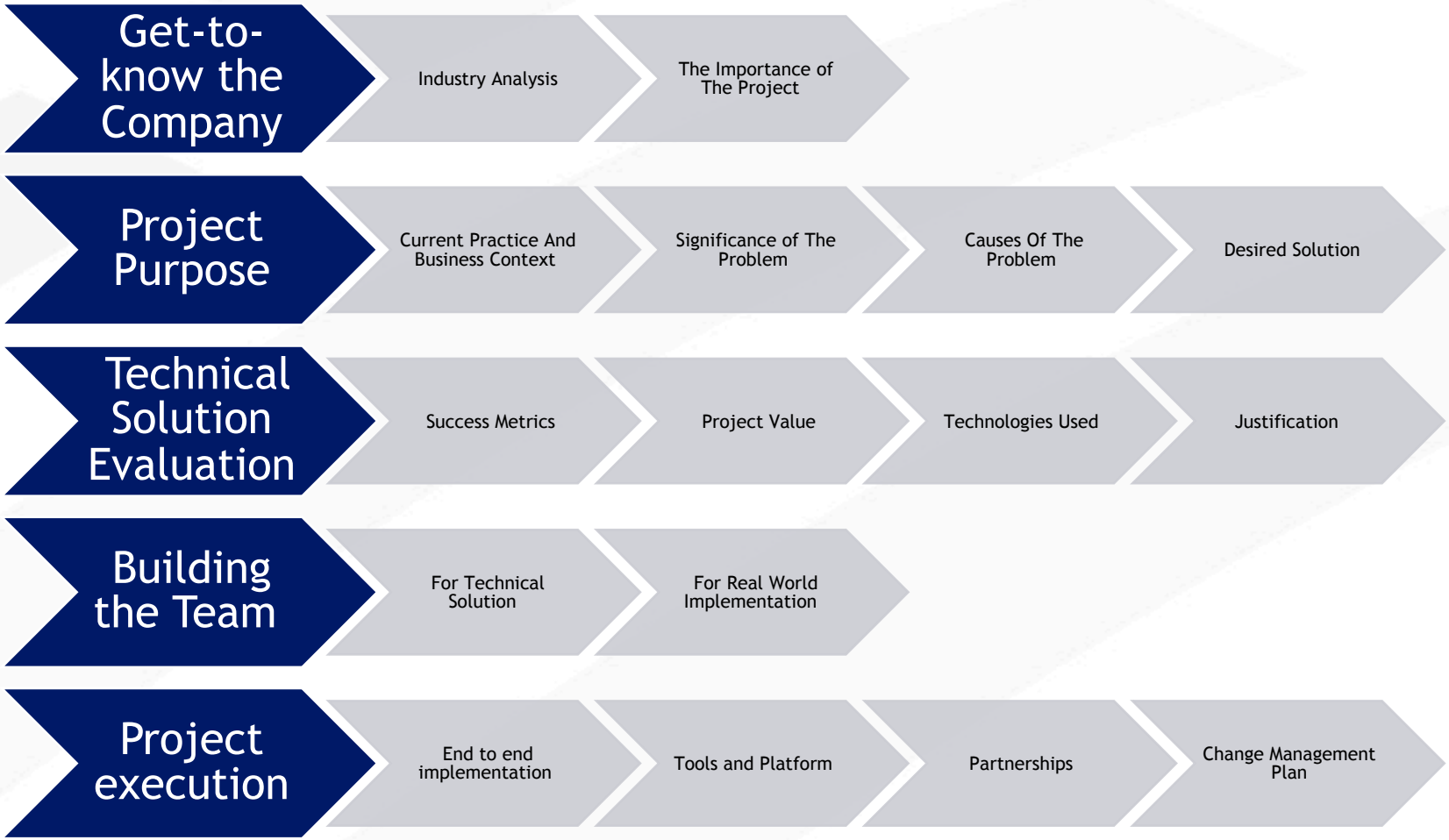
BANK OF AMERICA

# Credit Card Fraudulent Detection


Kartik Garg  Neo Liu

Zhiyi Zhao  Michelle Tan

# Agenda




BANK OF AMERICA

# Introduction Of The Industry

## The Banking Industry Overview

**What is the industry**

- Banks: systems of financial institutions

**What does the industry do**

- Offer clients the opportunity to open accounts for saving or investing purposes
- Provides resources for transactions and investments.

## Current trends in the industry

**Online Banking**

- Digital banking channels and services such as chatbots and mobile banking apps

**Mobile Banking**

- Access bank accounts using mobile devices.

**Statistics**

- In 2018: 61% V.S. In 2021: 65.3%
- **Increasing rate of more than 1 % per year**

BANK OF AMERICA

# Introduction of Bank of America

## History Overview

### Historical context

- Formed in 1998, San Francisco

### Current practice

- One of the largest banking and financial services corporations worldwide.

## Current Service

### Financial solutions and services.

- Individual consumers, small and middle-market businesses and large corporations

### Increasingly digital key features

- Including mobile check deposits and digital lending applications and online platforms

BANK OF AMERICA

# Company Financials - Five-year summary

| For the year | 2020 | 2019 | 2018 | 2017 | 2016 |
|---|---|---|---|---|---|
| Total Revenue | $ 85,528 | $ 91,244 | $ 91,020 | $ 87,126 | $ 83,498 |
| Net Income | 17,894 | 27,430 | 28,147 | 18,232 | 17,822 |
| Total assets | 2,683,122 | 2,405,830 | 2,325,246 | 2,268,633 | 2,190,218 |
| Total deposits | 1,632,998 | 1,380,326 | 1,314,941 | 1,269,796 | 1,222,561 |
| Earnings per common share | 1.88 | 2.77 | 2.64 | 1.63 | 1.57 |
| Total Shareholder Equity | 267,309 | 267,889 | 264,748 | 271,289 | 265,843 |

($ in millions, except per share information

BANK OF AMERICA

# Customer Capabilities And Online Segments

| Customer Capabilities | | |
|---|---|---|
| 66 million | $1.3T | $85.5B |
| consumer | In Consumer deposits (combined Consumer and GWIM businesses) | revenue generated by Consumer and small business clients |
| **Mobile Users Capabilities** | | |
| 44 Million (67%) | $0.65T (50%) | $36B (42%) |
| Mobile Users | of the payments made by consumers and small businesses using their credit cards. | Accomplished Online |

# Culture, Mission and Values

BANK OF AMERICA

# Credit Card Fraudulent Has Become a Trend

## Proliferation of Digital Banking

- Combined credit/debit card spend up $35 billion
- 2021 Combined credit / debit card spending of $201B, increased by 21% from 3Q20, where credit card volume went up by 26%

## Credit Card Fraud Caused Tremendous Loss for the Clients

- Mobile app fraud transactions have increased by over 600% since 2015

## An innovative Solution to Detect Fraud is Paramount

- An innovative Solution to Detect Fraud is Paramount

BANK OF AMERICA

# Cause Of Online Fraudulence

## Easy steps of access to stolen credit cards

- The personal identifiable information is easily stolen and sold (account takeover)
- Monetize from the account by transferring funds to other accounts or purchasing items online

## Current fraud processing methods are too late and hard to execute

- An investigation often crosses states/nations
- Evidence can be difficult to capture
- Ecommerce fraud may be perceived a low-priority crime

BANK OF AMERICA

# Value Estimation

**Current State**

- Revenue - **$22.8B**
- Cost - **$14.4B**
- Current fraud detection algorithm - **60%**

**Future State**

- Increase fraud detection precision to **75%**

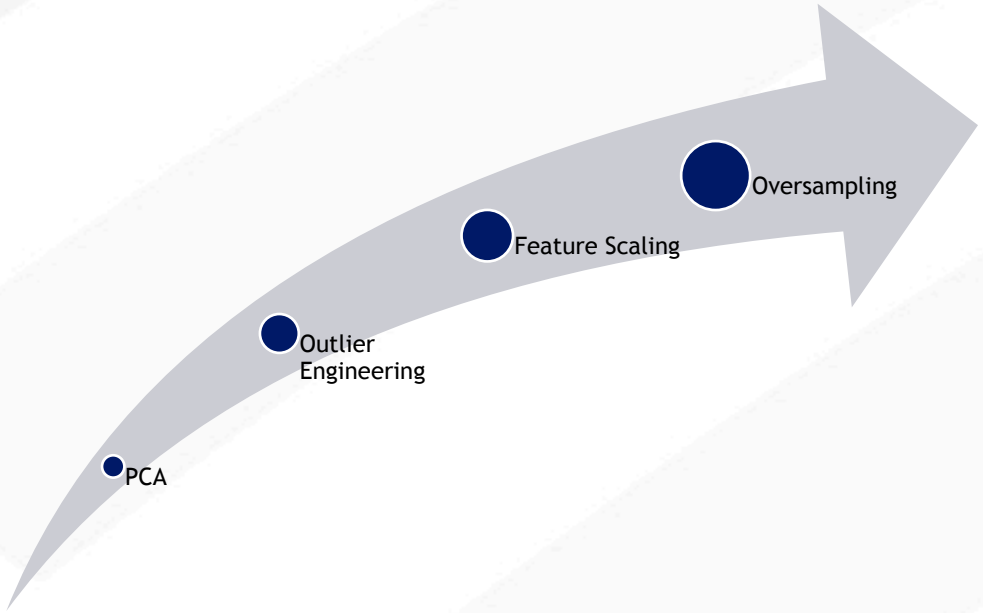
**Gaps**

- **Increase revenue**
  - Reinvesting **$3.4B** to grow market share
  - **Value $3.4B**
- **Decrease costs**
  - Current Cost – **$14.4B**
  - At **75%** efficiency cost will be – **$13.2B**
  - **Value $2.2B**
- **Execute Both**
  - **$1.1B Save cost**
  - **$1.1B reinvest**

BANK OF AMERICA

# Data Exploration

| | Time | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | ... | V21 | V22 | V23 | V24 | V25 | V26 | V27 | V28 | Amount | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0.0 | -1.359807 | -0.072781 | 2.536347 | 1.378155 | -0.338321 | 0.462388 | 0.239599 | 0.098698 | 0.363787 | ... | -0.018307 | 0.277838 | -0.110474 | 0.066928 | 0.128539 | -0.189115 | 0.133558 | -0.021053 | 149.62 | 0 |
| 1 | 0.0 | 1.191857 | 0.266151 | 0.166480 | 0.448154 | 0.060018 | -0.082361 | -0.078803 | 0.085102 | -0.255425 | ... | -0.225775 | -0.638672 | 0.101288 | -0.339846 | 0.167170 | 0.125895 | -0.008983 | 0.014724 | 2.69 | 0 |
| 2 | 1.0 | -1.358354 | -1.340163 | 1.773209 | 0.379780 | -0.503198 | 1.800499 | 0.791461 | 0.247676 | -1.514654 | ... | 0.247998 | 0.771679 | 0.909412 | -0.689281 | -0.327642 | -0.139097 | -0.055353 | -0.059752 | 378.66 | 0 |
| 3 | 1.0 | -0.966272 | -0.185226 | 1.792993 | -0.863291 | -0.010309 | 1.247203 | 0.237609 | 0.377436 | -1.387024 | ... | -0.108300 | 0.005274 | -0.190321 | -1.175575 | 0.647376 | -0.221929 | 0.062723 | 0.061458 | 123.50 | 0 |
| 4 | 2.0 | -1.158233 | 0.877737 | 1.548718 | 0.403034 | -0.407193 | 0.095921 | 0.592941 | -0.270533 | 0.817739 | ... | -0.009431 | 0.798278 | -0.137458 | 0.141267 | -0.206010 | 0.502292 | 0.219422 | 0.215153 | 69.99 | 0 |




- Principal components obtained with PCA due to confidentiality

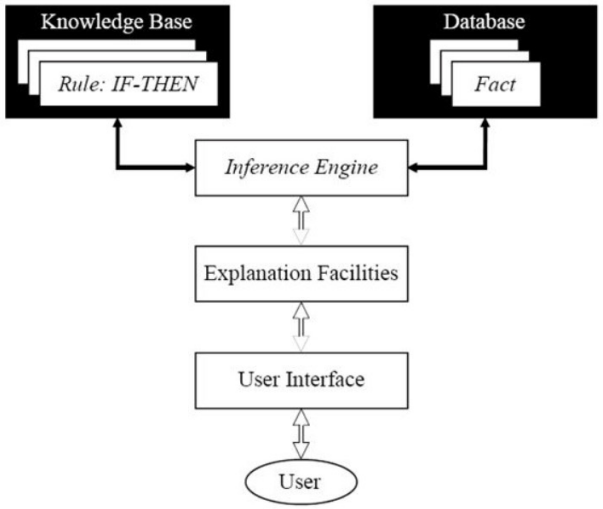
- Outlier Engineering – interquartile range
- Feature Scaling - Standardization

- Oversampling to achieve balanced dataset (SMOTE)

BANK OF AMERICA

# Models - Baseline models

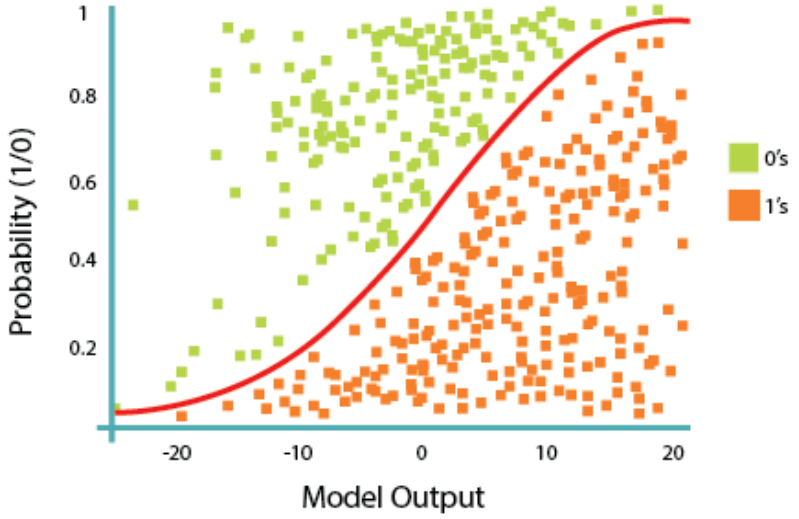
## Conventional Rule-based System




Not Machine Learning

- Made up of a set of rules
- Predictive Stage

## Logistic Regression



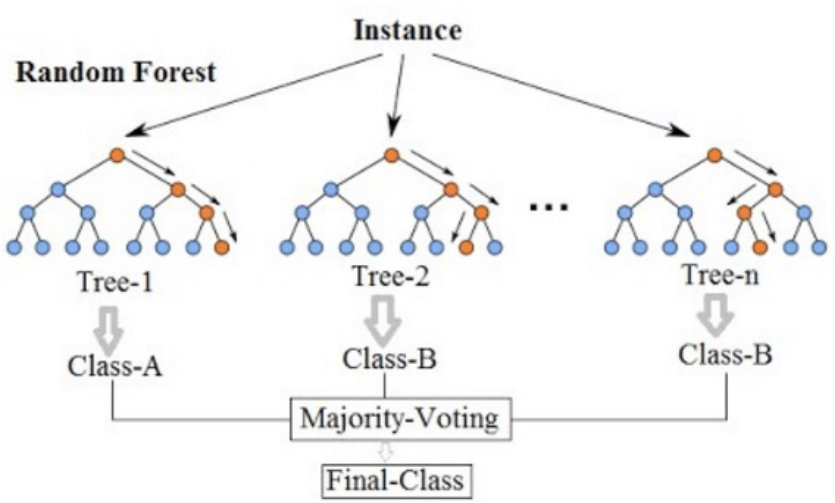

Supervised Machine Learning

- Modeling the probability of fraud
- Predictive Stage

BANK OF AMERICA

# Models - Decision Trees

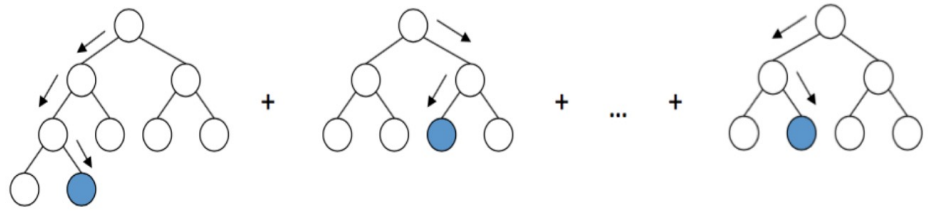
## Random Forest




Supervised Machine Learning

- An ensemble of decision trees
- Predictive Stage

## Light Gradient Boosting Machine (LightGBM)



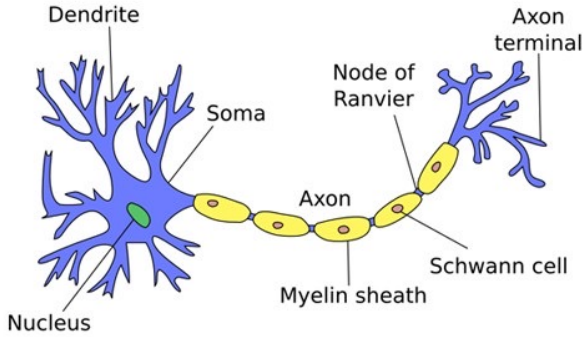
Supervised Machine Learning

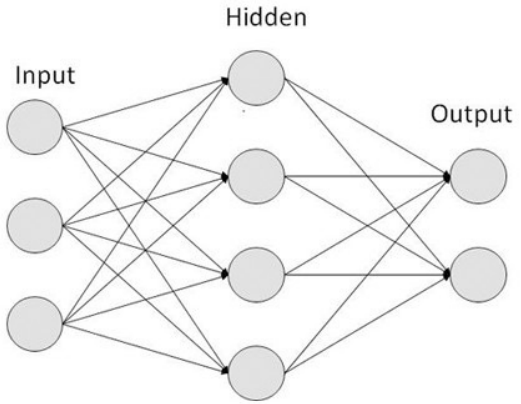
- Tree-based learning algorithms
- Predictive Stage

BANK OF AMERICA

# Models - Deep Learning

## Keras neural network model




Supervised Deep Learning

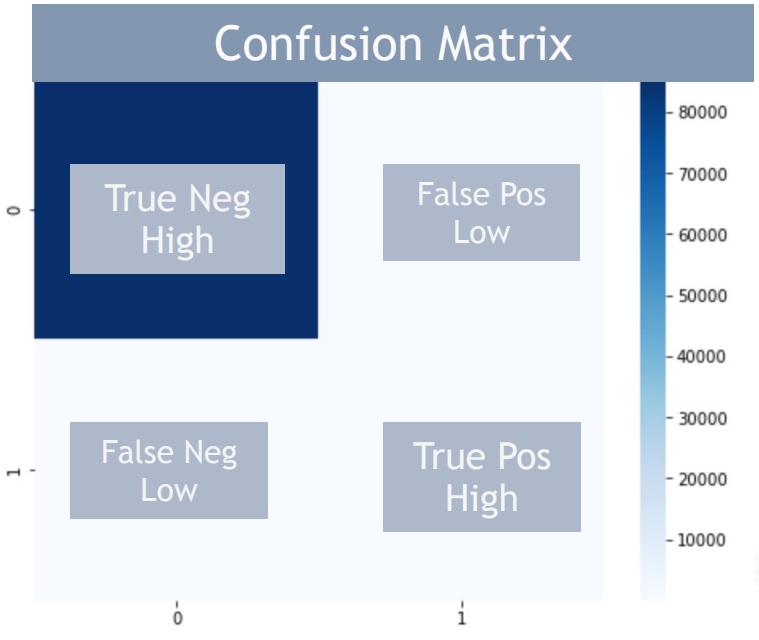
- A neural network is a simplified model of the way the human brain processes information
- Predictive Stage and Prescriptive Stage

BANK OF AMERICA

# Success Metrics

- Confusion Matrix
  - Keep the false positive rate low
  - Keep the false negative rate low

- Classification Report
  - Precision
  - Recall
  - F1 score




| Classification Report | | | | |
|---|---|---|---|---|
| | Precision | Recall | F1 Score | Support |
| Non-Fraud | | | | |
| Fraud | High | High | Target 0.75 | |

BANK OF AMERICA

# Model Selection & Justification

Conventional Rule-based System

Logistic Regression (with SMOTE )

Random Forest

LightGBM

Keras NN

| | Conventional Rule-based System | Logistic Regression | Random Forest | Light GBM | Keras NN |
|---|---|---|---|---|---|
| Recall (Sensitivity) | 12% | 62% | 69% | 82% | 79% |
| Precision | 35% | 88% | 94% | 75% | 89% |
| **F1 Score** | **18%** | **73%** | **79%** | **79%** | **84%** |

BANK OF AMERICA

# Updated Economic Value Estimation

Best performance for deep learning case : Keras NN

**Current State**

- Revenue - **$22.8B**
- Cost - **$14.4B**
- Current fraud detection algorithm - **60%**

**Future State**

- Increase fraud detection efficiency to **84%** ( instead of 75%)

**Gaps**

- **Increase revenue**
  - Reinvesting **$3.4B** to grow market share
  - **Value $3.4B -> 6.7B**
- **Decrease costs**
  - Current Cost – **$14.4B**
  - At 83% efficiency cost will be – **$12.05B**
  - **Value $2.35B -> 3.4B**
- **Execute Both**
  - **$1.7B Save cost ( instead of 1.1B)**
  - **$1.7B reinvest ( instead of 1.1B)**

BANK OF AMERICA

# Value Of Our Business Case

Improves both precision and recall value

+

Cancels fewer amount of non-fraud transactions

+

Detects fraud transactions with higher accuracy

=

Increasing Economic Value

Save Enormous Cost

Increase Retention Rate

Build Customer Royalty

BANK OF AMERICA

# Building the Right Team

Workflow


BANK OF AMERICA

# Building the Right Team

| Role | Headcount | Job Function |
|---|---|---|
| **Data Steward** | 1 | Data quality check |
| **Data Analyst** | 2 | Data acquisition, labeling, and dashboard creation |
| **Data Engineer** | 2 | Data warehouse/mart/ETL pipeline construction |
| **Data Scientist** | 4 | Data acquisition, cleaning, manipulation, and model train/test/tune, post-deployment monitor |
| **Data Science Lead** | 1 | Leads the fraud team to achieve the desired goal |
| **Software Engineer** | 2 | Fraud monitoring platform construction |
| **Domain Expert** | 1 | Provides insights on fraud life stages and formats |
| **Product Manager** | 1 | Creates product profiles and communicates with other stakeholders |

BANK OF AMERICA

# End to end project execution

# End To End Project Execution: Tools and Platform

## Data Lake

### AWS Simple Storage Service (S3)

- Amazon S3 is an object storage service that offers scalability, data availability, security, and performance.

## Model training

### Amazon Elastic MapReduce (EMR)

- It is a cloud big data platform for running large-scale distributed data processing jobs, interactive SQL queries, and machine learning applications

### Amazon SageMaker

- It is a cloud machine-learning platform that enables developers to create, train, and deploy machine-learning models in the cloud.

## Real time monitoring

### Amazon Kinesis data firehose

- It is an extract, transform, and load (ETL) service that reliably captures, transforms, and delivers streaming data to data lakes, data stores, and analytics services.

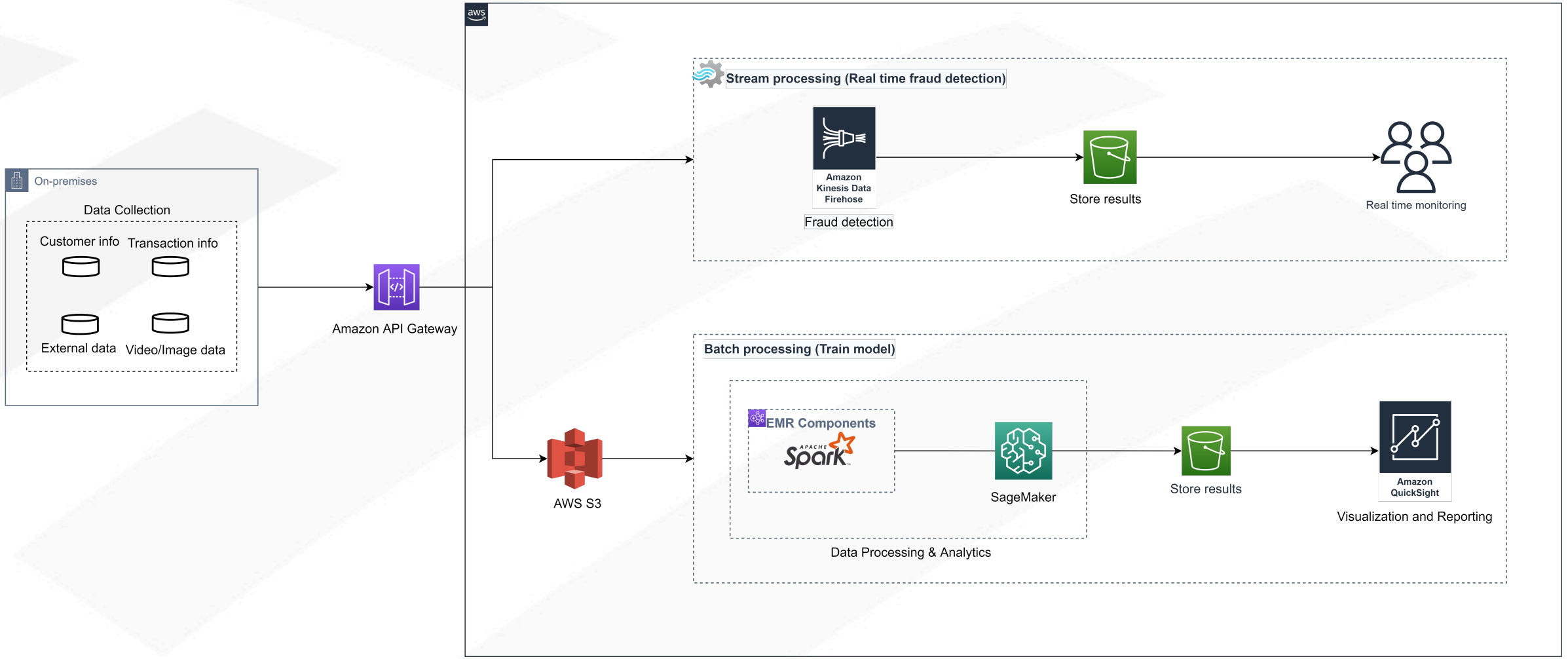
## Visualization

### Amazon QuickSight

- Amazon QuickSight is a machine learning-powered business intelligence service built for the cloud

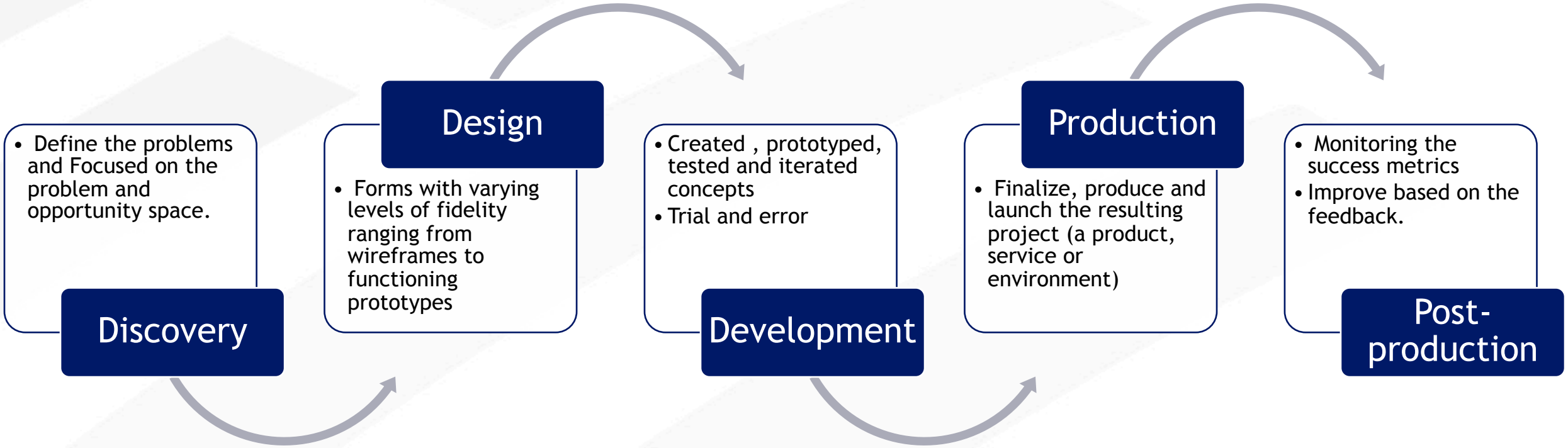
BANK OF AMERICA

# End To End Project Execution Architecture



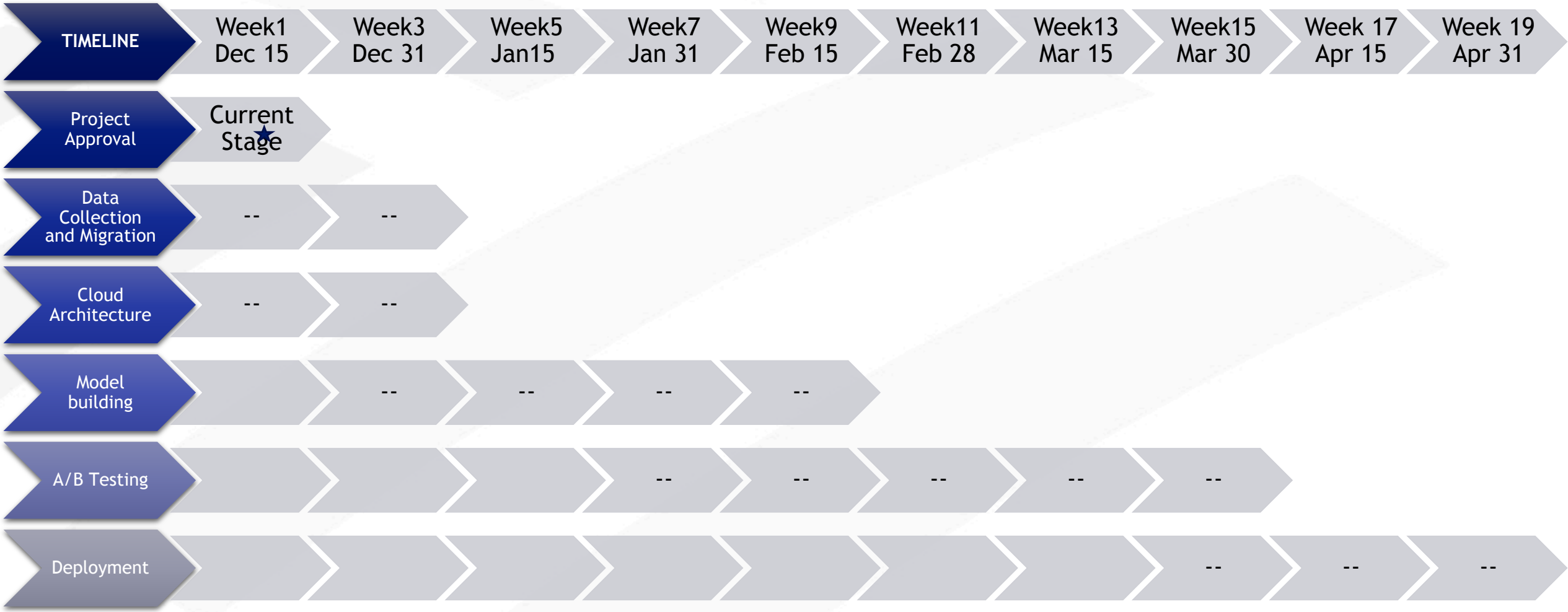

BANK OF AMERICA

# End To End Project Execution: Change Management Plan




BANK OF AMERICA

# End To End Project: Execution Timelines (Drive Adoption)




BANK OF AMERICA

# Thanks for listening!

## References


- "Overview Random Forest." Gaussian37, 3 Oct. 2018, https://gaussian37.github.io/ml-concept-RandomForest/.
- Serengil, Sefik. "A Gentle Introduction to LightGBM for Applied Machine Learning." Sefik Ilkin Serengil, 13 Oct. 2018, https://sefiks.com/2018/10/13/a-gentle-introduction-to-lightgbm-for-applied-machine-learning/.
- Ghallou, Ismail. "Build Your Perceptron Neural Net from Scratch." Medium, 22 Nov. 2018, https://medium.com/@ismailghallou/build-your-perceptron-neural-net-from-scratch-e12b7be9d1ef.
- https://feedzai.com/blog/fraud-attack-li fecycle/
- https://feedzai.com/blog/account-takeover/


## Q&A

BANK OF AMERICA