Self-Adaptive Financial Fraud Detection System

Group Number & Members: G9 (Aman Kumar and Manan Raheja)

1. Abstract

In this project, we aim to address the pressing issue of fraudulent transactions in the global financial industry. The objective is to develop a machine learning model that accurately identifies fraud while adapting to evolving fraud patterns. The proposed solution combines conventional machine learning models with real-time self-adaptive capabilities enabled by continuous performance monitoring and automated model re-training.

2. Problem Description

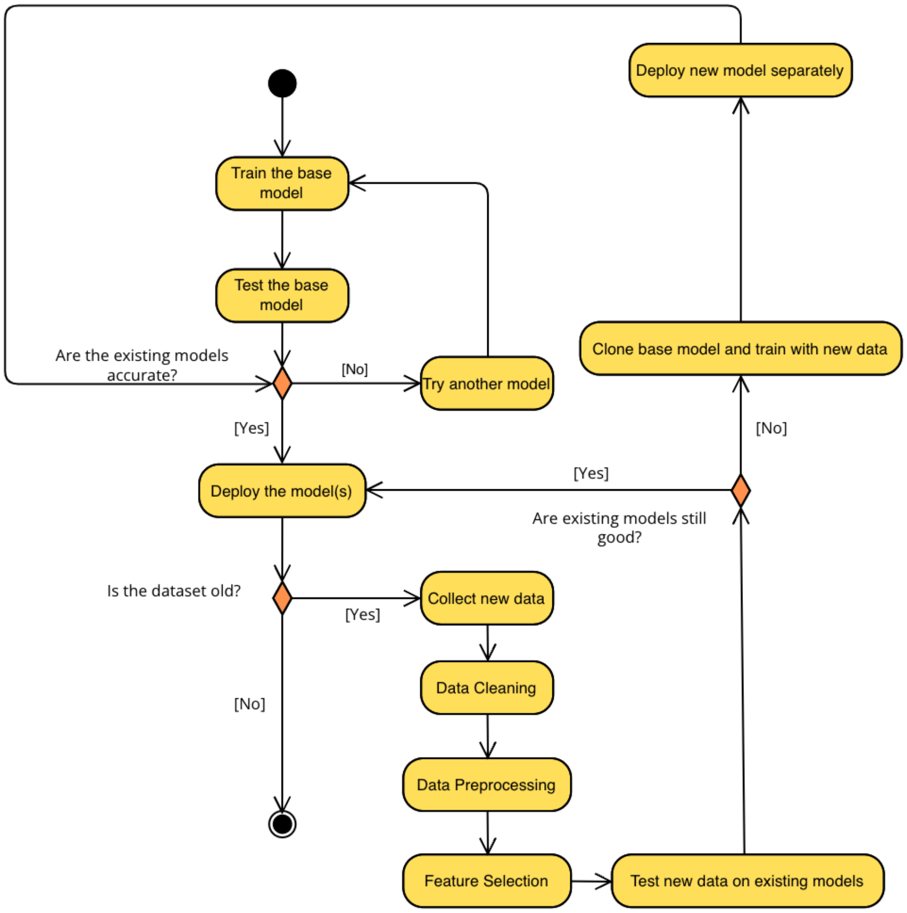
Financial transactions occur on the scale of millions of transactions per second, all around the world. Most of these transactions are also time critical and the financial institutions need to be certain whether the transaction being carried out is fraudulent or not. Fraudulent transactions are a global problem with ties to money laundering, terror financing, trading in contraband substances, human trafficking and other malicious purposes. Therefore, at the bank level, there is an indispensable and ever evolving need to have state-of-the-art fraud detection and prevention algorithms so that these illegal activities can be prevented and the customers remain protected from fraud. Very often, people knowingly commit anonymous transactions in order to prevent themselves from being tracked by certain corporations or individuals. This poses another challenge in determining the authenticity of transactions. The task is to develop a machine learning model that accurately determines whether a transaction was fraudulent or not, and keep it relevant with the latest kinds of patterns in fraud detection.

3. Proposed Solution

The conventional machine learning approach is usually effective in scenarios where the model is designed for use by a relatively small user base, does not necessitate frequent scalability adjustments, and is not intended for real-time prediction tasks. Instead, it primarily serves the purpose of making predictions based on pre-fed data inputs, offering a stable and well-contained solution for specific predictive analytics needs.

However, in case of fraud detection, to be able to perform effectively, we need to make sure that the model is not outdated and is aware of the latest fraud trends. Therefore, it is essential that the model self-adapts to the real-time financial information by learning the latest data. In order to make a self-adaptive system, we require identifying and monitoring the performance of our existing model and then improving the model performance by re-training the model when the performance drops a certain threshold .

As manual model training is not practical, we implement an automation pipeline to handle data validation, preparation, model training and feedback-loop for re-training. Through this process, the training and analysis of the model are taken care of automatically. We just have to check if proper data is available and make sure there isn't a skewed dataset so that the model is trained properly. This improves the scalability and real-time predictions of the model.



To maintain optimal performance, the system is continuously monitored. Any drop in performance triggers a retraining of the model with new data. This iterative process ensures that the model stays up to date and accurately reflects real-world changes, effectively serving its intended purpose.

4. Tools/Technologies Used

For data engineering and ML model development, mainly Python would be used. However, for model deployment and monitoring, we will work with Docker, Kubernetes and AWS Machine Learning, and machine-learning lifecycle management platforms like MLflow.

5. Project Schedule

The project is scheduled to be executed under the given timeline -

Week 1: Conducting research into state-of-the-art machine learning techniques for financial fraud detection. Finalizing the tools for machine-learning operations (MLOps) management. Create a robust system design. Setting up the development environment: obtaining dataset, defining model architecture.
Week 2: Build the machine learning pipelines: data cleaning, data preprocessing, feature selection, model training and evaluation, model optimizing/tuning. Conduct ML experiments to optimize model performance. Develop deployment scripts.
Week 3 & Week 4: Deploy the model. Setup model performance monitoring. Create the feedback loop to make the system self-adaptive.
Week 5: Report writing. Presentation preparation. Buffer for delays.

6. References

1. Sorournejad, S., Zojaji, Z., Ebrahimi Atani, R., & Monadjemi, A. H. (2016). A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective. arXiv preprint. https://arxiv.org/abs/1611.06439
2. Boulieris, P., Pavlopoulos, J., Xenos, A., et al. (2023). Fraud detection with natural language processing. Machine Learning, https://doi.org/10.1007/s10994-023-06354-5.
3. Kovach, Stephan & Ruggiero, W.V.. (2011). Online Banking Fraud Detection Based on Local and Global Behavior. In: Proc. of the Fifth International Conference on Digital Society. https://www.researchgate.net/publication/228616927_Online_Banking_Fraud_Detection_Based_on_Local_and_Global_Behavior
4. Lawal, Solomon, Fraud Detection and Prevention: A Synopsis of Artificial Intelligence Intervention in Financial Services Smart Card Systems (September 23, 2021). Available at SSRN: https://ssrn.com/abstract=4117507
5. Nexocode. (July 11, 2022). AI-Based Fraud Detection in Banking and Fintech: Use Cases and Benefits. Nexocode Blog. https://nexocode.com/blog/posts/ai-based-fraud-detection-in-banking-and-fintech-use-cases-and-benefits.