machine learning has become an essential tool for detecting fraudulent transactions in online marketing. Fraudulent activities in online marketing can cause significant financial losses to businesses and harm their reputation. Therefore, detecting fraudulent transactions has become a critical challenge for online marketing platforms. This paper presents an approach to fraud detection using machine learning techniques. The proposed approach uses supervised learning algorithms ie: Random Forest, KNN, and XGBoost to classify whether the transaction is fraudulent or not. The datasets used here is obtained from an online marketing platform and consists of various features such as the IP address of the user, time of the transaction, and device type used for the transaction.

The study also shows that the features that have the most significant impact on the detection of fraudulent transactions are the time of the transaction and the device type used. The time of the transaction is more crucial in detecting fraudulent transactions than the device type used. The study suggests that combining these features with other relevant features could further improve the performance.

Keywords: Machine learning, Online Marketing System, Deep Learning, Random Forest, supervised learning.

1. INTRODUCTION

In order to identify fraudulent transactions in online marketing, predictive analytics with machine learning has gained popularity. Online marketing fraud can be a serious issue for businesses, as it can result in significant revenue losses and reputational harm. Predictive analytics and machine learning techniques can aid in quickly and accurately identifying fraudulent activities thanks to the enormous amount of data that is available in online marketing transactions. First off, because fraudsters' behavior is dynamic, it can be difficult to spot it in online marketing transactions. Traditional rule-based approaches struggle to keep up with fraudsters' constant adaptation of their strategies for avoiding detection. Additionally, it is impossible to manually analyze the massive amount of transactions and data.

Predictive analytics can help in this situation. A statistical technique called predictive analytics makes predictions about the future based on historical data. Predictive analytics can be used to examine trends in historical data to spot potentially fraudulent transactions in the context of fraud detection.

Predictive models for fraud detection can be created with the aid of machine learning algorithms. Machine learning algorithms are created to recognize relationships and patterns in data and use them to predict outcomes for new data. These algorithms can be taught how fraudulent transactions differ from legitimate ones using historical data.

2. OBJECTIVE

To describe various cutting-edge technologies that are used to combat threats and frauds that occur when transactions are made online. The inclusion of feature engineering and its feature selection helps to demonstrate the superior performance of advanced technologies.

3. LITERATURE REVIEW

[1] Credit card fraud detection of methods like unsupervised learning models like neural networks and classification, from (IJERT) vol. 09, no. 04 (April 2020),L. Bhavja, V. Sasidhar Reddy, S. Karishma, 2020.

Nowadays, the number of online transactions has increased tremendously. Among them, online marketing transactions made up a large part. Therefore, transaction fraud detection applications are highly desirable in banking and financial operations. The purpose of online marketingtransaction fraud may be to obtain unauthorized funds from an account.

[2] Detecting Forgedsellers in online transactions Using a Support Vector Machine Approach. From International Journal of Engineering Trends and Technologies, Ruņgis, Shini. The money that spent on retail through e-commerce has continuously increased over time, demonstrating a clear change in consumer priorities away from concrete establishments and towards retail. One of the main drivers of this expansion in recent years has been online markets. There includes a detailed examination of fraudulent e-commerce buyers and their transactions, as well as a discussion of potential control and prevention measures. Trade fraud, which occurs on the seller's side of the market, is another type of fraud. A straightforward example of this form of fraud is a product or service that is advertised and sold for a low price but never delivered. In this research, a strategy to identify these dishonest merchants using machine learning approaches is proposed.

[3] (2019). Fraud detection using machine learning in e-commerce,Saputra, Adi and Suharjito, Suharjito. The amount of e-commerce transactions has increased along from the growth of Internet users. Moreover, the number of fraudulent internet transactions has increased. The goal is to assess appropriate machine learning techniques. The methods employed include decision trees, naive bays, random forests, and neural networks. Fraud protection in e-commerce should be built using machine learning. The accuracy ratings from the some of the ML algorithms utilising uncertainty matrices are 91%, 95%, 96%, and 96% respectively. The F1 mean may be increased from 67.9% to 94.5% and the G-Mean from 73.5% to 84.6% using the Synthetic Minority Intercept Method (SMOTE).

[4] "Credit card fraud detection using an unsupervised machine learning scheme",Kazu R.K. Dwivedi,A.K. Rai. International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India 2020. Credit cards are using for transaction of online as well as in offline. Because in the developments in communications and e-commerce. Thus, it is anticipated that system security will significantly reduce the likelihood of fraudulent transactions. In order to precisely and successfully identify these scams, several procedures are always required. The suggested approach performs efficiently than the K-Means clustering approach, isolation forest, local outer factor, and autoencoder (AE), among other approaches. The accuracy of the NN-based fraud detection approach is 99.87%, while the AE, IF, LOF, and K Means of the existing methods are, respectively, 97%, 98%, 98%, and 99.75%.

4. EXISTING SYSTEM

The current system uses models based on K-Means clustering and Logistic Regression to estimate fraudulent and non-fraudulent transactions. Due to its lengthy computational process, this method has poor recall and precision scores and also lacks robustness.

5. PROPOSED SYSTEM

The proposed system for predictive analytics with machine learning for fraud detection of online marketing transactions is designed to identify fraudulent activities in online marketing transactions. The system uses machine learning algorithms to categorize transactions as legitimate or fraudulent by observing various features such as the user's IP address, time of the transaction, and device type used. The system aims to provide accurate and efficient fraud detection capabilities to online marketing platforms.

The system consists of three main components: data collection, feature selection, and model training. In the data collection stage, the system collects transactional data from the online marketing platform. The data is then processed and cleaned to remove irrelevant or missing data. In the feature selection stage, the system selects the most relevant features that have the most significant impact on the detection of fraudulent transactions. Finally, in the model training stage, the system trains a machine learning model using the selected features and evaluates the model's performance using various evaluation metrics.

6. METHODOLOGY

1. Random Forest:

Popular machine learning algorithms like Random Forest are used for tasks like classification, regression, and others. This algorithm builds decision trees and based on their predictions a single prediction will be produced. The algorithm generates a group of decision trees, each of them are constructed using a random subset of the data and features, thus earning the name "Random Forest" for the system.

2. KNN Algorithm.

Machine learning algorithms for classification and regression tasks include KNN, also known as K-Nearest Neighbors. It does not rely on any prior knowledge of the distribution of the

underlying data and does not call for any training stage. Instead, it gathers all accessible data instances and stores them, then uses them to forecast the behavior of fresh data.

3. Decision Tree:

A decision tree is defined as a tree-like model that represents decisions and their possible situations, including chance events, resource costs, and utility. It is a type of supervised learning algorithm used in machine learning, data mining, and statistics.

4.XGBoost:

XGBoost is an open-source machine learning algorithm that is used for classification and regression tasks. It is designed to improve performance and speed over other gradient boosting algorithms. XGBoost works by iteratively adding decision trees to an ensemble, where each tree corrects the errors of the previous trees. The algorithm focuses on reducing both bias and variance in the model by combining weak learners into a strong learner.

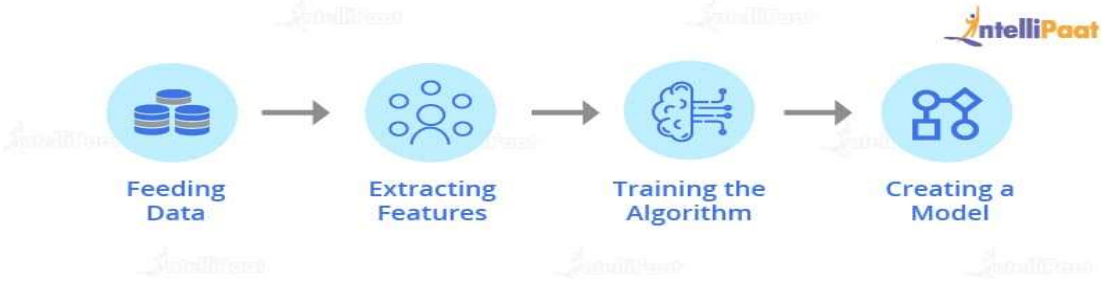

Fig 1: Data Flow

Predictive analytics with machine learning can be very effective for fraud detection in online marketing transactions. Here's a general flow for implementing such a system.
Feeding Data: The model is first fed with data. The quantity of data used to train the model affects how accurate it is; the model performs better with more data. You need to feed your model increasing volumes of data in order to identify scams unique to a given industry. This will train your model so that it can accurately identify fraud actions unique to your company.

Extracting Features: In machine learning, extracting features means selecting and transforming raw data into features which can be further used for model training and prediction. The goal of feature extraction is to identify relevant information from the input data and represent it in a way that can be easily understood by machine learning algorithms.

Training the Algorithm: In machine learning, a training algorithm is a method used to train a machine learning model to observe patterns and make predictions based on the given data. The training algorithm is responsible for optimizing the model's parameters and weights based on the training data, which is a set of labelled examples used to teach the model how to make predictions.

Creating a Model: creating a model refers to the process of building a mathematical representation that can learn from and make predictions on input data. A model is essentially a set of rules, parameters, and mathematical equations that define the relationships between data.

The purpose of creating a model is to find the best possible representation of the underlying patterns in the data that can be used to make accurate predictions on new.

7. RESULT & ANALYSIS

This is the home page shown on the screen. The user can load the dataset they wish to work with in next step.

Fig 2. Home page

Choose a model: The user can use the model to accurately apply to the dataset.

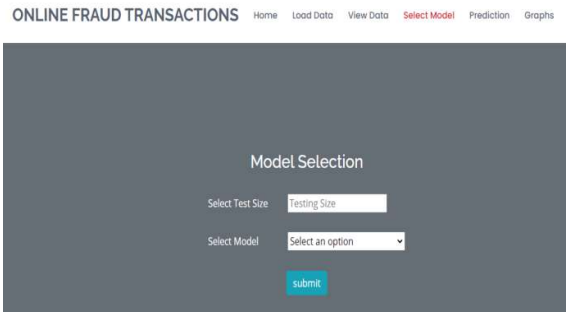

Fig. 3. Model Selection

Predictions: The user may insert random values.

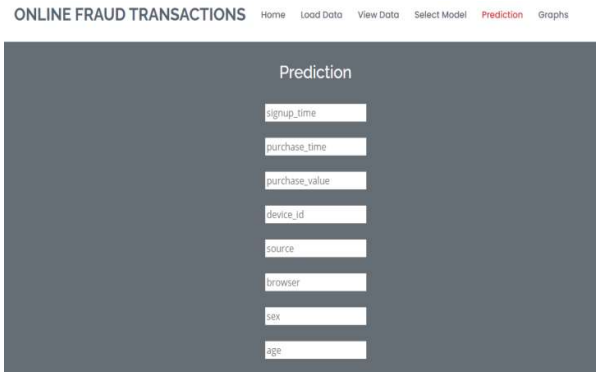

Fig. 4. Model Predection.

Graphs: By using the graphs, a user can assess the model's performance.

Fig. 5. Model Performance

8. CONCLUSION

To sum up, predictive analytics combined with machine learning is a potent method for spotting fraud in online marketing transactions. It is possible to spot patterns in the data that are suggestive of fraudulent activity by utilising machine learning algorithms, and then use that knowledge to stop additional fraudulent transactions. In addition to careful feature engineering and model selection, implementing such a system necessitates careful data collection, cleaning, and preprocessing. To make sure the model remains effective over time, it is crucial to continuously monitor and update it. Online marketing companies can defend themselves and their clients from fraudulent activity by putting in place a well-designed predictive analytics system, thereby enhancing the overall trust and security of the online market.

9. REFERENCES

[1] “Intelligent Approach to Credit Card Fraud Detection Using an OLightGBM S. J. Malebary,A. A. Taha.”, IEEE Access.

[2] “A Multiple Classifiers System for Anomaly Detection in Credit Card Data With Unbalanced and Overlapped Classes”,K. H NG,S. N. Kalid, K.C Khore,G. K.Tong., IEEE Access (2020)

[3] “An Experimental Study classification approaches of online transactions”, it is aspecial section on Advanced Software And Data Engineering For Secure Societies, Y. Taher, R. Haque, M. S Hacid, H. Zeineddine, IEEE Access (2019)

[4] “Credit Card Fraud Detection Using Adaboost And Majority Voting”,, K. Randhawa, C. K.Loo, C. P.Lim, A. K.Nandi, Ieee Access,

[5] “Credit Card Fraud Detection: A Realistic Modeling and a Novel Learning Strategy”,AndreaDalPozzolo, Giacomo Boracchi,CesareAlippi, Gianluca Bontempi Ieee, Accessed on (2018)

[6] "DataDriven Anomaly Recognition for Unsupervised Model-Free Fault Detection in Artificial Pancreas", L. Meneghetti, M. Terzi, S. DelFavero, G. A.Susto, C. CobelliIeee Transactions On Control Systems Technology, Accessed on (2018)