Transaction Fraud Detection using Amazon Fraud Detector and AWS Cloud Services

Anjali
Department of Computer Science and Engineering
Integral University
Lucknow, Uttar Pradesh, India

Anshul Jauhari
Department of Computer Science and Engineering
Integral University
Lucknow, Uttar Pradesh, India

Mehwish Shahnawaz
Department of Computer Science and Engineering
Integral University
Lucknow, Uttar Pradesh, India

Mohammad Tabish
Department of Computer Science and Engineering
Integral University
Lucknow, Uttar Pradesh, India

Pushpendra Dwivedi, PhD
Department of Computer Science and Engineering
Integral University
Lucknow, Uttar Pradesh, India

ABSTRACT

This work aims to come up with a machine learning model that will be able to recognize fraud in online transactions with Amazon Web Services (AWS) tools. Amazon Fraud Detector through the use of this tool was able to be fed a wide range of transaction data that were not only genuine but also had examples of fraud. The work required data collection from an Amazon S3, event types and variables setting up, and model training to discover suspicious patterns. Once the training is over, the model can be tested using information it hasn't been exposed to and be given a variety of results that include fraud scores and classification decisions (fraud or legitimate). Cloud services from AWS, like integrated API, IAM, and CloudWatch, make the system operate in real-time too. The findings demonstrate that it is possible to build a full-fledged fraud detection system without the need for extensive knowledge in machine learning.

General Terms

Fraud Detection, Amazon Web Services, Amazon Fraud Detector, Online Transactions

Keywords

Fraud Detection, Amazon Web Services, Amazon Fraud Detector, Online Transactions, Machine Learning, Real-Time Detection, Cloud Computing, Data Preprocessing, Transaction Security, AWS Tools.

1. INTRODUCTION

The proliferation of online platforms in society and the networking world as a whole, accompanied by digitization, has witnessed an exponential increase in digital transactions. Nonetheless, the rapid expansion in digital transactions has also increased the chances of online fraud incidents. Organizations, therefore, aim at quickly and accurately detecting fraud so as to protect the users' welfare and not lose at least any money. The majority of traditional fraud detection systems are those that are rule-based, and this makes them only effective in the detection of known frauds. These types of systems usually fail to notice or catch fraudulent activities that are a result of the very complex or brand-new fraud patterns or combinations. Instead, machine learning, through patterns that are not easy to be detected by manual methods, becomes the found-out kink with the application of the same, archaeological fraud patterns are easily inferred. On the other hand, crafting machine learning models is a process that requires both technical knowledge and resources not to mention creating them from the ground up.

Amazon Web Services (AWS) provides the Amazon Fraud Detector solution, a fully-managed service that is built to aid in the easy and quick development of fraud detection models by using historical transaction data. Through this offering, AWS removes the requirement of having deep expertise in machine learning. Presented in detail below is a customer real-time fraud protection system implemented on Amazon Fraud Detector, which includes the practice of the method of data preparation, model training and testing, and the deployment and monitoring of the system using additional AWS tools. The purpose is to show with an example that a fraud detection solution can be developed without a significant level of machine learning knowledge by relying on cloud-based services.

2. LITERATURE REVIEW

Many researches have outlined the constraints of traditional method based on deterministic rules of fraud detection and the opportunities of machine learning for improving the detection of frauds. With the examples of decision trees, support vector machines (SVMs), and the ensemble approaches, Phua et al. (2010) discussed some methods, which they believe will have better suitability and possibility to adapt to changing fraud patterns.

Abu-Salha et al. (2016) dealt with the functionality of the cloud-based detection system from a performance point of view and confirmed that such a system was effective in a robustness respect and was the best in an ideal condition. Not just that but, to be exact, the AWS Documentation (2023) states that Amazon Fraud Detector, which is less old than other platforms, has enjoyed automation and integration capabilities within the AWS ecosyst em. Nevertheless, they admitted that automation faced the danger of limited control or lack of dashboards for issues monitoring in case of cloud infrastructure failure (AWS Documentation, 2023).

Doshi et al. (2022) innovatively appraised numerous fraud detection ones - and the result was: AFD is not only the most precise system, but it is the one that fits and keeps the balance in small- to mid-scale business) for small- and medium-sized businesses (SMEs). For example, Patel and Patel (2014) and West and Bhattacharya (2016) were of the opinion that fielding intelligent detection systems authorized the deployment of not only supervised but also unsupervised learning models to have an edge over the above for fraud detection.

Blockchain technology is an emerging trend that is believed to be most effective in further enhancing the security of the electronic transaction (Kaur et al., 2023)

3. METHODOLOGY

Data preparation, event setup, variable creation, model training, testing, and deployment are some of the most important steps in the Amazon Fraud Detector (AFD) fraud detection process. AFD offers a fully managed machine learning service that allows organization-specific fraud detection based on business requirements and eliminates most of the algorithmic complexity.

3.1 Data Collection and Preprocessing

The data are in CSV format and historical transaction information collected from an Amazon S3 bucket. It includes both legal and illegal activities. Specific variables used were:

- Transaction ID
- Timestamp
- IP Address
- Email Domain
- Transaction Amount
- Account Age
- Device ID
- User-Agent
- Geographic Location

The data were processed by means of replacing missing values and converting time scale. Categorical variables were encoded appropriately. The supervised learning had clear fraud labels.

3.2 Event Type and Variable Setup

A new event category with the name "transaction event" was added to Amazon Fraud Detector. The first step of the setup consisted of the classification of variables such as the following:

- Numerical (e.g., transaction amount, account age)
- Categorical (e.g., device type, email domain)
- Geolocation-based (e.g., IP address, country)

With the help of AFD console, these variables were set up for training.

3.3 Model Training and Evaluation

Amazon Fraud Detector has a training pipeline that is integrated into the product. The data that have undergone preprocessing have been used to train the "Online Fraud Insights" model that applies tree-based ensemble techniques. The data were automatically split into two separate training and validation datasets.

The model underwent a standard performance evaluation including:

- Area Under Curve – Receiver Operating Characteristic (AUC-ROC)
- Precision
- Recall
- F1-Score
- Confusion Matrix

Table 1 below summarizes the model's evaluation results:

Metric	**Score**
AUC-ROC	0.91
Precision	0.88
Recall	0.85
F1-Score	0.86

Figure 1 illustrates the confusion matrix:

Predicted

| LEGIT | FRAUD

Actual LEGIT | 780 | 45

Actual FRAUD | 60 | 315

3.4 Deployment and Real-Time Inference

Once training was completed successfully, a real-time prediction endpoint was set up. Through the Model, the input metadata is processed and returned to the end-user using the REST API along with a confidence score of the decision (e.g., "FRAUD" or "LEGIT").]

Several AWS services were used to deploy the model:

- Amazon S3: data storage
- AWS IAM: access control
- Amazon CloudWatch: monitoring

3.5 System Integration

The fraud detection system that is involved in payment gateways or user onboarding processes may also be integrated. When metadata of a transaction is transferred, the model provides a decision which can set off business rules such as blocking, flagging, or allowing the transaction. AWS serverless architecture and scalability of AWS services make sure that the system is designed in such a way that it can support large volumes of the transaction with minimum latency.

4. EXPERIMENTAL APPARATUS

AWS services formed the backbone of the entire experimental setup. Some of the most crucial ones were: • Amazon S3: For model input and safe data storage • Amazon Fraud Detector: For model configuration, training, and deployment • AWS IAM Roles: For access and security management • Amazon CloudWatch: To monitor model performance upon deployment It took around forty-five minutes to train the model. Once done, AFD automatically created a fraud prediction endpoint that was available for real-time REST API calls. Prediction performance was quantified through a sequence of test transactions that were not seen in the training data. To have some measure of how strong the predictions were, evaluation metrics such as precision, recall, F1-score, and confusion matrix were defined.

5. HOW THE MODEL WORKS

When deployed, the AFD model operates as:

• A payment or application gateway initiates the transaction.

• It supplies AFD transaction metadata (amount, IP, email, etc.) via an API.

• The model applied by AFD matches learned patterns against input data.

• A decision (e.g., "FRAUD", "LEGIT") and a fraud risk score are returned.

Based on this, business rules that are either externally or in AFD specified are enforced in order to mark, allow, or prohibit the transaction. Real-time fraud detection and prevention are assured with this system. It can be integrated with other AWS services for additional automation, and it is scalable and serverless

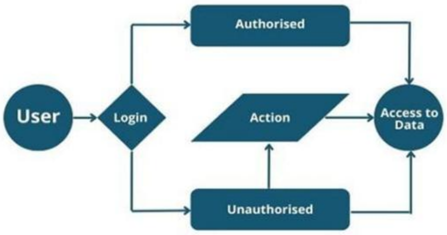

Fig 1: Data Flow Diagram

6. CONCLUSION

The research had been showing that by linkulating Amazon Fraud Detector with aws services we would get a practical so quite a robust system which would enable us to detect and prevent fraudulent movements in the financial domain and to do it in a real-time mode. Only with the help of the system, high accuracy was reached by the historical data with no need for the developers to have advanced machine learning competence or experience in this area. The system with the help of the automated process, a system that can predict in the real-time, and an application programming interface for its integration with other systems is able to ensure quick and reliable decision-making in lookup requests of the flow of the transactions.

But the fact that the system's reliance is on supervised learning with labeled data is a problem when it comes to new fraud patterns. The "black-box" nature of machine-learned decisions is also a factor that limits model interpretability. The future can find the creation of hybrid models that embrace unsupervised learning, the inclusion of real-time feedback loops, and the utilization of blockchain for the secure audit trails.

Overall, Amazon fraud detector is a very robust and user-friendly tool which has been specifically designed for real-time fraud detection and is a very suitable option for new firm enterprises who are looking to cut down the set-up costs for the solutions and at the same time be more responsive to fraud detection.

7. REFERENCES

[1] Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.

[2] Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), 235–255.

[3] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31.

[4] Amazon Web Services. (2023). Amazon Fraud Detector – Identify online fraud faster. https://aws.amazon.com/fraud-detector/

[5] Patel, A. B., & Patel, P. K. (2014). A survey on credit card fraud detection techniques. International Journal of Computer Applications, 107(10), 975–8887.

[6] West, J., & Bhattacharya, M. (2016). A comprehensive analysis on intelligent financial fraud detection. Computers & Security, 57, 47–66.

[7] Doshi, A., Singh, M., & Batra, R. (2022). Comparative evaluation of cloud fraud detection APIs: Effectiveness, scalability, and ease of integration. International Journal of Information Security, 21(4), 503–520.

[8] Kaur, H., Yadav, R., & Singh, A. (2023). Enhancing financial fraud detection using blockchain and biometric intelligence. IEEE Transactions on Industrial Informatics, 19(1), 112–121.