Team Name : Uncertainty Coders

Problem Statement : Real-Time High Scale Financial Fraud Risk Management

Brief about the Idea

Credit card fraud is a significant concern in the financial sector, with fraudsters constantly evolving their strategies to bypass conventional security measures. To combat this, a state-of-the-art solution has been developed that leverages an Artificial Neural Network (ANN) combined with Kafka, a distributed event streaming platform, to create a robust and adaptive credit card fraud detection system.

Opportunity : How different is it from any other existing ideas out there?

Real-Time Adaptation with ANN	Kafka Integration for Real-Time Streaming	Comprehensive Evaluation Metrics	Continual Monitoring and Response	Focus on Emerging Threats
Traditional rule-based systems and even some machine learning models struggle to keep up with rapidly evolving fraud tactics. The ANN's ability to learn from real-time data ensures that it can detect novel and evolving fraud patterns.	The integration of Kafka into the system architecture enables real-time data streaming and analysis. The utilization of Kafka ensures that the system can respond to incoming transactions instantaneously, enhancing its ability to catch fraudulent activities as they occur.	The system employs a range of evaluation metrics, including accuracy, precision, recall, F1 score, and ROC AUC score. This comprehensive evaluation provides a holistic view of the model's performance compared to systems that might rely on a single metric.	The integration of Kafka and the ANN creates a system that can operate 24/7, continuously monitoring transactions and making real-time predictions. Traditional systems might operate on a periodic basis, leaving windows of vulnerability between updates.	By emphasizing the need to detect evolving threats, the system addresses a critical weakness in some existing methods. Fraud tactics change rapidly, and a solution that can adapt in real-time is better equipped to stay ahead of fraudsters.

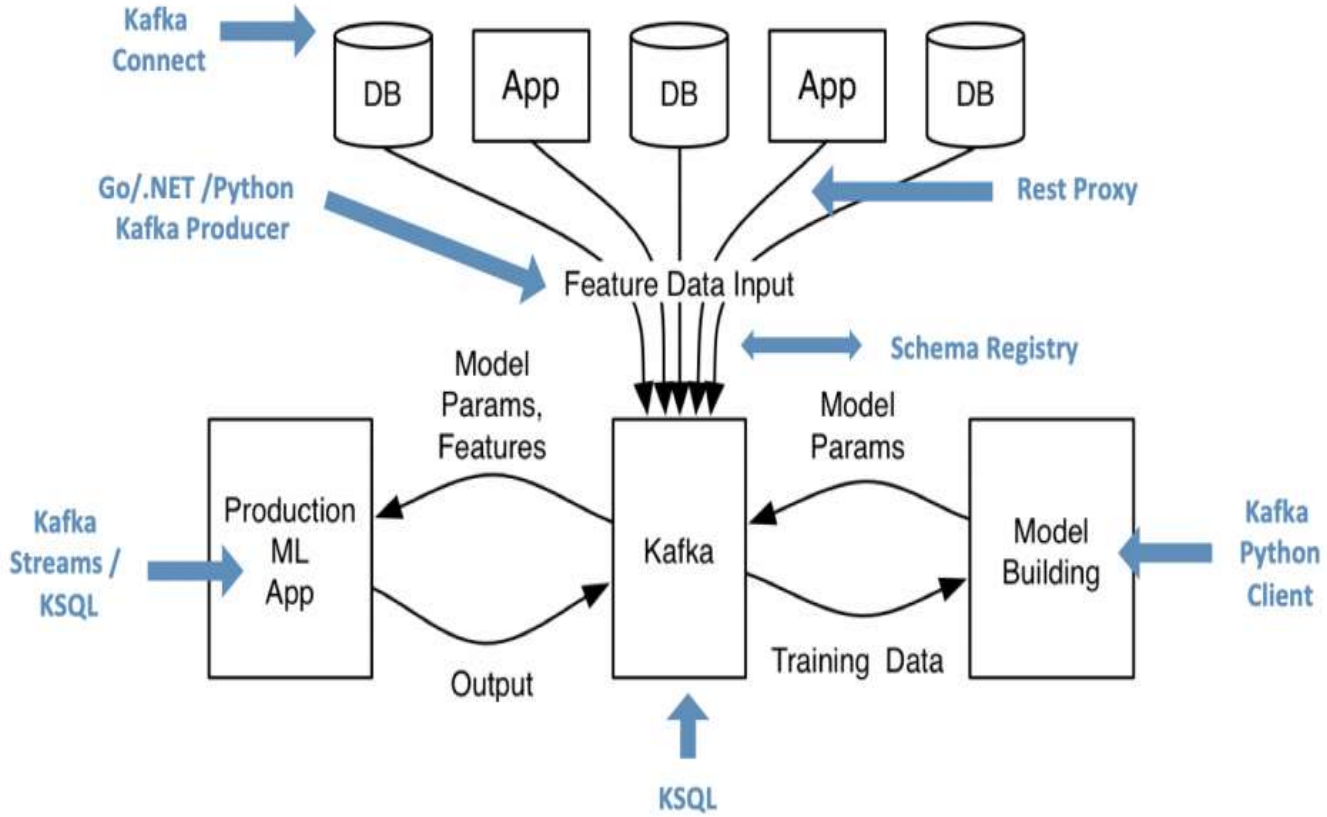
Opportunity : How will it be able to solve the problem?

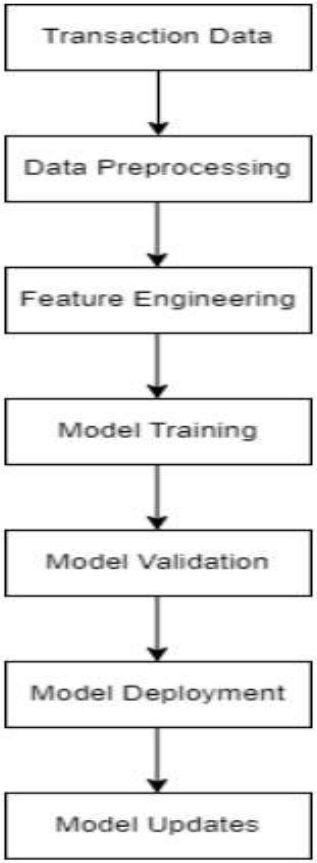
We have deployed highly scalable, open source data transaction streamer of Kafka. Kafka is used by various huge corporations like Netflix for enormous traffic channelization. Our model, converts the json data into a preprocessed format that our model further predicts upon. Our ann model shows a accuracy of 99.6 % and has been tuned to reduce amount of false positives.

Anomaly detectors are not as scalable and can only predict based on the patterns learned from training data. In the modern world fraud threats evolve over time and hence there is a need for models that learn based on incoming data aswell, learning new threats and evolving overtime, hence ANN models are most accurate for this application. The training data and ANN model is as close to real world as possible with consideration of identity. Kafka is open source , highly scalable real time data straming network, which has been utilised.

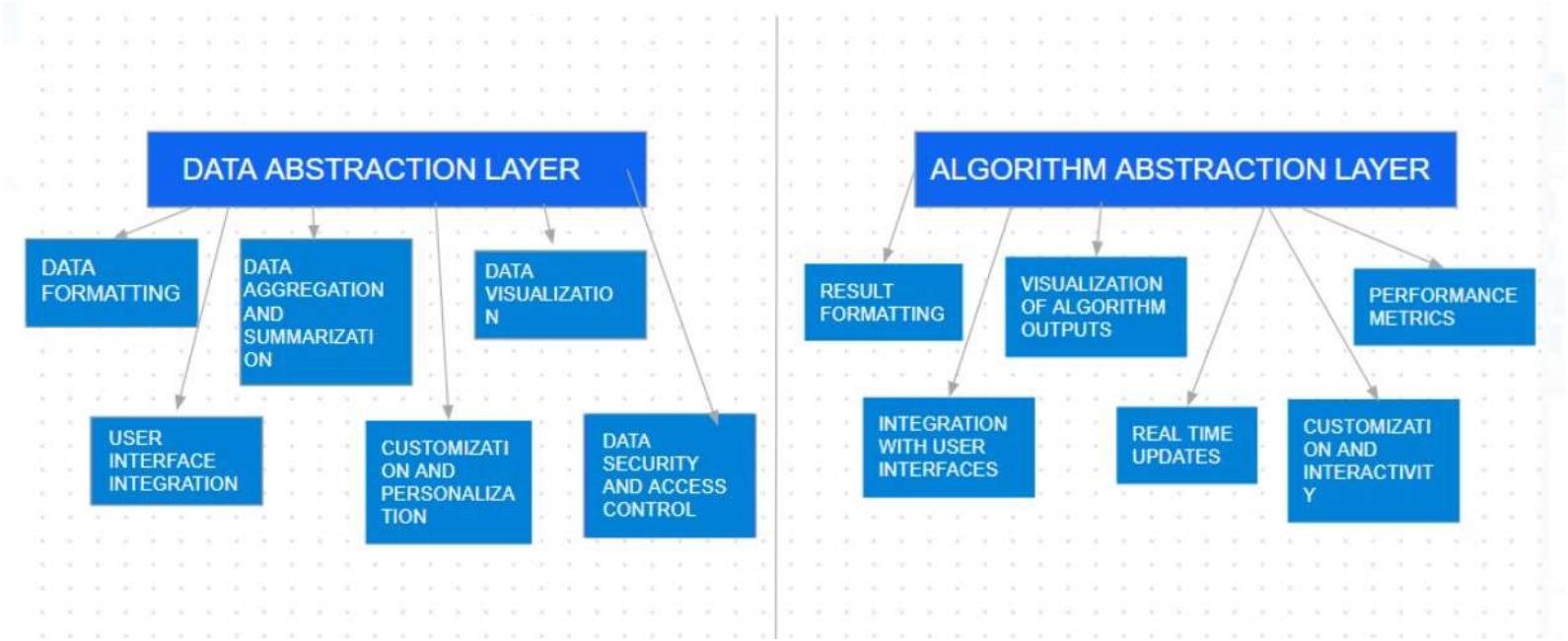
List of features offered by the solution :

It is always better to add a few visual representations (drawings/sketches/illustrations etc.) to your presentation, it adds to the power through which it reaches the audience.

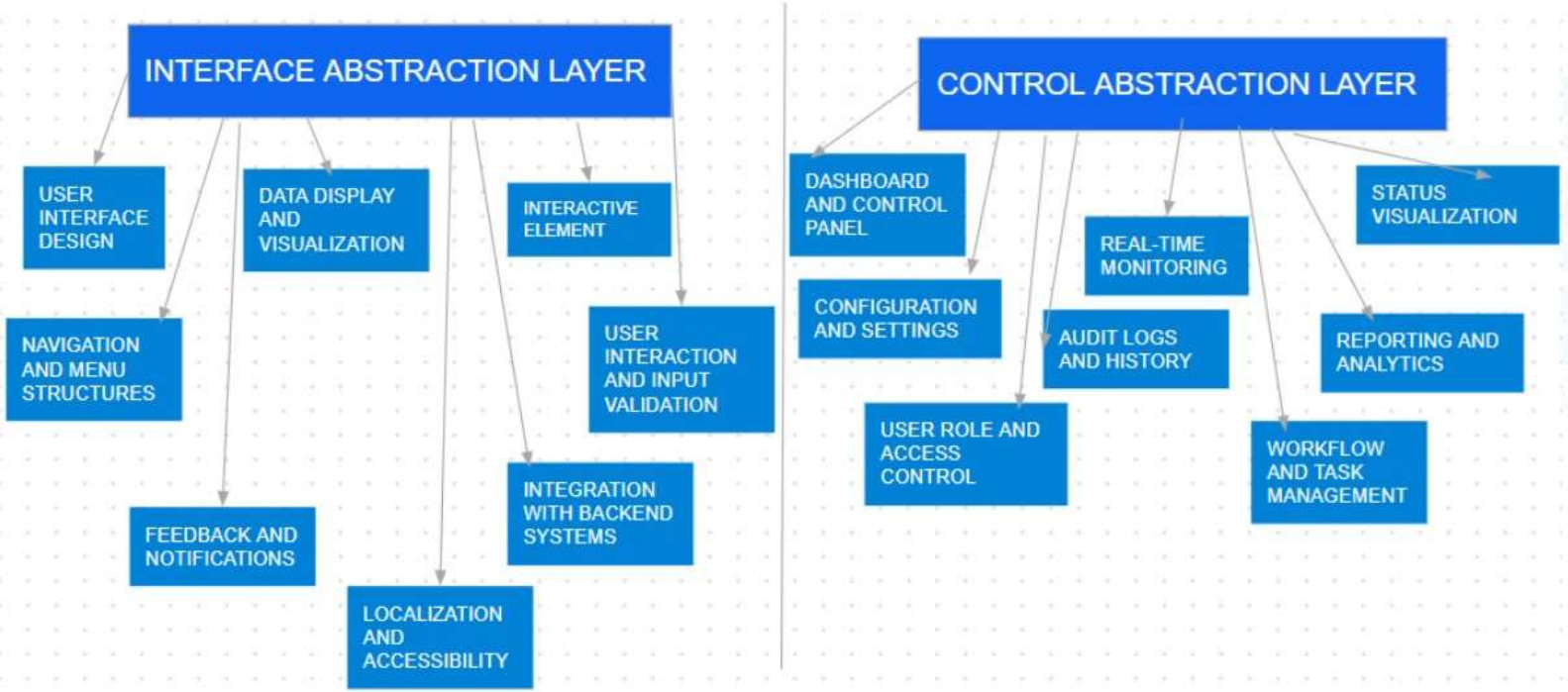

Kafka Connect
DB
App
DB
App
DB
Go/.NET /Python Kafka Producer
Rest Proxy
Feature Data Input
Schema Registry
Model Params, Features
Model Params
Kafka Streams / KSQL
Production ML App
Kafka
Model Building
Kafka Python Client
Output
Training Data
KSQL


Transaction Data
Data Preprocessing
Feature Engineering
Model Training
Model Validation
Model Deployment
Model Updates

Architecture



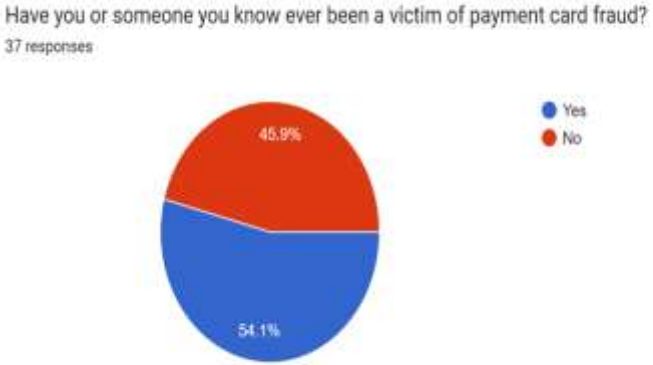
Architecture

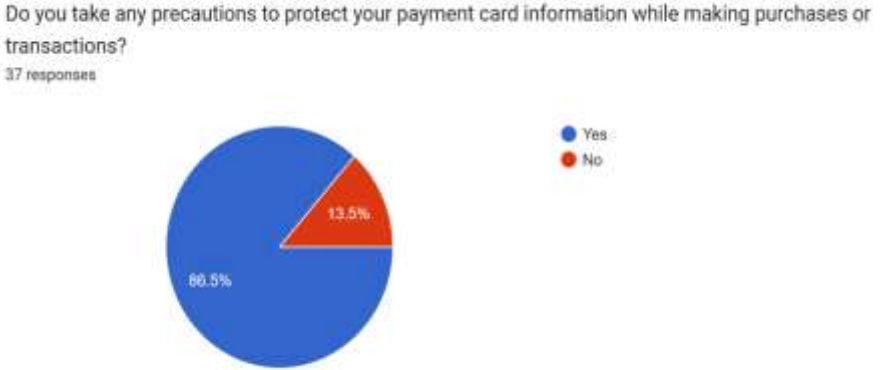


Business Logic of the solution :

We conducted a survey which resulted in the following-

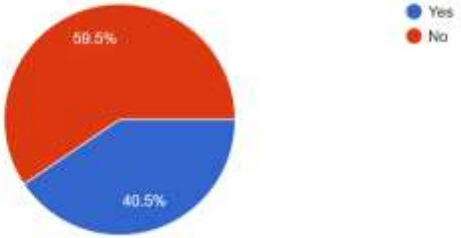
We conducted a survey which resulted in the following-





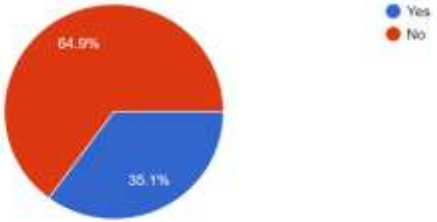
Have you ever reported a suspicious transaction or unauthorized charge on your payment card?

37 responses



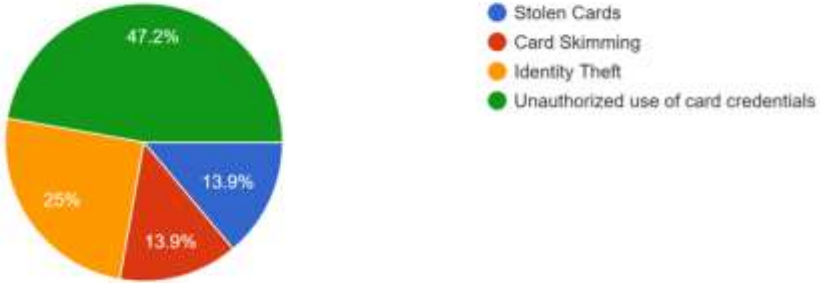
Have you ever encountered a situation where your payment card was declined due to suspected fraud?

37 responses



According to you which mode is most frequently used for such frauds?

36 responses

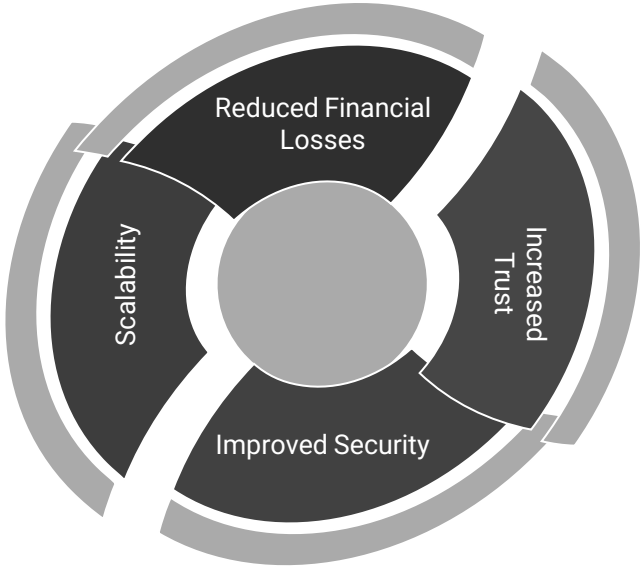


Business Logic of the solution :

We researched and concluded the following business logic of the solution:

Improved Security:
By detecting fraudulent transactions, the proposed solution will help improve the overall security of the payment card ecosystem. This will benefit both businesses and individuals.

Scalability:
The proposed solution can be scaled to other industries, such as the banking sector, where fraud detection is a critical issue. This will result in the adoption of efficient and accurate fraud detection solutions, which will lead to the prevention of financial loss and increased trust in the banking sector.



Reduced Financial Losses:
The adoption of an ML-based fraud detection solution can lead to a significant reduction in financial losses caused by payment card fraud. This, in turn, will benefit businesses and individuals.

Increased Trust in Payment Card Ecosystem:
With the adoption of an efficient fraud detection solution, users will have increased trust in the payment card ecosystem. This will lead to a more significant number of individuals adopting the use of payment cards and the online payment system, leading to an increased digital economy.

Business Logic of the solution :

We further researched and concluded the following:

Value Proposition:
Our real-time credit card fraud detection solution offers financial institutions a cutting-edge tool to proactively combat evolving fraud tactics. By integrating an adaptive Artificial Neural Network (ANN) with Kafka's real-time data streaming capabilities, we provide continuous monitoring and immediate response, enhancing fraud prevention accuracy and efficiency.

Customer Segments:
Financial Institutions: Banks, credit card companies, and other financial entities seeking advanced fraud detection solutions
Payment Processors: Companies handling credit card transactions on behalf of merchants, needing robust security measures.
Revenue Streams:

Subscription Model: Offer tiered subscription plans based on the volume of transactions and level of service, providing ongoing revenue.
Per-Transaction Fee: Charge a fee for each transaction analyzed by the system.
Customization and Integration: Charge for tailoring the solution to specific customer needs and integrating it into existing systems.
Key Activities:

Software Development: Continually enhance and optimize the ANN and system architecture.
Research and Learning: Keep the ANN up-to-date with emerging fraud patterns through continuous learning.
Customer Support: Provide technical assistance and guidance to customers.

Consultative Approach: Understand customer needs and tailor the solution to their specific fraud detection requirements.
Customer Support: Provide technical support and troubleshooting as needed.
Regular Updates: Keep customers informed about system upgrades and enhancements.
Cost Structure:

Research and Development: Invest in ongoing R&D to enhance the ANN's capabilities.
Data Storage and Processing: Cover costs associated with data storage and real-time processing.
Personnel: Employ data scientists, engineers, and customer support staff.
Partnerships:

Data Providers: Partner with data providers to access transaction data for analysis.
Cloud Service Providers: Collaborate with cloud providers for scalable and cost-effective infrastructure.
Consultants: Partner with industry consultants to offer holistic fraud prevention strategies.
Unique Selling Points:

Dynamic Adaptation: The solution evolves with emerging fraud tactics through continuous learning.
Immediate Response: Real-time analysis and response prevent fraud as it happens.
Comprehensive Monitoring: 24/7 vigilance ensures no fraudulent transaction goes unnoticed.
Customizable and Scalable: Tailored to customer needs and capable of handling varying transaction volumes.

Technology used :

Artificial Neural Network (ANN): The core of the solution is the ANN, a machine learning model inspired by the human brain's neural networks. The ANN processes and learns from data to make predictions, allowing it to adapt to evolving fraud patterns over time.

Kafka: Kafka is a distributed event streaming platform that facilitates real-time data ingestion and analysis. It allows the solution to process incoming credit card transaction data as it occurs, enabling immediate response and prevention.

Python: The solution is implemented using the Python programming language, which offers a wide range of libraries and tools for machine learning, data preprocessing, and integration with other technologies.

TensorFlow and Keras: TensorFlow, an open-source machine learning framework, and Keras, a high-level neural networks API, are used to develop, train, and evaluate the ANN model efficiently.

Pandas: The Pandas library is utilized for data manipulation and analysis, enabling efficient preprocessing and feature engineering.

Technology used :

Scikit-learn: Scikit-learn provides various tools for data preprocessing, feature scaling, and evaluation, enhancing the model's performance.

Cloud Infrastructure: Cloud computing resources are employed for scalable and cost-effective data storage, processing, and deployment of the solution.

Data Streaming Architecture: Technologies for real-time data streaming architecture are incorporated to ensure seamless integration of Kafka and the ANN, allowing continuous processing of incoming transactions.

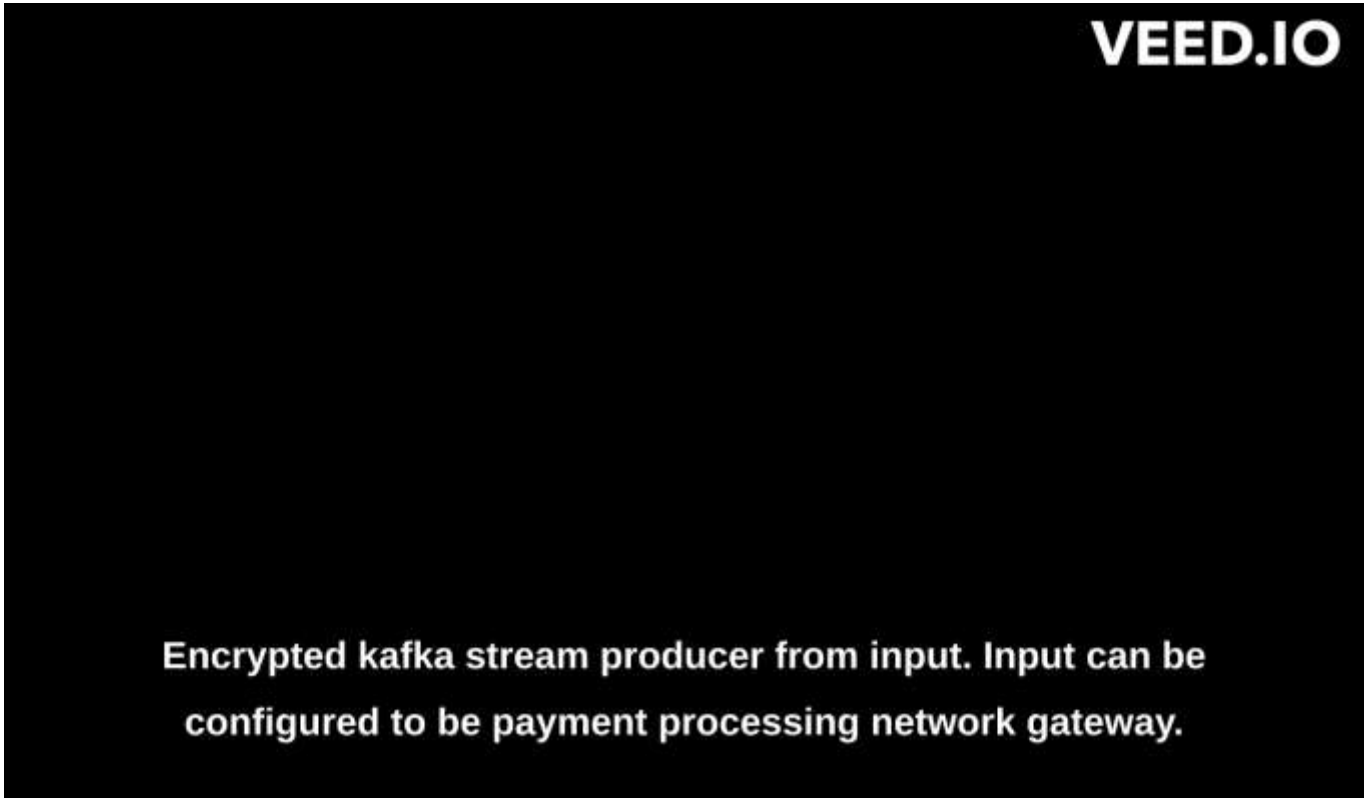
Visualization Libraries: Libraries such as Matplotlib and Seaborn are used for creating data visualizations to analyze and understand the processed data.

Label Encoding and Standardization: Label encoding and standardization techniques are employed from the Scikit-learn library to preprocess categorical variables and numerical features.

Estimated cost of/after implementing the solution :

The estimated cost of implementing the proposed credit card fraud detection solution would encompass expenses related to research and development efforts for optimizing the Adaptive Neural Network (ANN) and dynamic feature engineering, data storage and processing in cloud environments, cloud computing resources for hosting the solution and processing real-time transactions, personnel salaries for data scientists and engineers, data streaming infrastructure setup and maintenance, customization and integration efforts, security measures implementation, ongoing maintenance to ensure ANN effectiveness, and potential costs associated with external data providers for enrichment. The overall cost would be influenced by factors such as system complexity, data volume, technology choices, and team size.

Demo Video/Live Demo :-



GitHub Link (Codes should be public and available after hackathon also)

https://github.com/sharma-kshitij-ks/Pitch-to-SBI-Hackathon

SBI PCI FCC

GLOBAL FINTECH FEST

PITCH TO SBI HACKATHON

THANK YOU

Powered by H2S