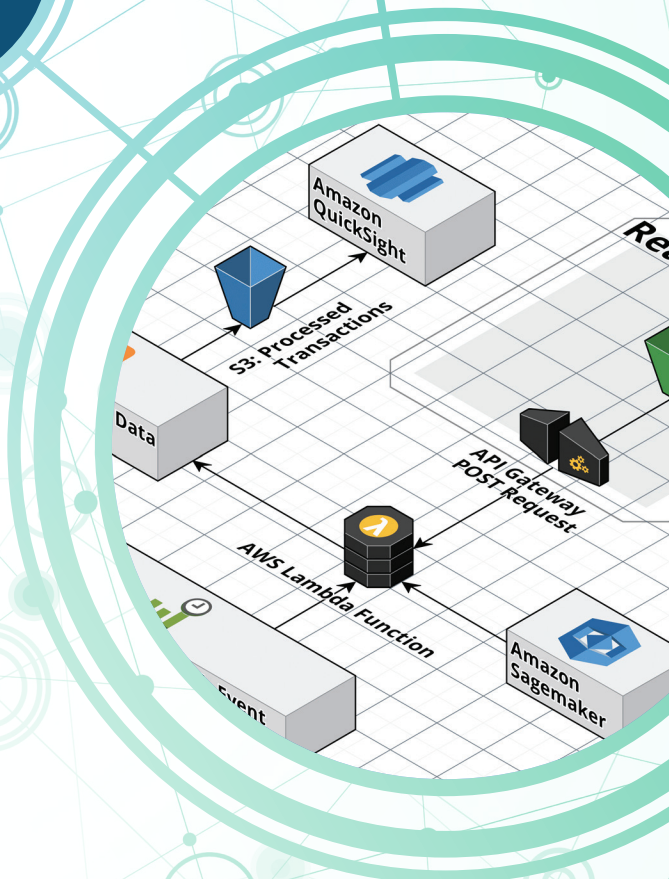
Unissant

Unissant White Paper Series

Fraud Detection Using Machine Learning on Cloud



Introduction

Businesses and public around the globe are losing billions of dollars every year and it's increasing due to fraudulent transactions. In a PwC survey conducted in 2018, 49% of the companies asked had dealt with fraud. This is an increase from the 2016 study where 36% of companies were affected.

The market for fraudulent transactions or data is relevant to almost every major industry. Below are few use cases:

- Financial Fraud: In 2018 alone, $24.26 Billion was lost due to payment card fraud worldwide. Firms in banking industry face fraudulent scenarios such as a scamster/hackers using another person's credit card to make big purchases or make repeated easy to miss small charges, etc.
- Insurance Fraud: A claimant may declare their car was damaged in flooding, but their social media shows they were at a completely different location. Firms are now utilizing internal data such as customer social media, call center notes, personal details, and voice recordings to gain insight into potentially fraudulent claims.
- Telecom Fraud: Fraud in this industry is generally demonstrated when customers don't pay bills and have suspicious long-distance calling patterns within six months of joining the service. Firms in the telecommunication industry analyze call record data coupled with customer data in order to build a complete profile.
- CMS Medicare and Medicaid Fraud: Fraud in Medicare and Medicaid is observed when providers bill for services that weren't provided through phantom billing, upcoding, etc. CMS has been employing several data mining algorithms and detection models to detect scenarios such as medicines prescribed that are not supposed to be prescribed at the same time in single prescription, etc.

FRAUD DETECTION – Approach

Traditional methods of data analysis have been used to detect fraud for a long time, but the techniques used to commit fraud are also getting more and more sophisticated. Fraud often consists of many instances or incidents involving repeated transgressions using the same method. Fraud instances can be similar in content and appearance but usually not identical. Thus, effective fraud detection requires complex and time-consuming investigations that deal with different domains of knowledge like finance, economics, business practices, and law.

A popular method for detecting fraud currently are rule-based engines. There are several problems with these methods. Below are some of the biggest challenges with preexisting rule-based systems:

- Rules are effective only when they are actively monitored and managed by dedicated fraud teams. This means fraud teams need to be staffed to review manual review queues, rejects and chargebacks and recalculate thresholds that can be codified in a rule. This makes fraud prevention reactive in a fast-moving business environment than being pro-active in preventing frauds.
- Rules are trained to do what is told without adding intelligence. They follow a binary view of whether the rules criteria are met or not. They do not dynamically adjust themselves for normal behavior or seasonal fluctuations. This would result in more false positives and undetected frauds.

- Rules that are based on a single channel/device do not provide a holistic view of consumer activity across multiple channels. 68% of fraud today is a cross channel. Thus, the importance of building a complete customer profile across channels.
- Incorrect and poorly coded rules increase manual review queues and continue to result in high fraud rates.

Machine Learning as the Solution

Machine Learning provides a much more flexible approach to address constantly evolving fraud Tactics, providing a better solution than the traditional methods that utilize rigid active rule-based systems.

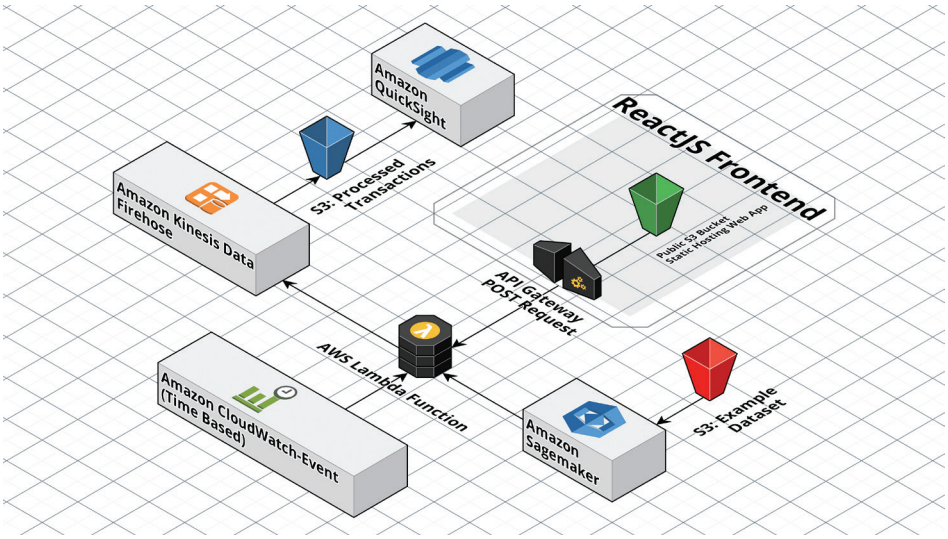
- Wide Parameters: Machine learning and artificial intelligence solutions may be classified into two categories making use of a wider range of parameters: supervised and unsupervised learning. These methods seek for accounts, customers, suppliers that behave unusually in order to output suspicion scores, rules, or visual anomalies depending on the method.
- Limitations: Despite the methods used the output gives us only an indication of fraud likelihood. No stand-alone statistical analysis can assure that a particular transaction is a fraudulent one, but they can identify them to high degrees of accuracy.

Advancement in Machine Learning libraries has been happening at a rapid pace all the way from hardware to software libraries. Machine learning training is very computationally intensive involving lot of experimentation. Advances in Cloud computing have made it possible overcome this limitation by providing access to a scalable infrastructure with access to highly powerful computing resources. , Recent machine learning libraries make use of several hardware level features such as GPUs for deep learning, etc.

Example Implementation

At Unissant, we have developed a POC in AWS environment for detecting fraud in financial domain. A we used a publicly available anonymized credit card transaction dataset that was collected and analyzed during a research collaboration of Worldline and the Machine Learning Group of Université Libre de Bruxelles on big data mining and fraud detection. After gathering required data for building a ML model, data is prepared for ingestion. On that data, various algorithms are tested using libraries such as SKLearn and AWS's Linear Learner Model to find best scoring algorithm and trade-offs. For identifying fraud, the metrics such as Sensitivity, F1 Score are used. Sensitivity Measures Identifying Fraud. F1 Score Measures Legitimate Customers.

We implemented the solution in AWS environment as shown in the high-level architecture shown below:



We used the following services from AWS:

- CloudWatch: This event is configured to run every minute and triggers a Lambda Function that processes transactions from the dataset
- Lambda Function: The lambda function is set to process transactions from the dataset and invokes the SageMaker endpoint
- SageMaker: Utilizing the Linear Learner model predicts whether transactions are fraudulent.
- Kinesis: The delivery stream loads the processed transactions into another Amazon S3 bucket for storage.
- QuickSight: Pulls the transactions from S3, for visualization, reporting, ad-hoc queries, and more detailed analysis.
- API Gateway: Request from Lambda function uses API Gateway to call onto the SageMaker which returns a prediction in JSON format to the frontend.

Conclusion

Though the example implementation was created for detecting fraud in credit card transactions, the implementation can be extended and scaled to detect fraud in other domains. This could be made by preprocessing a dataset and utilizing the created model. After this is done frontend can be easily be reconfigured.

About Unissant

Unissant is an advanced data analytics and business transformation services provider with expertise in healthcare and health IT, finance, national security, and energy. The company delivers innovative solutions to assist government agencies and private sector businesses in tackling their biggest challenges. Founded in 2006, Unissant is a prime contractor on various government vehicles such as CIO-SP3, GSA PSS, GSA HealthIT SIN, and GSA 8(a) STARS II and is a CMMI Level 3, ISO 9001 & 27001 certified company headquartered in Herndon, Virginia with a satellite office in San Antonio, Texas. In March 2017, Unissant received the Government Project of the Year award by Small and Emerging Contractors Advisory Forum (SECAF). Unissant has experience with implementing machine learning solutions and full stack implementation on cloud.

Unissant is a Select Consulting Partner Public Sector and is leveraging this relationship with AWS to strengthen our ability to support our customers in their Digital Transformation journey. We are already supporting several of our customers in the public sector with their Machine Learning implementations, cloud adoption strategy and implementation.

Copyright

Restricted Rights Legend

For additional information on Unissant full range of services, please visit our website at www.unissant.com or call us at 703.889.8500.