

Fraud Detection System Employing Machine Learning Techniques for Credit Card Transactions

G V S Abhishek Varma¹, G V N Akshay Varma², Adari Kethaan³

Student CSE Department, Maharaj Vijayaram Gajapathi Raj College of Engineering, AP, India^{1,3}

Student CSE Department, Sri Sivasubramaniya Nadar College of Engineering, TN, India²

Abstract: *This research paper presents a comprehensive demonstration of the Microsoft Fabric data science workflow, illustrating its effectiveness through an end-to-end example centered on the development of an advanced fraud detection system. The primary objective of this study is to construct a robust and efficient fraud detection mechanism by leveraging the capabilities of machine learning algorithms trained on historical data encompassing instances of fraudulent activities. The overarching aim is to discern intricate patterns inherent in fraudulent events, thus empowering the system to swiftly and accurately identify and flag such activities in case of their recurrence. This paper expounds upon a meticulously designed workflow that encompasses a series of pivotal steps, including the installation of custom libraries tailored to the task, meticulous data loading and preprocessing, a comprehensive exploratory data analysis phase aimed at extracting meaningful insights, the intricate process of training a machine learning model using Scikit-Learn and Flow, the critical step of selecting and registering the most performant model, and, finally, the seamless deployment of the trained model for real-time scoring and prediction. The core of the presented workflow centers on the concept of a lake house, wherein the data is sourced from a public blob and subsequently stored for comprehensive analysis. This architectural paradigm underscores the significance of unified data storage, offering a coherent platform for seamless integration and manipulation. The research paper emphasizes that this approach not only elevates the efficiency of data handling but also lays the foundation for consistent and structured analysis, ultimately enhancing the accuracy and applicability of the subsequent machine learning stages. By tackling the complexities of a real-world scenario involving fraud detection, this research paper underscores the versatility and adaptability of the Microsoft Fabric framework. While the primary focus is on the development of a robust fraud detection system, the significance of this approach reverberates across diverse data science endeavors. The intricate process of custom library installation, data ingestion, preprocessing, and model deployment provides an invaluable resource for practitioners navigating the multifaceted landscape of data science workflows within the Microsoft Fabric ecosystem. In a world characterized by an exponential surge in data generation and a corresponding demand for actionable insights, the presented paper serves as more than just a solution to the specific problem of fraud detection. It morphs into a foundational template, empowering data scientists, researchers, and industry professionals with a structured approach to harnessing the potential of complex datasets. The step-by-step elucidation of the workflow equips practitioners with a tangible guide, offering insights into handling intricacies that often accompany large-scale data analysis. Through its detailed exposition, the paper bridges the gap between theory and practice, allowing readers to not only comprehend the theoretical underpinnings of the Microsoft Fabric framework but also practically implement and adapt its methodologies to suit a myriad of data science challenges. In conclusion, this research paper transcends the confines of a traditional study on fraud detection by encompassing a holistic exploration of the Microsoft Fabric data science workflow. While effectively addressing the practical challenge of developing an advanced fraud detection system, it simultaneously contributes a substantive and versatile resource for the data science community at large. As organizations across domains seek to extract maximum value from their data assets, the workflow detailed in this paper*

offers a beacon of guidance, ensuring that the journey from data to insights is traversed with precision and efficacy.

Keywords: Fraud Detection, Data Science Workflow, Machine Learning, Microsoft Fabric

REFERENCES

- [1]. Abhishek Varma G V S, Akshay Varma G V N."Dynamic User Routing for Paid and Free Users in Web Applications using Content Delivery Network (CDN)", Volume 11, Issue VII, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 409-414, ISSN : 2321-9653, www.ijraset.com
- [2]. <https://github.com/varma59/Creditcard>
- [3]. PER, ORIGINALRESEARCH PA. "Bat optimization algorithm for wrapper-based feature selection and performance improvement of android malware detection." (2021)..
- [4]. Ramesh, Mr P., and Mr T. Narayan Rao. "The Sensitivity of Service Quality on Customers towards Overall Satisfaction."
- [5]. Delamaire, Linda, Hussein Abdou, and John Pointon. "Credit card fraud and detection techniques: a review." Banks and Bank systems 4.2 (2009): 57-68..