Supervised Anomaly Detection & Real-time Alerts Using aws

Team 8
Zheming Lian, Chuchen Xiong, Shunshun Miao,
Pahal Patangia, Tabassum Fazel, Jiahui Jiang

Outline

Current Scenario & Industry Applications

Limitations in Current Setup

- Traditional anomaly detection system are rule based
- Do not conform to the evolving data ecosystems
- Siloed data sources - limited view
- High number of false positives - costs are punitive
- Latent response

What our framework can bring

- Streaming data
- Real-time Alerts
- Incremental Training
- Daily Updated Dashboard

Business Use Case & Application

Banking

Transaction Fraud Detection

Manufacturing

Identifying Defect Machines

Health Care

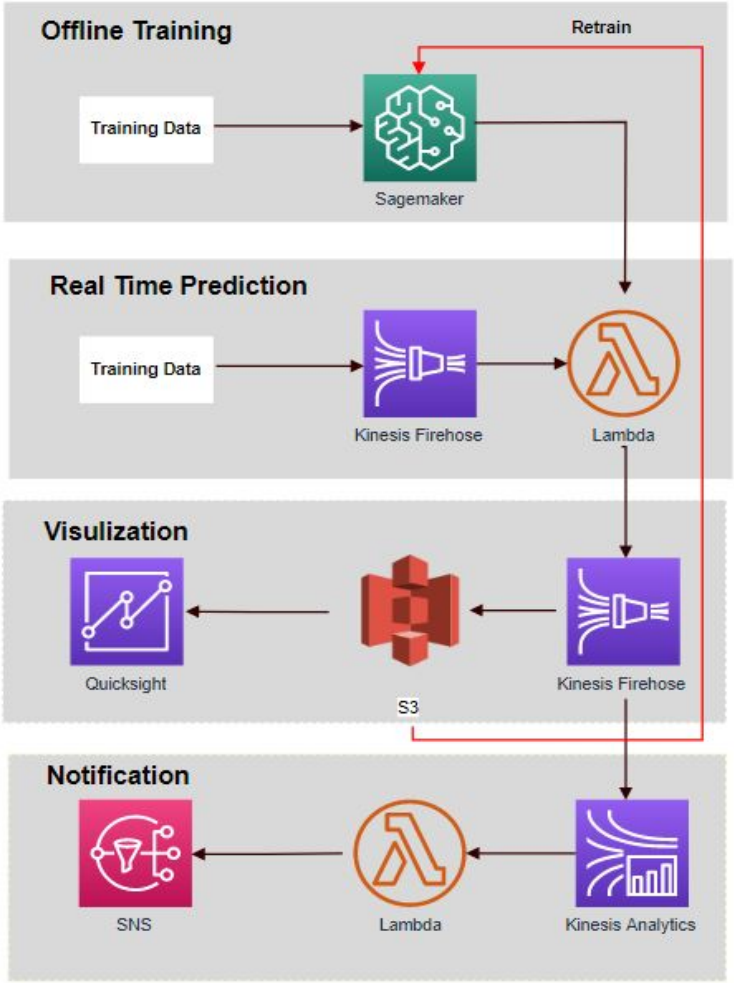
Diagnosis of Diseases

Cyber Security

Discern Security Threats

Solution Overview

Process Diagram



AWS Services utilized

Computation

SageMaker

AWS Lambda

- ML training
- Data processing
- Invoke services

Streaming processing

- Capture and Load
- Process with SQL

Storage

S3

- Historical data
- Service endpoint
- Prediction result
- Incoming data

Others

SNS

QuickSight

- Notification
- Dashboard

Application Demo

- Credit Card Fraud Detection

Data Overview

Question to answer: How to recognize fraudulent credit card transactions?

Variables

Original Variables

Customer Profile

- Salary
- Number of cards
- ...

Transaction record

- Average amount
- Transaction time
- ...

PCA Dimension Reduction

PCA Variables

- V1
- V2
- ...

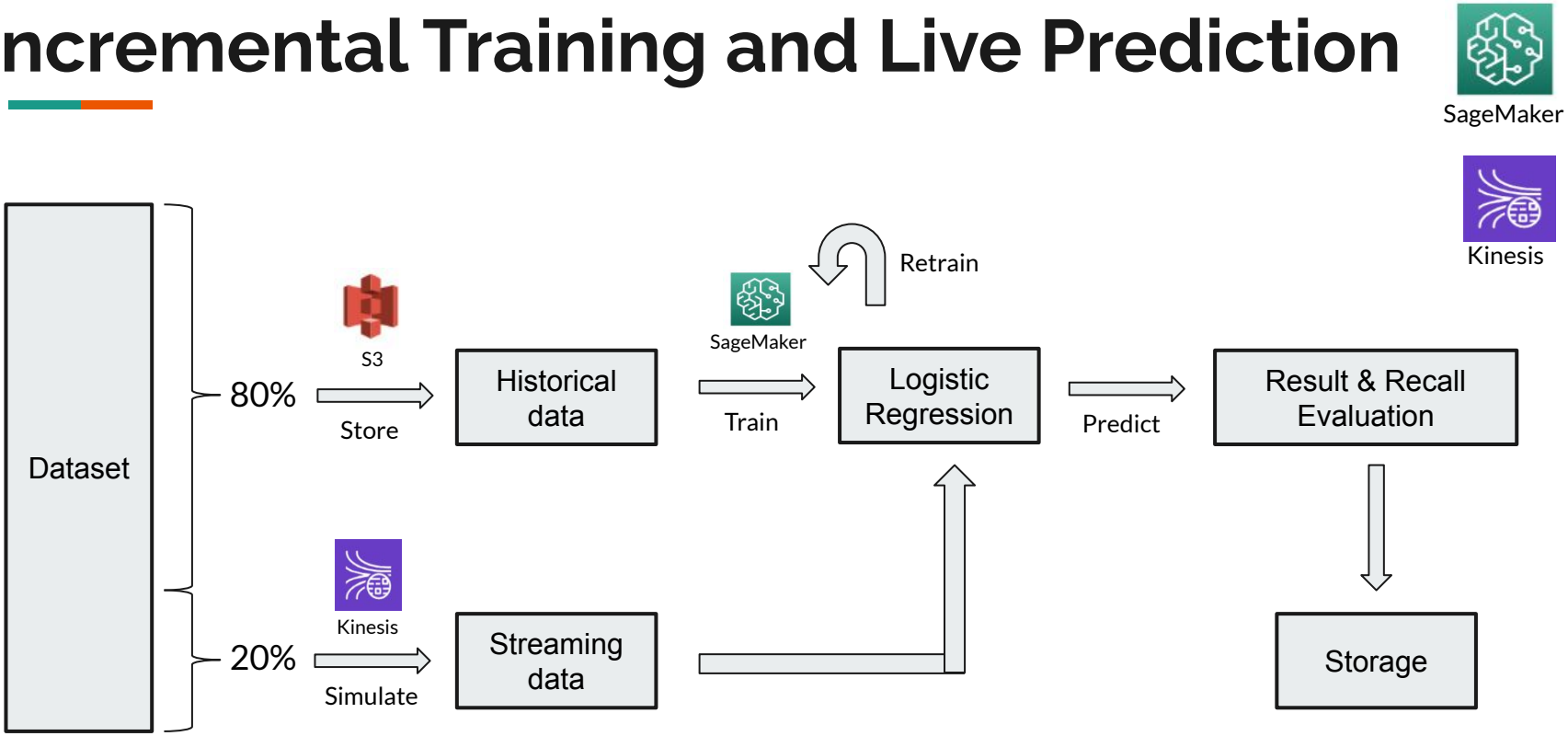
Label

0: Not Fraud	99.8%
1: Fraud	0.2%

V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	Class
-0.99139	-0.31117	1.468177	-0.4704	0.207971	0.025791	0.403993	0.251412	-0.01831	0.277838	-0.11047	0.066928	0.128539	-0.18911	0.133558	-0.02105	0
-0.59522	-4.28925	0.389724	-1.14075	-2.83006	-0.01682	0.416956	0.126911	0.517232	-0.03505	-0.46521	0.320198	0.044519	0.17784	0.261145	-0.14328	1
0.489095	-0.14377	0.635558	0.463917	-0.1148	-0.18336	-0.14578	-0.06908	-0.22578	-0.63867	0.101288	-0.33985	0.16717	0.125895	-0.00898	0.014724	0
0.717293	-0.16595	2.345865	-2.89008	1.109969	-0.12136	-2.26186	0.52498	0.247998	0.771679	0.909412	-0.68928	-0.32764	-0.1391	-0.05535	-0.05975	0
0.507757	-0.28792	-0.63142	-1.05965	-0.68409	1.965775	-1.23262	-0.20804	-0.1083	0.005274	-0.19032	-1.17558	0.647376	-0.22193	0.062723	0.061458	0

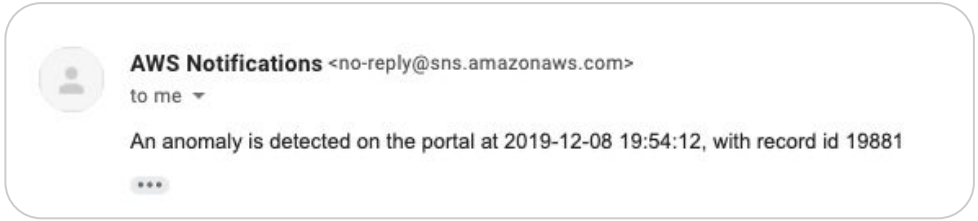
Data Source: https://www.kaggle.com/mlg-ulb/creditcardfraud

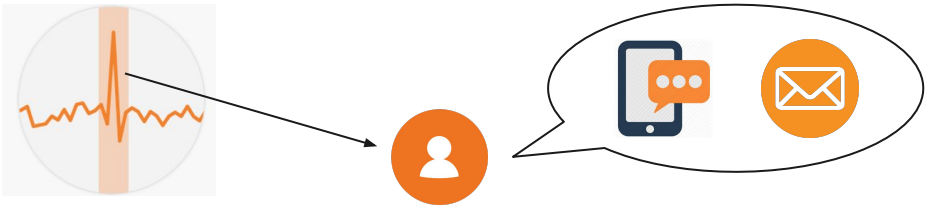
Incremental Training and Live Prediction



Real-time Notification

SNS



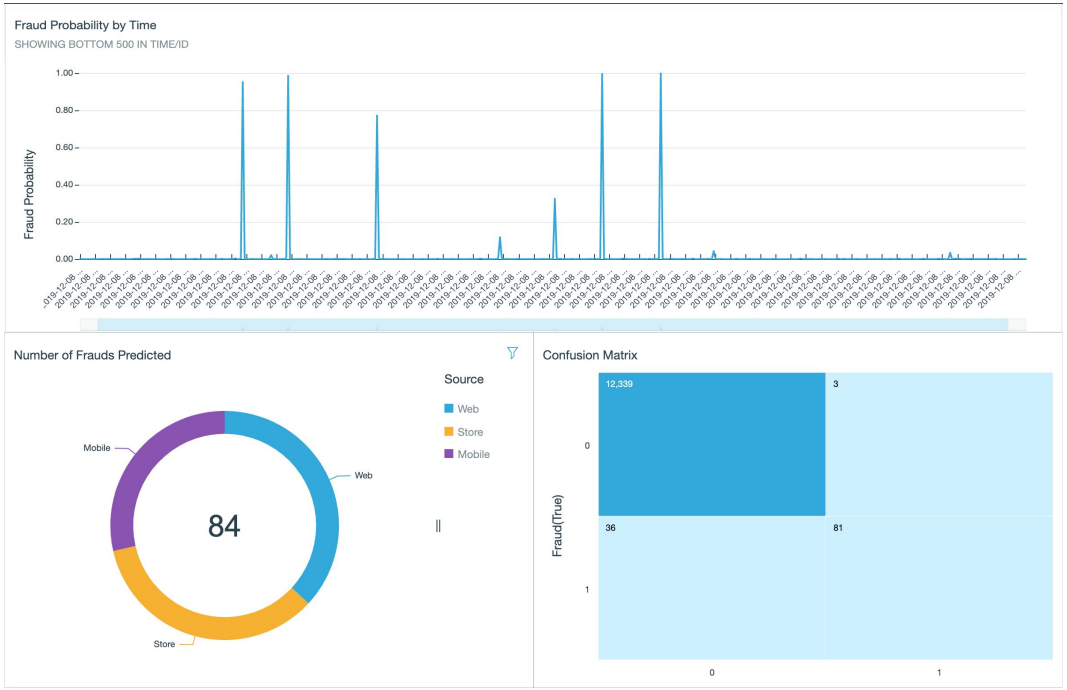


- Extract the anomaly data with Kinesis Analytics
- Lambda respond to the anomaly stream and process SNS notifications
- SNS topics fan out messages to subscribers and users receive real time alert

Daily Visualizations

QuickSight

- Summarize events with easy-to-understand, daily based data visuals
- Monitor model performance on daily live data



Sample Dashboard

Limitations & Next Steps

Limitations

- SageMaker has limited built-in algorithm
- Quicksight visualization options are limited
- Limited computational resources (e.g training instance type, enterprise package)

Next Steps

- Apply SDK to implement advanced prediction model
- If needed, connect to Power BI or other software for real time visualization