ONLINE FRAUD DETECTION

A Project Report

submitted in partial fulfillment of the requirements

of

Edunet Foundation certification on Artificial intelligence with Cloud computing

by

CHIKKAM SARIKA

MORTHA MAHASREE

NAGIREDDY SWATHI PURNIMA

VADREVU SIRI VEERA BHAVYA

VENTRU CHITRA KALA SRI JANAKI

Under the Esteemed Guidance of

Name of Guide

ABDUL AZIZ SIR

ACKNOWLEDGEMENT

We would like to take this opportunity to express our deep sense of gratitude to all individuals who helped us directly or indirectly during this thesis work.

Firstly, we would like to thank my supervisor, RAMAR BOSE SIR for being a great mentor and the best adviser I could ever have. His advice, encouragement and critics are source of innovative ideas, inspiration and causes behind the successful completion of this dissertation. The confidence shown on me by him was the biggest source of inspiration for me. It has been a privilege working with him from last one year. He always helped me during my thesis and many other aspects related to academics. His talks and lessons not only help in thesis work and other activities of college but also make me a good and responsible professional.

ABSTRACT

Online fraud detection systems operate as a crucial line of defense against the ever-evolving tactics of fraudsters. These systems continuously analyze transaction data in real-time, using machine learning models to discern patterns, anomalies, and deviations from typical spending behavior

ONLINE FRAUD DETECTION

CHAPTER 1

INTRODUCTION

1.1. Problem Statement:

Anomaly Detection: Build an anomaly detection system that identifies unusual patterns in credit

card transactions to detect fraudulent activity.

1.2. Problem Definition:

In this fraud detection project, the selected machine learning algorithm is the **Decision Tree Classifier**. Decision trees are a widely used classification algorithm known for their simplicity and interpretability. They work by partitioning the dataset into subsets based on feature values, making decisions at each internal node to classify data into different classes.

1.3. Expected Outcomes:

To develop an effective anomaly detection system for credit card transactions aimed at identifying potential fraudulent activities, a systematic approach is essential. The first step involves collecting a comprehensive dataset comprising both legitimate and fraudulent transactions, ensuring it mirrors real-world scenarios with diverse transaction types, amounts, and user behaviors. Subsequently, the data undergoes preprocessing, addressing

missing values, scaling numerical features, and encoding categorical variables. The dataset is then split into training and testing sets for robust model evaluation.

CHAPTER 2

LITERATURE SURVEY

2.1. Paper-1

1.Information Retrieval Systems:

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.Salton, G., Wong, A., & Yang, C. S. (1975). A Vector Space Model for Automatic Indexing. Communications of the ACM, 18(11), 613620.

2.Decision Classifier:

In this fraud detection project, the selected machine learning algorithm is the **Decision Tree Classifier**. Decision trees are a widely used classification algorithm known for their simplicity and interpretability. They work by partitioning the dataset into subsets based on feature values, making decisions at each internal node to classify data into different classes. Decision trees are capable of handling both categorical and numerical data, which makes them suitable for the diverse nature of transaction data. While they can easily become too complex and prone to overfitting, techniques like pruning are often employed to prevent this and improve generalization. Decision trees offer transparency in understanding how decisions are made, making them useful for interpreting and explaining the rationale behind fraud detection predictions.

3.User Interface in Information Retrieval:

Hearst, M. A. (2009). Search User Interfaces. Cambridge University Press.Toms, E. G. (2000). The quest for effectiveness in interactive information retrieval. Journal of the American Society for Information Science, 51(7), 604-632.Streamlit Web Interface:Allaire, J., & Chang, W. (2019). shiny: Web Application Framework for R. R package version 1.3.2.9000.

CHAPTER 3

PROPOSED METHODOLOGY

3.1 System Design

Designing an effective anomaly detection system for credit card transactions involves a holistic approach to ensure accuracy, efficiency, and real-time monitoring. The system architecture typically includes components such as data collection, preprocessing, feature engineering, model selection, training, evaluation, deployment, and continuous monitoring.

3.2 Modules Used

In the development of an anomaly detection system for credit card transactions, various modules and libraries are typically employed to streamline the implementation process. Python, being a versatile programming language for data science and machine learning, is often the language of choice. Popular libraries such as scikit-learn are utilized for data preprocessing, feature engineering, and model selection. Scikit-learn provides a wide range of implementation-ready algorithms, including Isolation Forests, One-Class SVM, and Gaussian Mixture Models, facilitating the exploration and selection of the most suitable anomaly detection method.

In summary, a combination of Python libraries, machine learning frameworks, and web development tools forms the toolkit for building a

robust anomaly detection system, enabling the seamless integration of data processing, model training, deployment, and real-time monitoring functionalities.

3.3 Data Flow Diagram

The data flow design for the anomaly detection system in credit card transactions involves a series of interconnected steps to efficiently process information from raw data to actionable insights. Initially, the data flows into the system through a data collection module, where credit card transactions, both legitimate and potentially fraudulent, are gathered from various sources. Subsequently, the data undergoes preprocessing, managed by libraries like pandas and NumPy in Python. This step involves cleaning the data by handling missing values, scaling numerical features, and encoding categorical variables to prepare it for further analysis.3.4 Advantages

The implementation of an anomaly detection system for credit card transactions offers several significant advantages in the realm of fraud prevention and financial security. One of the primary benefits is the system's ability to detect potential fraudulent activities early on, thanks to its capacity to identify unusual patterns or deviations from normal transaction behavior. This early detection enables prompt intervention and mitigation measures, preventing financial losses and safeguarding both financial institutions and cardholders. Moreover, the adaptability of these systems to evolving fraud patterns ensures their effectiveness over time, as they can continuously learn and adjust to new tactics employed by fraudsters.

3.5 Requirement Specification

3.5.1 Hardware Requirements:

Processor: (Specify processor requirements)

Memory: (Specify memory requirements)

Storage: (Specify storage requirements)

3.5.2 Software Requirements:

The successful implementation of an anomaly detection system for credit card transactions necessitates a set of software requirements that collectively support the various stages of data processing, modeling, and deployment. First and foremost, a robust programming language such as Python is often chosen for its versatility and extensive libraries in data science and machine learning. Libraries such as scikit-learn, pandas, and NumPy are essential for tasks like data preprocessing, feature engineering, and model selection. For neural network-based approaches, TensorFlow or PyTorch serve as indispensable frameworks. The deployment phase typically involves web frameworks like Flask or Django to create APIs facilitating real-time transaction monitoring. Data processing and alerting mechanisms benefit from tools such as Apache Kafka or ELK (Elasticsearch, Logstash, Kibana). Moreover, version control systems like Git, integrated development environments (IDEs) such as Jupyter Notebooks, and containerization tools like Docker contribute to efficient collaboration and deployment practices.

CHAPTER 4

IMPLEMENTATION AND RESULT

The implementation of the anomaly detection system for credit card transactions began with the collection of a diverse dataset encompassing legitimate and fraudulent transactions. Python, leveraging libraries like scikit-learn and pandas, played a central role in data preprocessing, where missing values were addressed, numerical features were scaled, and categorical variables were encoded. Feature engineering extracted pertinent information such as transaction amount and user-specific details, enhancing the model's ability to identify anomalies. For the anomaly detection model, the Isolation Forest algorithm from scikit-learn was selected, trained on legitimate transactions to establish a baseline. Evaluation metrics, including precision, recall, and AUC-ROC, showcased the model's efficacy in distinguishing anomalies in a test dataset.

Upon achieving satisfactory results in the model evaluation phase, the system progressed to deployment. Using Flask for creating APIs, the model was integrated into a production environment, allowing real-time monitoring of incoming credit card transactions.

CHAPTER 5

CONCLUSION

In conclusion, the development and implementation of an anomaly detection system for credit card transactions represent a robust and proactive approach to enhancing financial security. Leveraging a diverse set of tools and technologies, including Python, scikit-learn, Flask, and continuous monitoring mechanisms, this system has demonstrated its efficacy in early fraud detection and prevention. The careful selection of the Isolation Forest algorithm and the thorough evaluation of the model on a diverse dataset showcased the system's ability to adapt to evolving fraud patterns while maintaining a low false positive rate.

ADVANTAGES:

The advantages of implementing an anomaly detection system for credit card transactions are multifaceted and contribute significantly to enhancing financial security:

1. **Early Fraud Detection:**
   - Anomaly detection systems excel at identifying unusual patterns, enabling the early detection of potentially fraudulent activities before significant financial losses occur.
2. **Adaptability to Evolving Patterns:**
   - These systems continuously learn and adapt to changing fraud patterns, ensuring effectiveness against new tactics employed by fraudsters over time.
3. **Reduced False Positives:**
   - By leveraging advanced algorithms and machine learning, anomaly detection systems can be fine-tuned to minimize false positives, avoiding unnecessary disruptions for legitimate cardholders.
4. **Real-time Monitoring:**
   - The ability to monitor transactions in real-time allows for immediate response to potential fraudulent activities, preventing further unauthorized transactions.

SCOPE:

The scope of implementing an anomaly detection system for credit card transactions is vast and extends across multiple dimensions. Primarily, it encompasses the proactive identification and mitigation of fraudulent activities, offering financial institutions a robust defense mechanism against evolving and sophisticated fraud tactics. The system's adaptability to changing patterns ensures its relevance over time, allowing it to stay ahead of emerging threats.

REFERENCES

Books:

Author(s). (Year). Title of the Book. Publisher.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.

Journal Articles:

Author(s). (Year). Title of the Article. Title of the Journal, Volume(Issue), Page Range.

Robertson, S., & Walker, S. (1994). Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.

GITHUB LINK :

(https://github.com/SARIKA-1/Online_Fraud_Detection)

VIDEO LINK:

https://youtu.be/H8_8OY_5rho?si=HDFnHllChzvbwqiv

APPENDIX