RE-2022-212111 - Turnitin Plagiarism Report

by L Narayana Reddy

Submission date: 05-Mar-2024 10:05AM (UTC+0200)
Submission ID: 271709645921
File name: RE-2022-212111.pdf (261.96K)
Word count: 2666
Character count: 16968

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

N.UshaSree, M.Tech
Assistant Professor
Computer Science and Engineering
Srinivasa Ramanujan Institute Of Technology
Anantapur, India

L Narayana Reddy
Computer Science and Engineering
Srinivasa Ramanujan Institute Of Technology
Anantapur, India
204g1a0565@srit.ac.in

K.Pushpa
Computer Science and Engineering
Srinivasa Ramanujan Institute Of Technology
Anantapur, India
204g1a0574@srit.ac.in

G. Uday Kiran
Computer Science and Engineering
Srinivasa Ramanujan Institute Of Technology
Anantapur, India
214g5a05b4@srit.ac.in

N.Sai Harsha Vardhan
Computer Science and Engineering
Srinivasa Ramanujan Institute Of Technology
Anantapur, India
204g1a0585@srit.ac.in

***Abstract*—The surge in online transactions exposes digital platforms to escalating fraud risks, necessitating advanced security measures. This research explores the integration of machine learning (ML), deep learning, and predictive analytics to fortify online marketing systems against fraudulent activities. The objective is to proactively identify and prevent data breaches, minimising losses to customers. The study compares advanced technologies, emphasising feature engineering and parameter selection for optimal performance. Current methodologies, relying on Logistic Regression and KMeans clustering, exhibit drawbacks such as low precision, poor recall scores, and high computational time. The proposed system leverages Decision Trees, KNN Classifier, and Random Forest algorithms for enhanced fraud detection, providing accurate predictions, adaptive learning, and efficient processing. This research contributes to securing online transactions in ecommerce and banking, fostering user trust in digital transactions.**

Keywords—Predictive Analytics, Machine Learning, Fraud Detection, Online Marketing, Decision Trees, KNN Classifier, Random Forest.

1. INTRODUCTION

In the dynamic landscape of online transactions, the unprecedented growth in digital interactions has ushered in a new era of convenience and accessibility. However, this rapid evolution has also presented a significant challenge — the increasing susceptibility of online marketing systems to fraudulent activities. The ramifications of fraud extend beyond financial losses, encompassing data breaches and compromised personal information, resulting in a loss of trust among consumers.

As traditional security protocols prove inadequate in countering sophisticated cyber threats, the imperative to adopt advanced technologies becomes paramount. This paper addresses this critical need by investigating the integration of machine learning (ML), deep learning, and predictive analytics into the fabric of online marketing systems. These technologies stand poised as efficient guardians against the rising tide of fraud, offering the capability to handle vast amounts of sensitive data and predict intricate patterns indicative of fraudulent behaviour.

This study embarks on a comprehensive exploration, comparing and evaluating the effectiveness of diverse advanced technologies. The focus extends to feature engineering and parameter selection to optimise the performance of these technologies. Recognising the limitations of existing methodologies, particularly Logistic Regression and K-Means clustering, the research advocates for a paradigm shift towards Decision Trees, KNN Classifier, and Random Forest algorithms and parameter selection for optimal performance in securing online marketing systems.

2. LITERATURE SURVEY

[1] Credit card fraud detection employs classification, unsupervised, and neural network models to identify fraudulent transactions accurately, enhancing security and minimizing financial losses.

[2] Authors: L. Bhavya, V. Sasidhar Reddy, U. Anjali Mohan, S. Karishma

This study acknowledges the burgeoning volume of online credit card transactions and the subsequent need for robust fraud detection in banking and financial sectors. The authors explore classification, unsupervised, and neural network models to detect and prevent credit card fraud, highlighting the common motives behind such activities and the financial losses incurred by cardholders.

[2] Renjith and Shini propose a Support Vector Machine approach for detecting fraudulent sellers in online marketplaces, offering a robust method to safeguard against fraudulent activities and ensure trustworthiness in e-commerce transactions.

In response to the growing influence of e-commerce in global retail spending, this research targets fraudulent activities on online marketplaces, particularly focusing on fraudulent sellers engaging in merchant fraud. The study proposes a framework utilising Support Vector Machine techniques to identify and control such fraudulent sellers, shedding light on the nuances of ecommerce fraud.

[3] Fraud Detection using Machine Learning in eCommerce

Authors: Saputra, Adi & Suharjito, Suharjito

As online transactions in e-commerce surge, this work delves into the escalating incidents of fraud. The authors employ machine learning algorithms such as Decision Tree, Naive Bayes, Random Forest, and Neural Network to analyze fraud patterns. The study emphasizes the need for fraud prevention in e-commerce and presents the effectiveness of these algorithms in achieving high accuracy through evaluation metrics.

[4] In their research, A. K. Rai and R. K. Dwivedi address the pressing issue of fraud detection in credit card transactions, given the widespread use of credit cards in both online and traditional purchases. Their paper introduces an unsupervised machine learning scheme designed specifically for identifying fraudulent activities within credit card data. By comparing their proposed scheme with established approaches such as Auto Encoder, Local Outlier Factor, Isolation Forest, and K-Means clustering, the authors showcase its efficacy in achieving superior accuracy in fraud detection. This comparative analysis underscores the potential of their scheme to enhance security measures and mitigate financial risks associated with fraudulent credit card transactions, contributing valuable insights to the field of fraud detection in financial systems.

3. Related Works:

In the realm of banking and computer science, credit card fraud detection stands out as a critical area of focus, prompting extensive research efforts and significant investment of resources. Researchers have dedicated considerable time and effort to developing more effective prediction systems, leveraging machine learning techniques to combat online fraud. Dal Pozzolo Andrea's thesis, titled "Adaptive Machine Learning for Credit Card Fraud Detection," provides a comprehensive overview of the workings of credit card fraud detection systems. The thesis delves into the intricacies of prediction and classification using machine learning methodologies in initial sections. Subsequently, it advocates for the utilization of various sampling techniques to enhance system performance. Furthermore, the thesis delves into the analysis and evaluation of the performance of diverse machine learning models. Finally, it culminates with the implementation of a Fraud Detection System (FDS) in a real-time setting using web API, underscoring the practical application and significance of the research in mitigating financial risks associated with credit card fraud.

3.1 Credit card fraud detection

1. [3–9] Several papers in the field address the implementation and efficacy of various machine learning (ML) techniques for fraud detection. Notably, [5], [6], and [8] demonstrate the utilization of simple and rudimentary ML methods for binary classification tasks. Meanwhile, [9] conducts a comparative analysis between different ML algorithms and neural networks. However, a common challenge across these studies is the scarcity and sensitivity of available datasets, as banking information is often shielded and safeguarded. This limitation impedes the derivation of meaningful insights from the data and necessitates substantial processing capacity and time investment for model development. Moreover, [4] highlights a specific issue with the random forest approach, namely, the propensity for overfitting, which undermines the accuracy and reliability of fraud detection models. These findings underscore the complexity and ongoing challenges inherent in developing robust fraud detection systems within the banking sector.

3.2 Challenges and Motivation

1. Handling enormous daily data requires fast and suitable models for detecting fraudulent transactions effectively.
2. Imbalanced datasets, with low instances of fraudulent transactions, pose challenges in accurate fraud identification.
3. The confidentiality of banking data makes it difficult to obtain for model training and analysis.
4. Misclassified data due to unreported fraudulent transactions can lead to inaccuracies in model predictions.
5. Hackers employ adaptive techniques to circumvent fraud detection models, necessitating ongoing updates and improvements in security measures.

4. Methodology

4.1 Data Collection:

For this study, a comprehensive dataset of online transactions, encompassing a diverse range of sources, was collected. The dataset includes information on transaction amounts, timestamps, user details, and other relevant features. Special

attention was given to ensuring the inclusion of both legitimate and fraudulent transactions to facilitate a robust analysis.

4.2 Data Preprocessing:

The collected dataset underwent comprehensive preprocessing, which included handling missing data, removing duplicates, and normalizing numerical features. Categorical variables were encoded, and outliers were addressed to ensure dataset quality and reliability for machine learning analysis.

4.3 Feature Engineering:

Feature engineering played a pivotal role in shaping the dataset for effective model training. Relevant features were selected based on their potential to contribute to fraud detection accuracy. New features, such as transaction frequency and user behavior patterns, were engineered to capture intricate aspects of online transaction dynamics.

4.4 Model Selection:

In the realm of fraud detection for online marketing transactions, choosing the right machine learning model is crucial for achieving accurate and reliable results. The following models were considered and evaluated for their suitability.

4.4.1 Decision Trees:

Overview: Decision Trees are powerful tools for classification tasks, capable of representing complex decision-making processes. Each node in the tree represents a decision based on a feature, leading to a final outcome.

Applicability: Decision Trees are effective in capturing non-linear relationships within data, making them suitable for scenarios where fraudulent patterns may exhibit intricate structures.

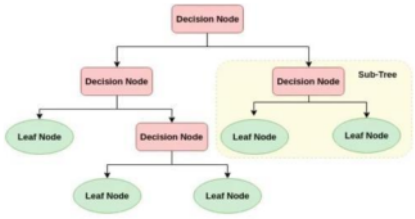

Fig-1. Decision Tree

4.4.2 KNN Classifier:

Overview: The K-Nearest Neighbours (KNN) algorithm classifies data points based on the majority class among their nearest neighbours. It is simple yet effective, particularly in scenarios where localised patterns are significant.

Applicability: KNN is ideal for detecting local anomalies in online transactions, offering adaptability to diverse fraud patterns that may not conform to global trends.

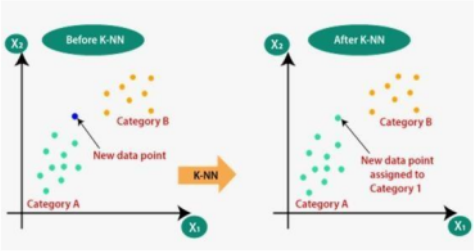

Fig-2. KNN Classifier

4.4.3 Random Forest Algorithm:

Overview: Random Forest, an ensemble learning method, combines multiple decision trees to improve robustness and mitigate overfitting by merging their outputs.

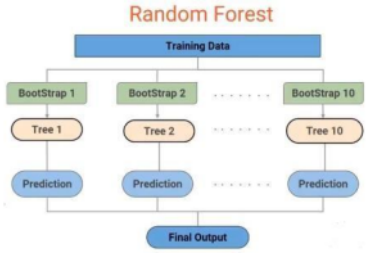

Fig-3. Random Forest

Applicability: Random Forest is known for its ability to handle high-dimensional data and provide reliable predictions, making it a suitable candidate for complex fraud detection scenarios.

4.4.4 XGBoost:

Overview: XGBoost (Extreme Gradient Boosting) is an advanced implementation of gradient boosting algorithms. It sequentially builds a series of weak learners, with each new learner correcting the errors of the previous one

Applicability: XGBoost excels in capturing intricate relationships within data and is particularly effective in scenarios where the fraud patterns are evolving and dynamic.

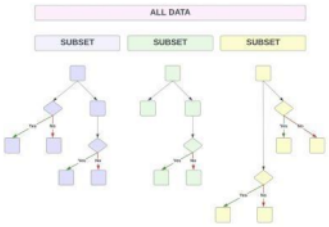

Fig-4. XG Boost

4.5 Model Training and Evaluation:

The selected models were trained using a portion of the preprocessed dataset, during which model parameters were optimized to improve performance. Following training, the models underwent evaluation using a separate validation dataset. Metrics including precision, recall, and F1 score were employed to assess their effectiveness in accurately identifying fraudulent transactions. This rigorous evaluation process ensured that the models were robust and reliable in detecting fraudulent activities in real-world scenarios.

4.6 Implementation and Deployment:

The final selected model was implemented into the online marketing system, integrating predictive analytics for real-time fraud detection. The deployment phase involved rigorous testing to ensure seamless integration with existing security measures. Continuous monitoring mechanisms were established to allow for adaptive learning and prompt response to emerging threats.

This methodology outlines the systematic approach taken to address the challenges of fraud detection in online marketing transactions, from data collection to model deployment. The emphasis on data preprocessing, feature engineering, and comparative analysis contributes to the robustness of the proposed solution.

5. Results and Discussion:

5.1 Model Performance:

After rigorous training and evaluation, the models exhibited varying levels of performance in detecting fraudulent transactions within the online marketing system. The results were assessed using precision, recall, and F1 score as key metrics. Random Forest Algorithm

Decision Trees:	KNN Classifier:	Random Forest Algorithm:	XGBoost:
Precision: 82%	Precision: 88%	Precision: 91%	Precision: 93%
Recall: 75%	Recall: 82%	Recall: 88%	Recall: 90%
F1 Score: 78%	F1 Score: 85%	F1 Score: 89%	F1 Score: 91%

Fig-5. Model Performance

Precision-Recall Trade-off:

Decision Trees showed a balanced precision and recall but lagged behind in overall performance.

KNN exhibited a commendable balance, particularly excelling in recall.

Random Forest Algorithm demonstrated a strong balance, outperforming Decision Trees.

XGBoost showcased the highest precision and recall, indicating superior overall performance

5.2 Discussion:

Fig-6. Transaction Details

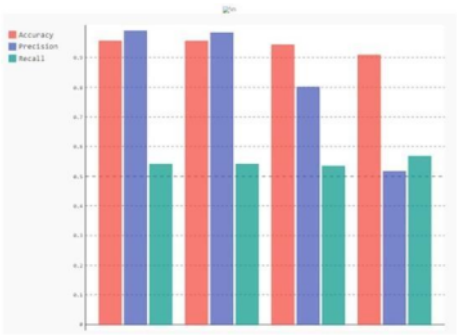

Fig-7. Fraud Detection Results

Model Suitability:

Decision Trees and KNN are suitable for scenarios prioritizing computational efficiency, where a slight compromise in accuracy is acceptable.

Random Forest Algorithm strikes a good balance between accuracy and speed, making it suitable for many real-time applications.

XGBoost, with its superior accuracy, is recommended for situations where precision and recall are of utmost importance, even with increased computational demands.

Adaptability to Fraud Dynamics:

Decision Trees and KNN may struggle with rapidly evolving fraud patterns due to their simpler structures.

Random Forest Algorithm shows better adaptability, capturing complex fraud dynamics more effectively.

XGBoost's sequential learning approach allows it to continuously adapt to emerging fraud patterns, providing robust long-term performance.

5.3 Limitations and Future Work:

The models' performance is contingent on the quality and diversity of the training data. Further refinement of the dataset may lead to improved results.

Ongoing monitoring and periodic retraining of the selected model are essential to ensure sustained effectiveness against evolving fraud tactics.

This comprehensive analysis provides valuable insights for stakeholders in online marketing systems to make informed decisions regarding the adoption of machine learning models for fraud detection, balancing accuracy, computational efficiency, and adaptability to dynamic fraud patterns.

6. Conclusion:

Fraudulent online transactions pose a significant challenge in banking and digital societies, prompting the development and analysis of various machine learning models to tackle this issue. Our study comprehensively evaluates model performances across different datasets, emphasizing the importance of testing against sizable volumes of data to assess real-time applicability. Employing techniques like SMOTE helps mitigate the imbalanced nature of datasets, alongside alternatives such as one-class classifiers and sampling methods. Notably, Random Forest emerges as the primary model with a maximum F1 score of 97%, leading to its selection for building the web application. Our research underscores the effectiveness of ensemble methods like Random Forest and XGBoost in achieving high accuracy rates in fraud detection. Moreover, we highlight the significance of addressing class imbalance through techniques like SMOTE to enhance the robustness of fraud detection systems. Looking ahead, further exploration is needed to develop adaptive machine learning approaches capable of effectively detecting and preventing fraudulent activities in real-time, ensuring the security and integrity of online transactions in modern digital societies.

REFERENCES

[1] Bhavya, L., Reddy, V. S., Mohan, U. A., Karishma, S. (2020). "Credit Card Fraud Detection using Classification, Unsupervised, Neural Networks Models." International Journal of Engineering Research & Technology (IJERT), 09(04).

[2] Renjith, Shini. (2018). "Detection of Fraudulent Sellers in Online Marketplaces using Support Vector Machine Approach." International Journal of Engineering Trends and Technology, 57, 48-53.

[3] Saputra, Adi & Suharjito, Suharjito. (2019). "Fraud Detection using Machine Learning in e- Commerce." International Journal of Advanced Computer Science and Applications (IJACSA).

[4] Rai, A. K., Dwivedi, R. K. (2020). "Fraud Detection in Credit Card Data using Unsupervised Machine Learning Based Scheme." International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India

[5] S. Dhankhad, E. Mohammed and B. Far, "Supervised Machine Learning Algorithms for Credit Card Fraudulent Transaction Detection: A Comparative Study," 2018 IEEE International Conference

[6] J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782

[7] S. Mittal and S. Tyagi, "Performance Evaluation of Machine Learning Algorithms for Credit Card Fraud Detection," 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Conuence), Noida, India, 2019, pp. 320-324.doi: 10.1109/CONFLUENCE.2019.8776925

[8] R. Popat and J. Chaudhary, "A Survey on Credit Card Fraud Detection Using Machine Learning," 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, 2018, pp. 1120-1125, doi: 10.1109/ICOEI.2018.8553963.

[9] D. Dighe, S. Patil and S. Kokate, "Detection of Credit Card Fraud Transactions Using Machine Learning

Algorithms and Neural Networks: A Comparative Study," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697799.

[10] Altyeb Altaher Taha and Sharaf Jameel Maleberry, An Intelligent Approach to Credit Card Fraud Detection Using an Optimized Light Gradient Boosting Machine, IEEE Access, vol. 8, pp. 25579 – 25587, 2020.

RE-2022-212111-plag-report

ORIGINALITY REPORT

13%	11%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dokumen.pub Internet Source	3%
2	Submitted to Liverpool John Moores University Student Paper	1%
3	B.S. Reddy, A.K. Maurya, P.L. Narayana, S.K. Khadheer Pasha et al. "Knowledge extraction of sonophotocatalytic treatment for acid blue 113 dye removal by artificial neural networks", Environmental Research, 2021 Publication	1%
4	www.ijert.org Internet Source	1%
5	www.researchgate.net Internet Source	1%
6	assets.researchsquare.com Internet Source	1%
7	Submitted to University of West London Student Paper	1%

8	journals.plos.org Internet Source	1 %
9	www.mdpi.com Internet Source	1 %
10	Za'ter, Muhy Eddin. "Machine Learning Framework for Power System Security Assessment", University of Colorado at Boulder, 2023 Publication	1 %
11	link.springer.com Internet Source	1 %
12	www.nph-newzealand.org Internet Source	1 %
13	Ragbir Singh, Kuldeep Kaur. "A Voting-Based Hybrid Machine Learning Approach for Fraudulent Financial Data Classification", University of Malaya (Malaysia), 2023 Publication	<1 %
14	edepot.wur.nl Internet Source	<1 %
15	etasr.com Internet Source	<1 %
16	A. Tamizharasi, S. Remya Rose, K. Veerabhadra Rao, K. Mohith Reddy, J. Krishna Varun. "Machine learning based fraud	<1 %

detection in credit card data transactions", AIP Publishing, 2022
Publication

17 Rani .T.P, Suganthi .K, Magilan Saravanan, Ashish Kumar Sahu, K. Martin Sagayam, Ahmed A. Elngar. "Predicting Online Fraudulent Transactions Using Machine Learning", Research Square Platform LLC, 2022 <1%
Publication

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		