Sprint 1-COMP 680

Determining the fraudulent return orders in reverse supply chain

CSUN | CALIFORNIA STATE UNIVERSITY NORTHRIDGE

Presented By

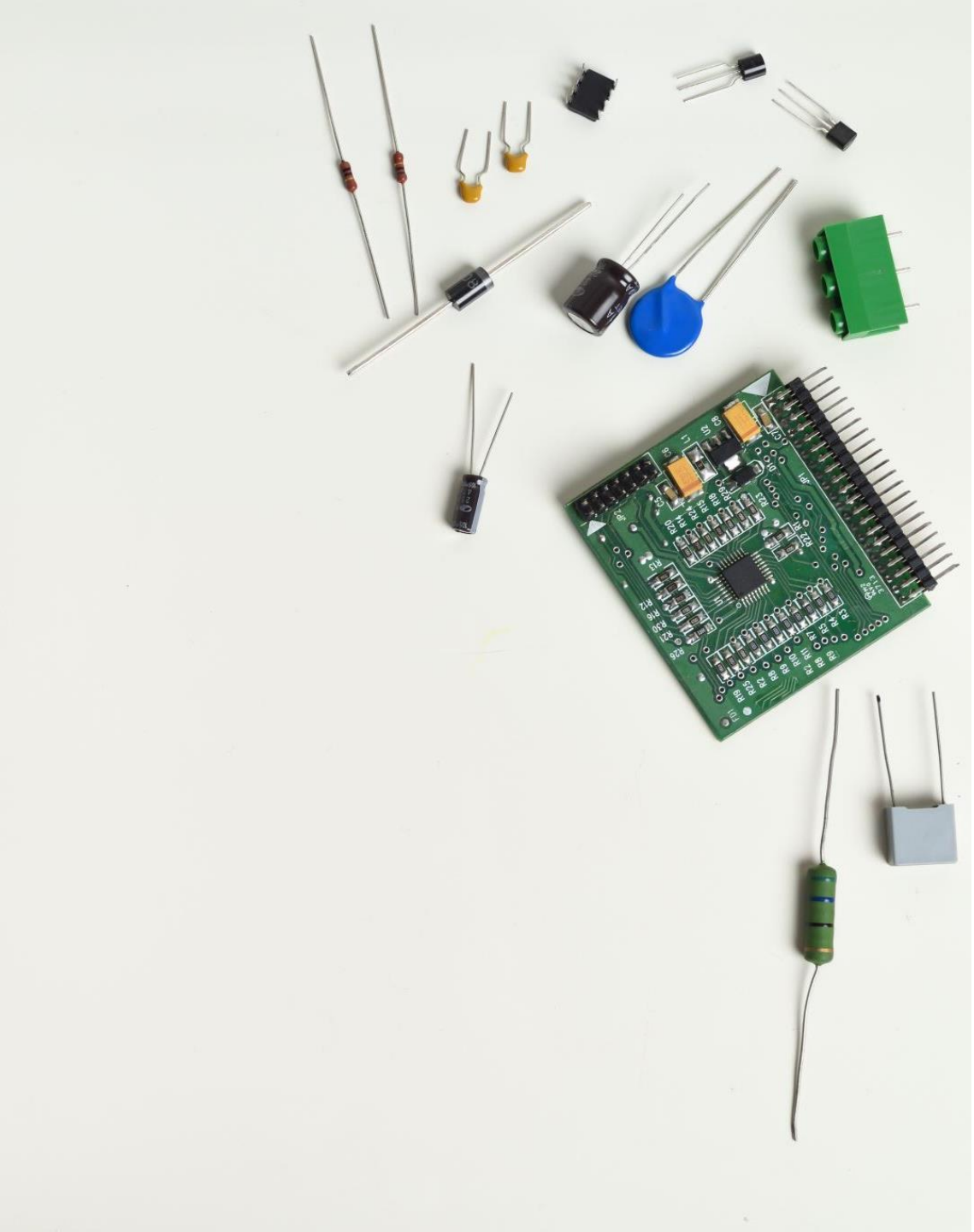
Venkat Sairam Ravala
Sowmya Kasu
Shamera Anjum Paleti
Linga Srikanth Rohan
Guna Yaswanth Gadde

Business Goal/Objective

- Online shopping has grown a lot, and they were offering many ways to buy and get orders. But, there's a issue with people returning or exchanging items in dishonest ways, which is costing online stores a lot of money.

- Right now, big retail companies are struggling to identify these fraud orders, which is a big problem. To stop fraud, it usually requires many workers and a lot of manual validations, which is difficult and time consuming.

- In this project, we want to study new technologies and implement them to figure out which orders might be fraud. This way, we can quickly refund honest customers and stop the fraud.



Team Responsibilities

- Venkat Sairam Ravala(Lead, Backend development,Integration)
- Sowmya Kasu (Regression algorithms)
- Shamera Anjum Paleti(Database management)
- Linga Srikanth Rohan(Backend development-Kafka)
- Guna Yaswanth Gadde (Backend development-kafka)

Sprint 1 Objective

- Research and Selection: Investigate current technologies and methodologies used for fraud detection in online retail, focusing on AI and machine learning and Data analysis algorithms . Select the most promising techniques based on effectiveness and implementation feasibility.

- Data Collection and Preparation: Gather historical transaction data from the company, including both legitimate and fraudulent return orders. Clean and preprocess the data for analysis, ensuring privacy and compliance with data protection regulations.

Sprint 1 Objective

- Preliminary Analysis: Conduct an exploratory data analysis to identify patterns and anomalies in the transaction data related to fraudulent activities.
- Model Development: Begin developing an initial model to predict fraudulent transactions based on historical data patterns.
- Technical Capabilities: Assess the current technical infrastructure and identify any gaps or needs for the implementation of new technologies.

API’s available for Fraud orders detection

- IPQualityScore Fraud Detection API: This API suite offers a comprehensive set of tools for real-time risk analysis and fraud prevention. It's designed to identify fraudulent users, suspicious payments, and abusive behavior across websites and applications. The service emphasizes its ability to block fraudulent payments, prevent fake account sign-ups, and detect bots and account takeover attempts with high accuracy .

- Amazon Fraud Detector: This fully managed service by AWS allows businesses to identify potentially fraudulent activities quickly. Utilizing machine learning and insights from Amazon's 20+ years of experience in handling fraud, the service can be used to flag suspicious online payments, new account fraud, and more, without requiring previous ML experience. It's particularly beneficial for businesses looking for a customizable fraud detection model that can be easily deployed and managed.

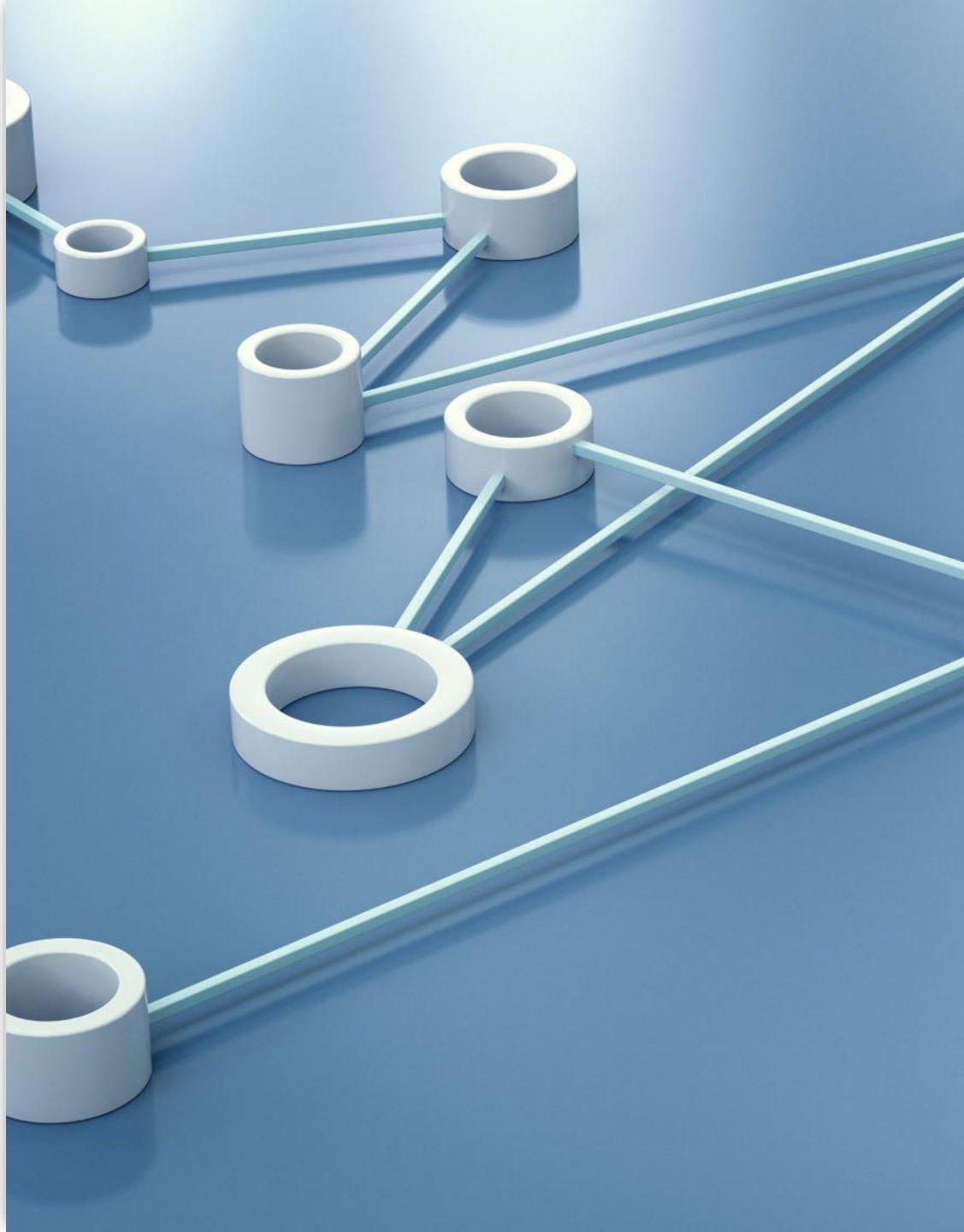
- Stripe: Besides being a payment processing platform, Stripe offers tools and techniques for e-commerce fraud prevention and detection. It leverages machine learning and AI to analyze transactions and identify potentially fraudulent activities. Stripe's approach includes multi-factor authentication, secure payment gateways, SSL certificates, and IP tracking, among others. These measures help in reducing false positives and enhancing the security of online transactions.

User Stories

- Creating a comprehensive project plan for Sprint 1, to have a clear understanding of the objectives, timelines, and their responsibilities.

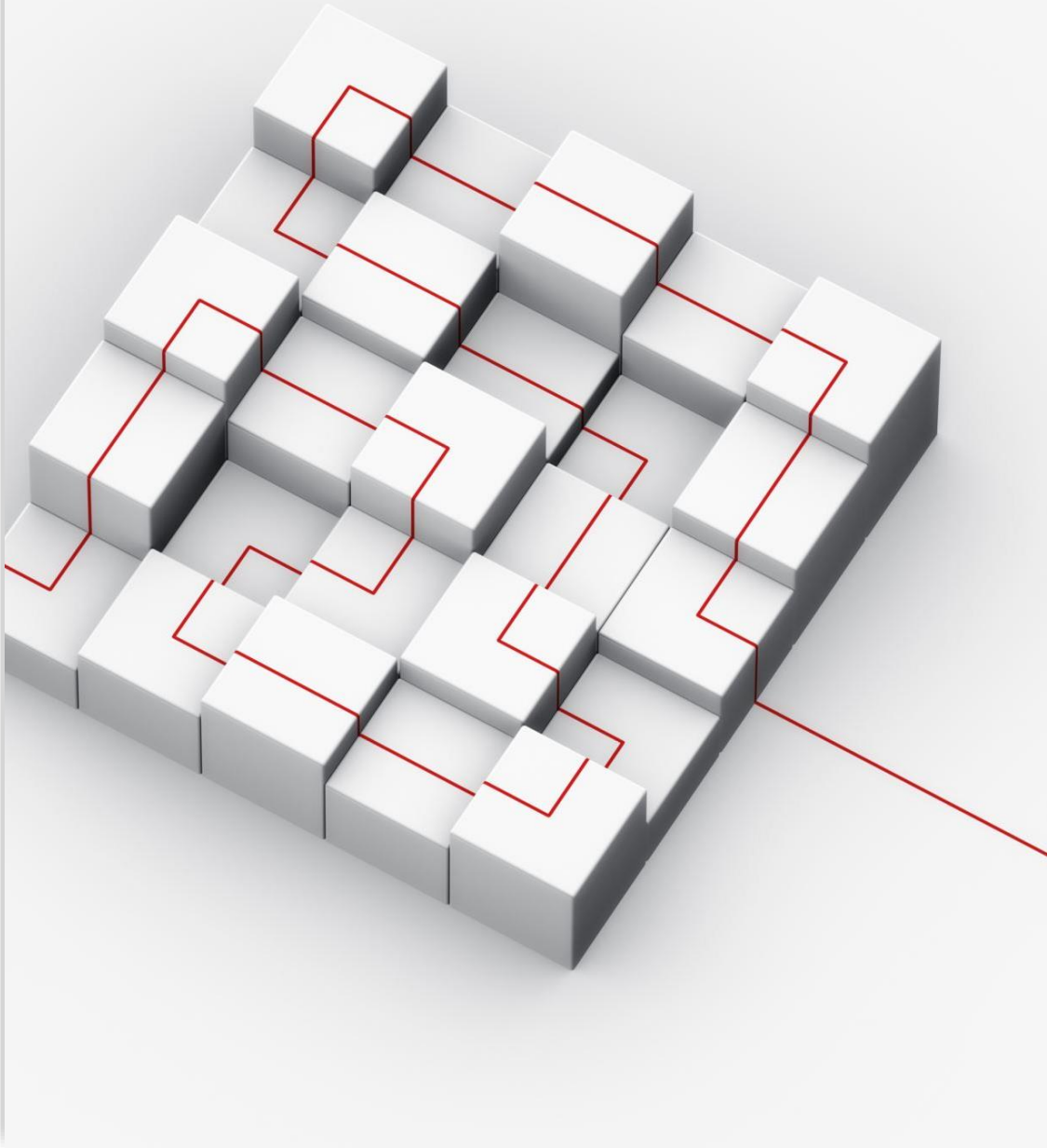
- Set up Kafka for real-time data streaming from the database to our fraud detection service, ensuring that we can process return order data efficiently for timely fraud detection.

- Conduct preliminary research on the Random Forest algorithm and other data analysis techniques, so that we can choose the most effective method for fraud order prediction based on our specific needs and data characteristics.



User Stories

- Gathering of a dataset for returned orders, marking them as fraud or not, so that we have a solid foundation for training our predictive model.
- Develop criteria for calculating the return score of orders, so that we can systematically assess the risk level of each return request.
- Design the initial architecture for integrating the fraud detection model with our current system, ensuring that it supports real-time processing and seamless data flow.
- Prepare the development environment and tools necessary for implementing the chosen data analysis algorithms, so that we can start the development process smoothly and efficiently.

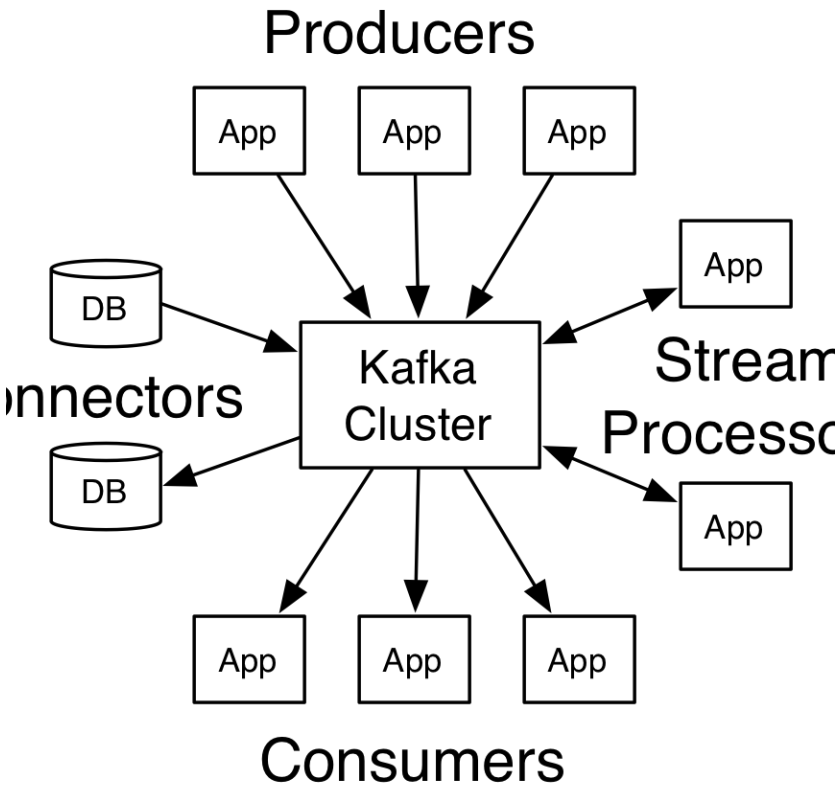


Technical Task Breakdown

- Data Collection and Preprocessing
- Kafka Integration for Real-Time Data Streaming
- Research on Random Forest Model implementation
- Return Score Calculation
- Analysis on project integration with existing system
- Programing language,Framework to be used
- Tools and database to be used

Kafka

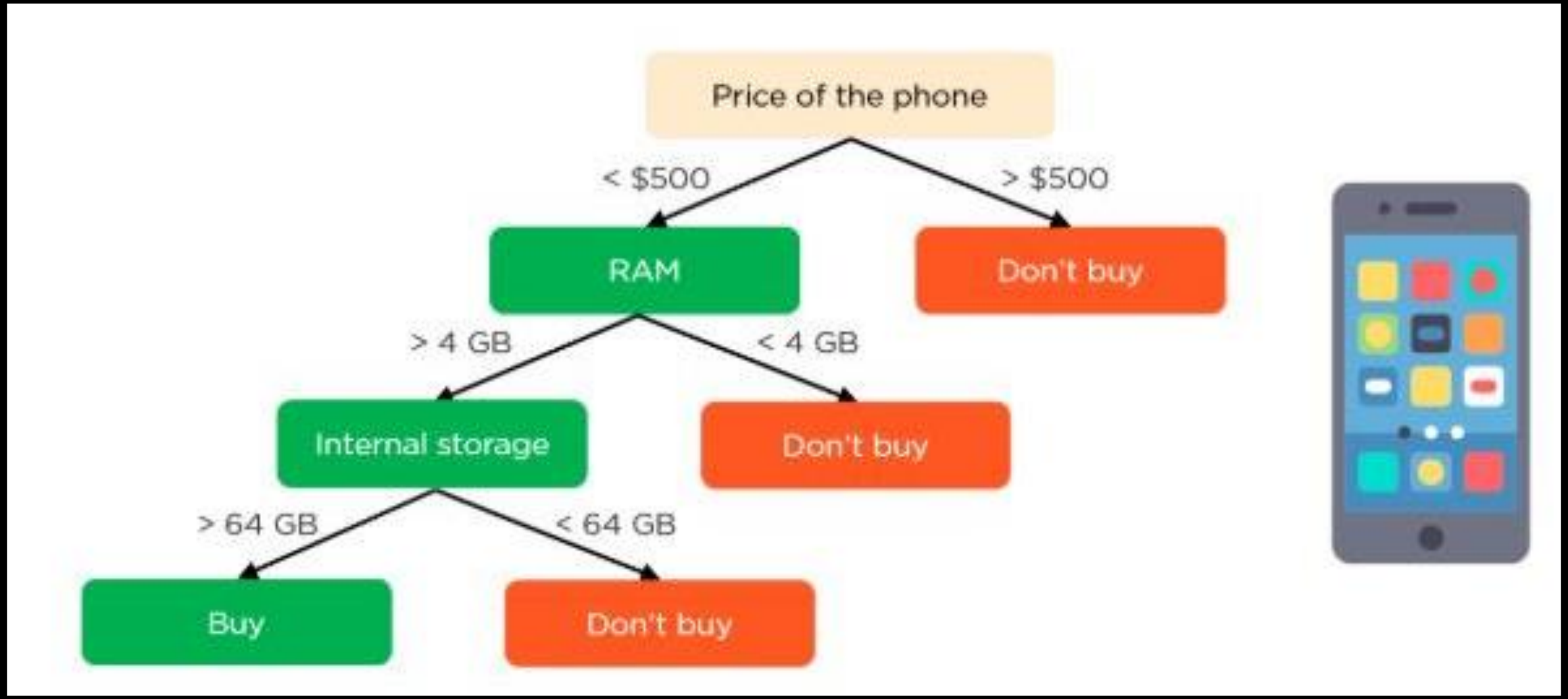
- Kafka is an opensource tool for handling real-time data.
- It helps separate the systems that create messages from those that read them, making it better at dealing with live data compared to traditional messaging systems like ActiveMQ.
- Kafka is reliable, can grow with your needs, and keeps data safe.
- It's great for analyzing data as it comes in, organizing logs, keeping track of website clicks, spotting fraud, and managing large amounts of data in real time.



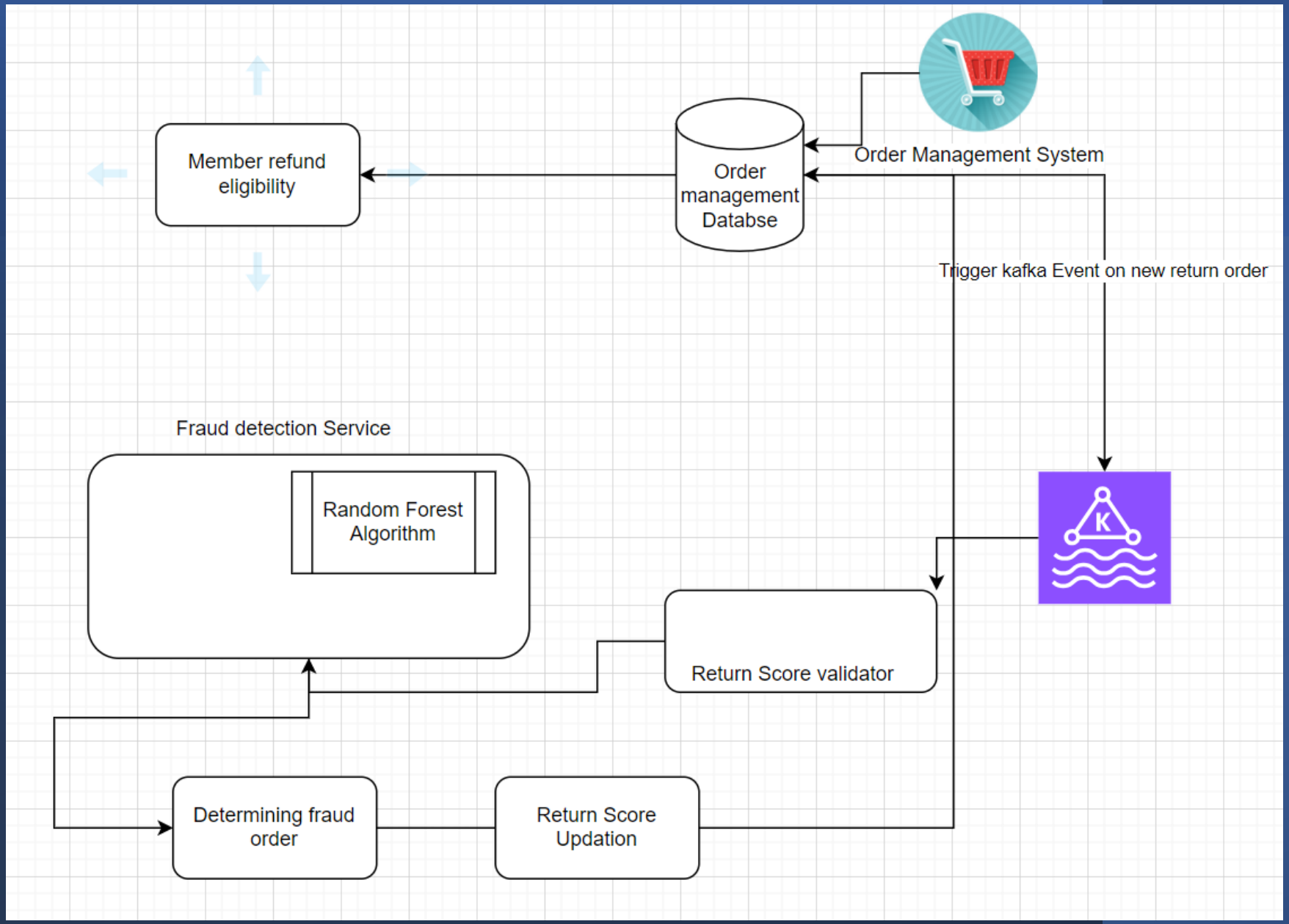
Random Forest Algorithm

- Random forest is a popular supervised machine learning algorithm—used for both classification and regression problems. It is based on the concept of ensemble learning, which enables users to combine multiple classifiers to solve a complex problem and to also improve the performance of the model.

- The random forest algorithm relies on multiple decision trees and accepts the results of the predictions from each tree. Based on the majority votes of predictions, it determines the final result.

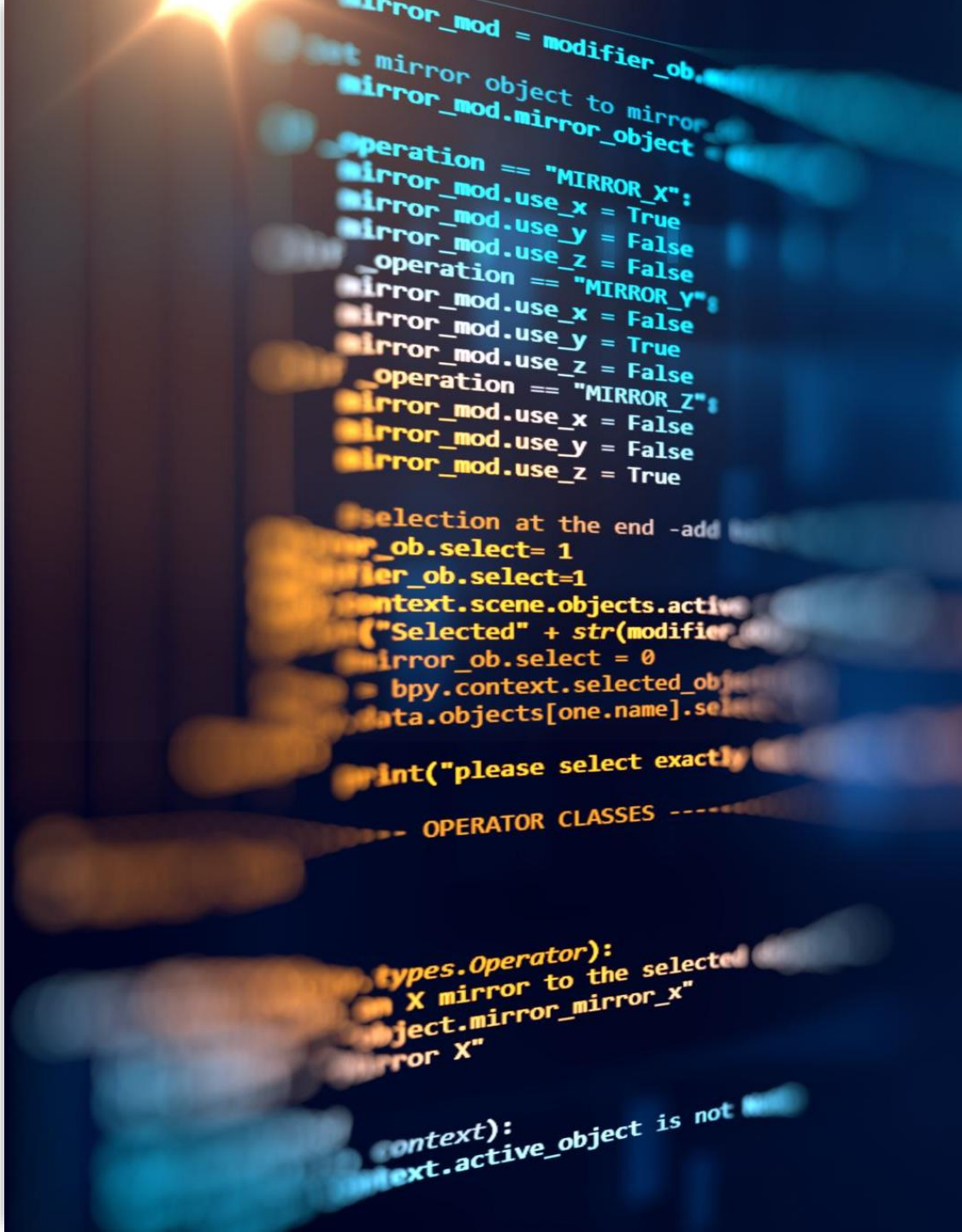


Random Forest Algorithm


Member refund eligibility
Order management Databse
Order Management System
Trigger kafka Event on new return order
Fraud detection Service
Random Forest Algorithm
Return Score validator
Determining fraud order
Return Score Updation

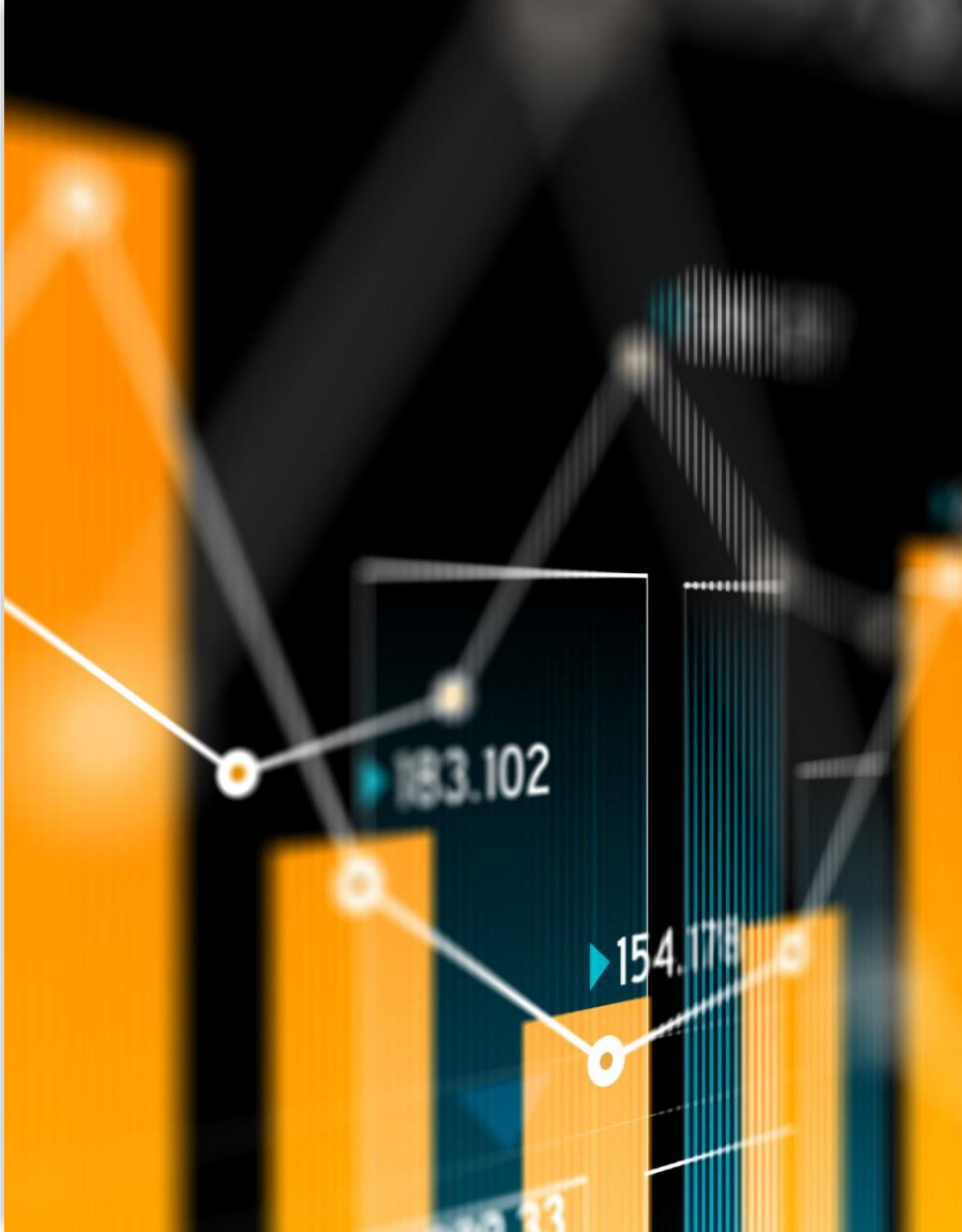
Implementation Approach

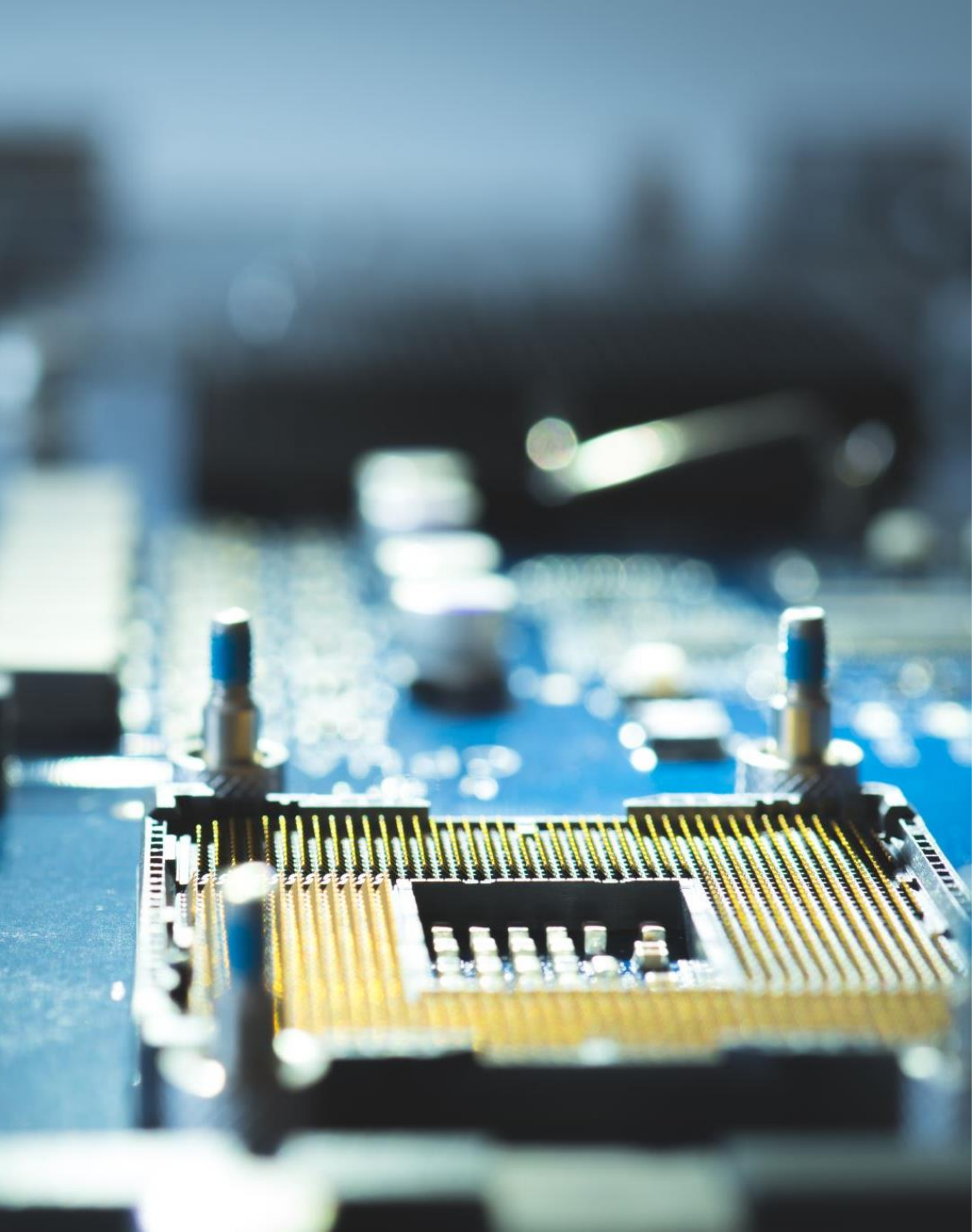
- Read data from database when ever new return request is raised.
- Analyse return is LVR or NON LVR
- If NON LVR- no action , it has to follow regular reverse supply chain process.
- If LVR, Check the return score for that customer.
- If return score not available ,mark the score to 100 and eligible for refund
- If return score <50 , follow regular reverse supply chain process and manual fraud analysis will be done.



Implementation Approach

- If return score not available ,mark the score to 100 and trigger fraud detection flow.
- If return score>50 , trigger fraud detection flow.
- Once fraud detection is done , if fraud reduce return score by (-10 points)





- Programing language to be used

  Java, R programming or Python programming
- Frameworks:
  Spring Boot
- Tools and database to be used

  R studio, Intellij, MySql Workbench

Dataset Sample

OrderID	CustomerID	ProductID	OrderDate	ReturnDate	ProductCategory	QuantityReturned	PurchaseAmount	ReasonForReturn	CustomerAccountAge	PreviousReturns	PreviousFraudReports	DeliveryType	ReturnCondition	RefundIssued	IsFraud
1	5d40a4d1-3819-47f6-b835-a3ef7224468e	69a9a051-47ba-4e72-9f60-c1a714fe8269	5/6/2022	########	Books	4	92.07	Not as described	340	0	0	Standard	Damaged	142.19	0
2	219c15b0-5ba5-4ead-a561-f58e45da841e	e4a58400-4f0a-46c3-82a5-4e63d9ad99b8	########	2/11/2024	Beauty & Health	4	61.24	No longer needed	899	12	1	Express	Opened	227.59	1
3	ecf9b807-d8e2-4bbd-980a-a287f47b45ce	1c4cf1dc-5ace-407d-897b-4c07d1c1acea	5/9/2022	11/7/2022	Home & Garden	4	321.85	Wrong item	915	13	0	Express	Opened	175.91	0
4	6a351023-68d6-40fa-aaa6-3278d3a9a224	f548cdf1-a681-4d58-98dc-3cfafcbd9586	4/21/2022	7/7/2022	Beauty & Health	4	356.17	No longer needed	183	12	0	Standard	Damaged	90.51	0

Sprint 1 Demo

Q&A