FRAUDX

FRAUD X

SMART. SAFE. SECURE

gradccfd@gmail.com

Egyptian Russian University

ERU

الجامعة المصرية الروسية

FRAUD X

SMART.SAFE.SECURE

Optimizing Financial Security Risk Management Using Artificial Intelligence Techniques

gradccfd@gmail.com

Table Of Content

1. Introduction

1. Introduction

- Rise in e-commerce and digital payments post-COVID
- Global credit card fraud losses: $28.6B (2019) → $49.3B (2030)
- Traditional detection systems: high false positives & outdated rules

Our solution:

- AI techniques for addressing the challenge and enhancing predictive power

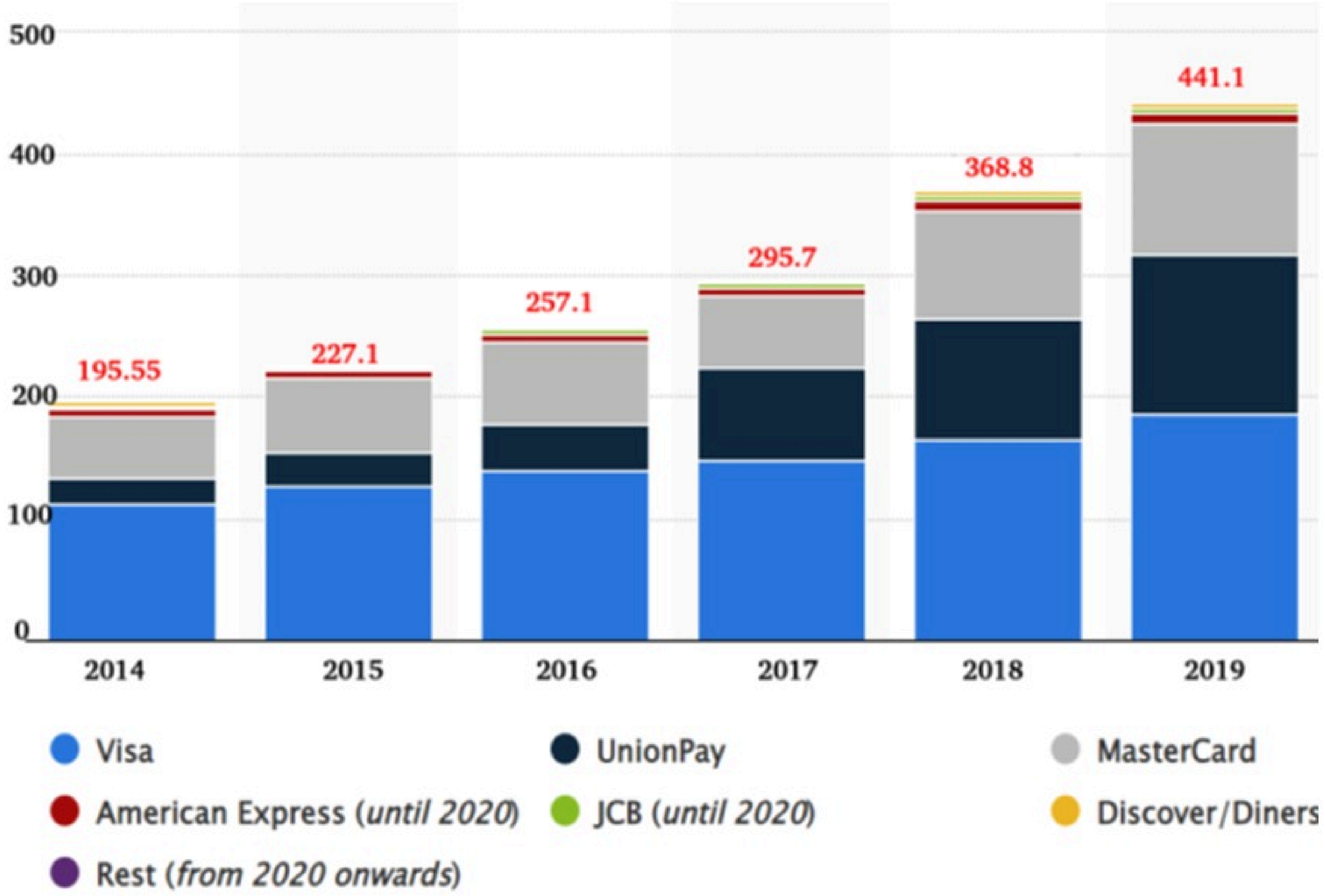
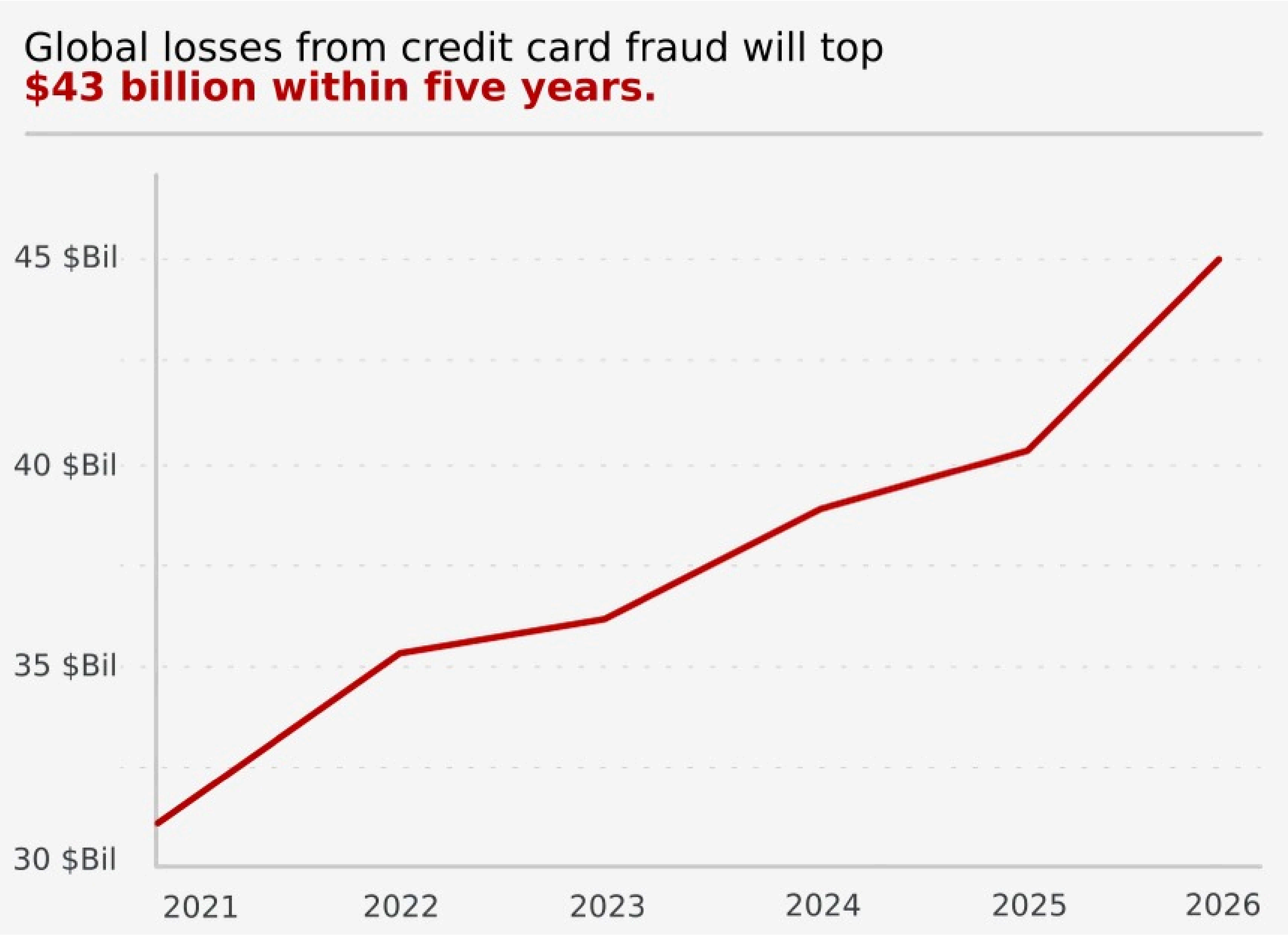


Figure 1 Number of transactions in billions.

2. PROBLEM

2. Problem

- Surge in digital transactions → increased fraud risk
- Challenges: credit card fraud, identity theft, unauthorized access
- Traditional systems can't keep up with fraud evolution
- Need for intelligent, real-time, adaptive fraud detection



3. Gaps

BIBLIOMETRIC ANALYSIS OF LITERATURE REVIEW

Papers

77

Filter by Year

2021 | 2022 | 2023 | 2024

Fig 1 : Papers by Year

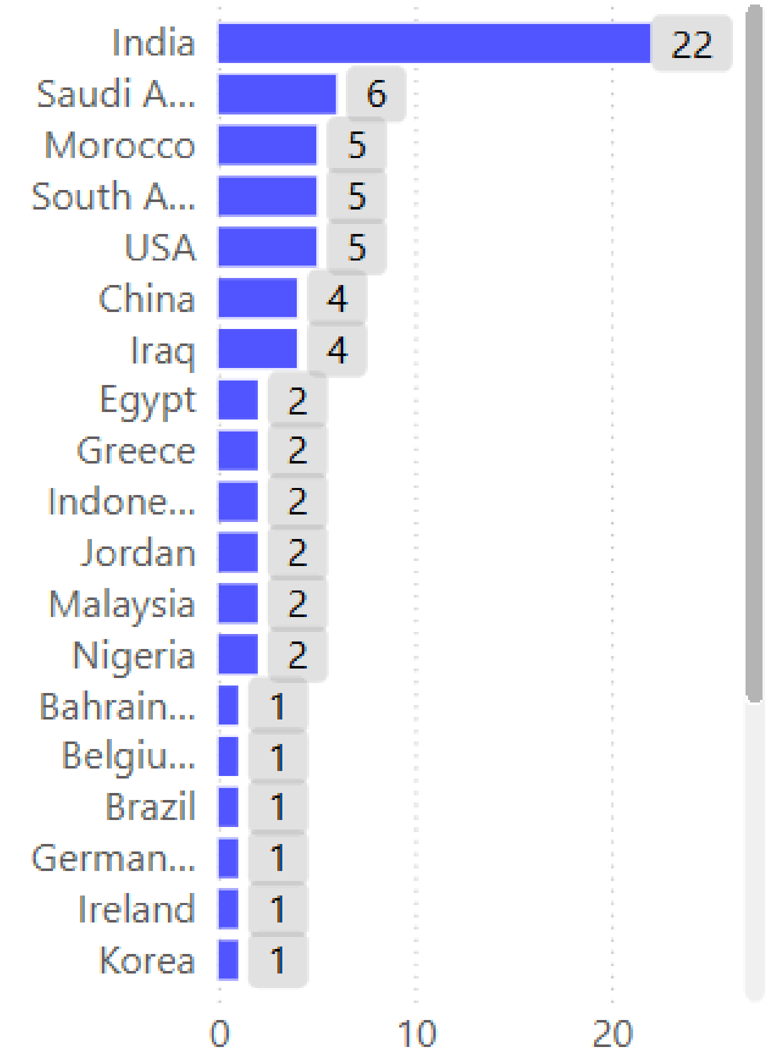


Fig 2 : Type of Papers

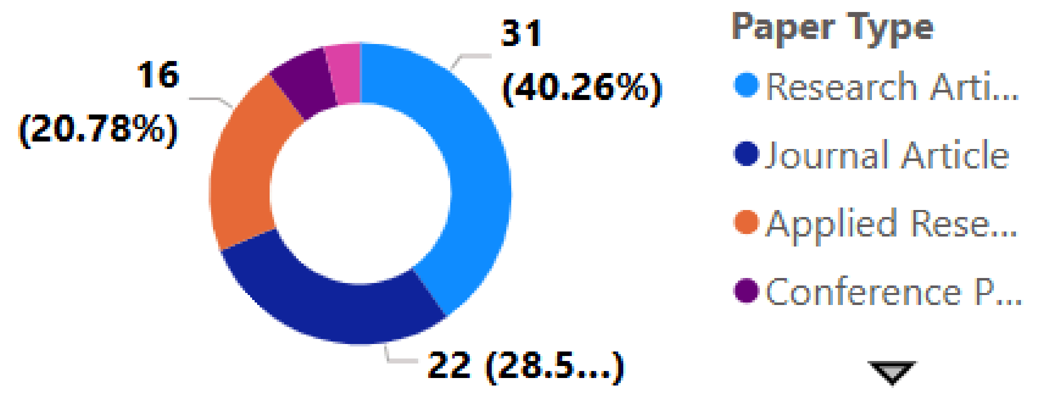
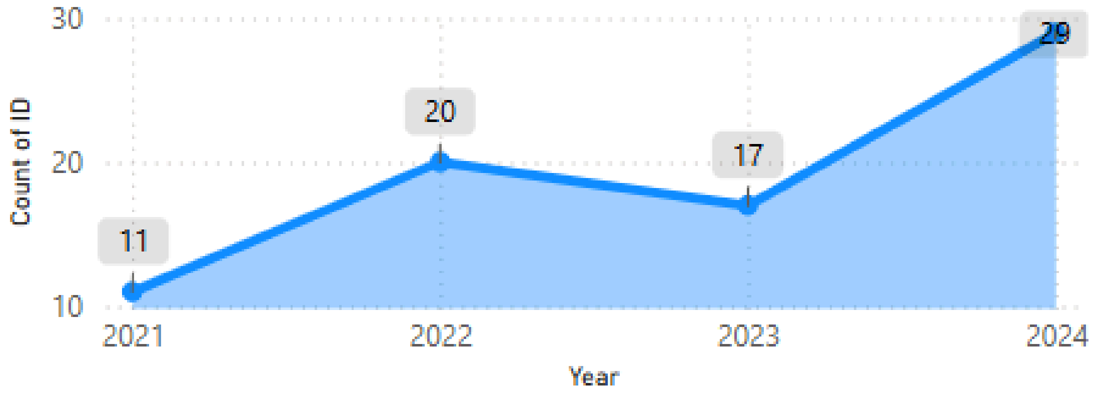
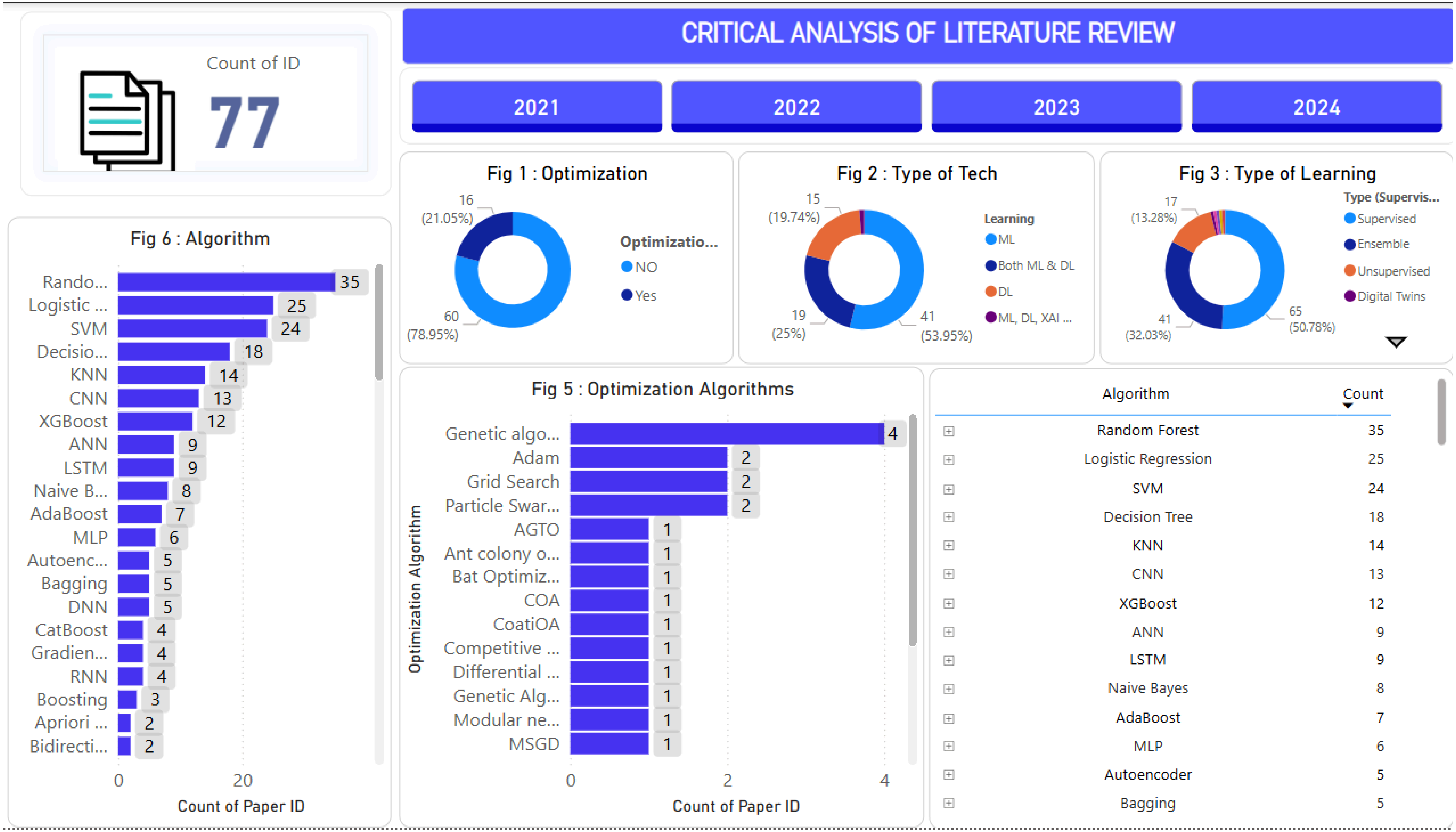


Fig 3 : Papers by Year



ID	Paper Type	Country	Journal
A13	Research Article	India	Procedia Computer Science
A14	Applied Research	Saudi Arabia	The International Arab Journal of Information Technology
A15	Applied Research	India	International Journal on Recent and Innovation Trends in Computing and Communication
A16	Applied Research	Nigeria	International Journal of Advanced Computer Science and Applications
A17	Applied Research	India	International Journal on Recent and Innovation Trends in Computing and Communication
A18	Applied Research	India	Computer Systems Science & Engineering
A19	Applied Research	South Africa	Applied Sciences
A20	Applied Research	South Africa	IEEE
A21	Research Article	Morocco	IEEE


CRITICAL ANALYSIS OF LITERATURE REVIEW
Count of ID
77
2021
2022
2023
2024
Fig 1 : Optimization
16 (21.05%)
60 (78.95%)
Optimizatio...
NO
Yes
Fig 2 : Type of Tech
15 (19.74%)
19 (25%)
41 (53.95%)
Learning
ML
Both ML & DL
DL
ML, DL, XAI ...
Fig 3 : Type of Learning
17 (13.28%)
41 (32.03%)
65 (50.78%)
Type (Supervis...
Supervised
Ensemble
Unsupervised
Digital Twins
Fig 6 : Algorithm
Rando... 35
Logistic ... 25
SVM 24
Decisio... 18
KNN 14
CNN 13
XGBoost 12
ANN 9
LSTM 9
Naive B... 8
AdaBoost 7
MLP 6
Autoenc... 5
Bagging 5
DNN 5
CatBoost 4
Gradien... 4
RNN 4
Boosting 3
Apriori ... 2
Bidirecti... 2
0
20
Count of Paper ID
Fig 5 : Optimization Algorithms
Optimization Algorithm
Genetic algo... 4
Adam 2
Grid Search 2
Particle Swar... 2
AGTO 1
Ant colony o... 1
Bat Optimiz... 1
COA 1
CoatiOA 1
Competitive ... 1
Differential ... 1
Genetic Alg... 1
Modular ne... 1
MSGD 1
0
2
4
Count of Paper ID
Algorithm Count
Random Forest 35
Logistic Regression 25
SVM 24
Decision Tree 18
KNN 14
CNN 13
XGBoost 12
ANN 9
LSTM 9
Naive Bayes 8
AdaBoost 7
MLP 6
Autoencoder 5
Bagging 5

GAPS

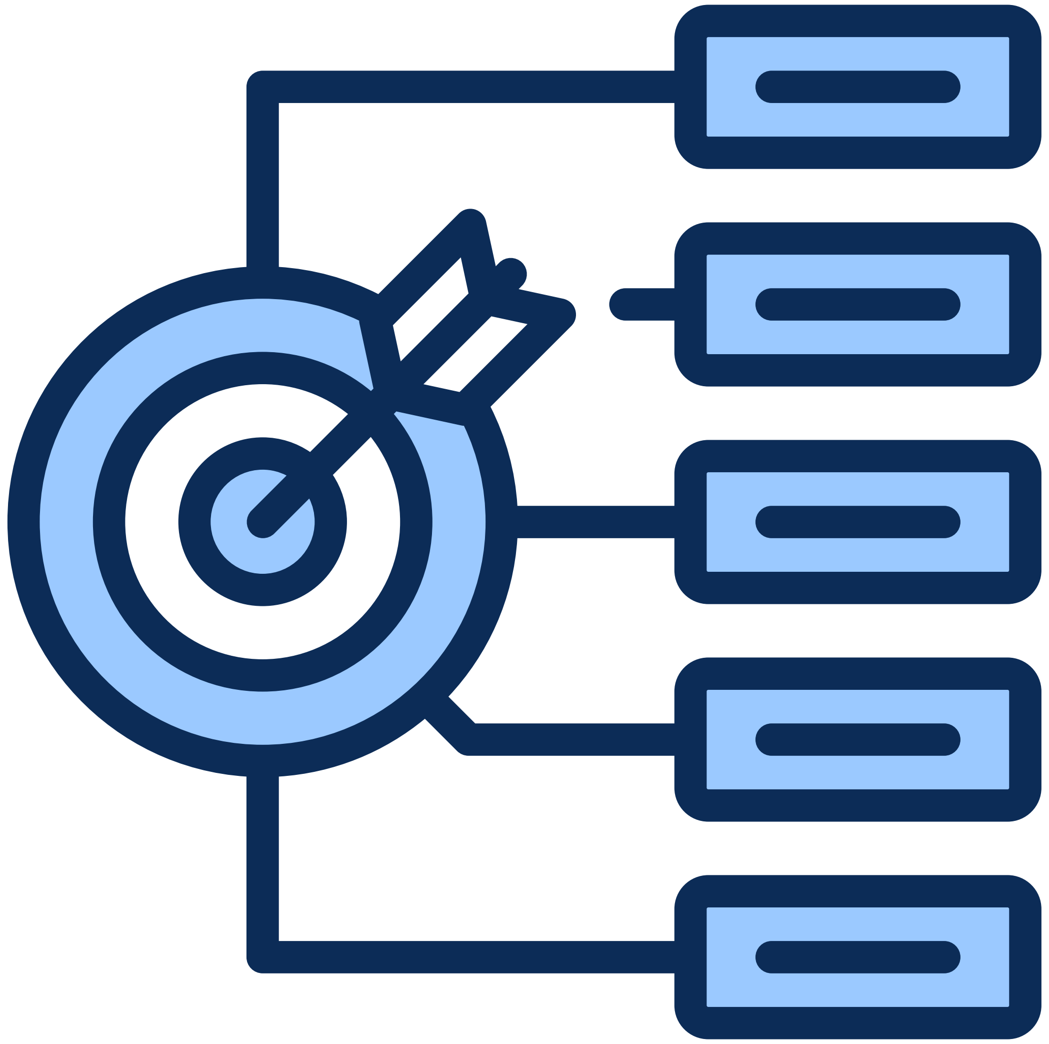
Business gaps:

- Undetected fraud → revenue loss
- Declining customer trust
- Delayed fraud response
- High false positives

Technical gaps:

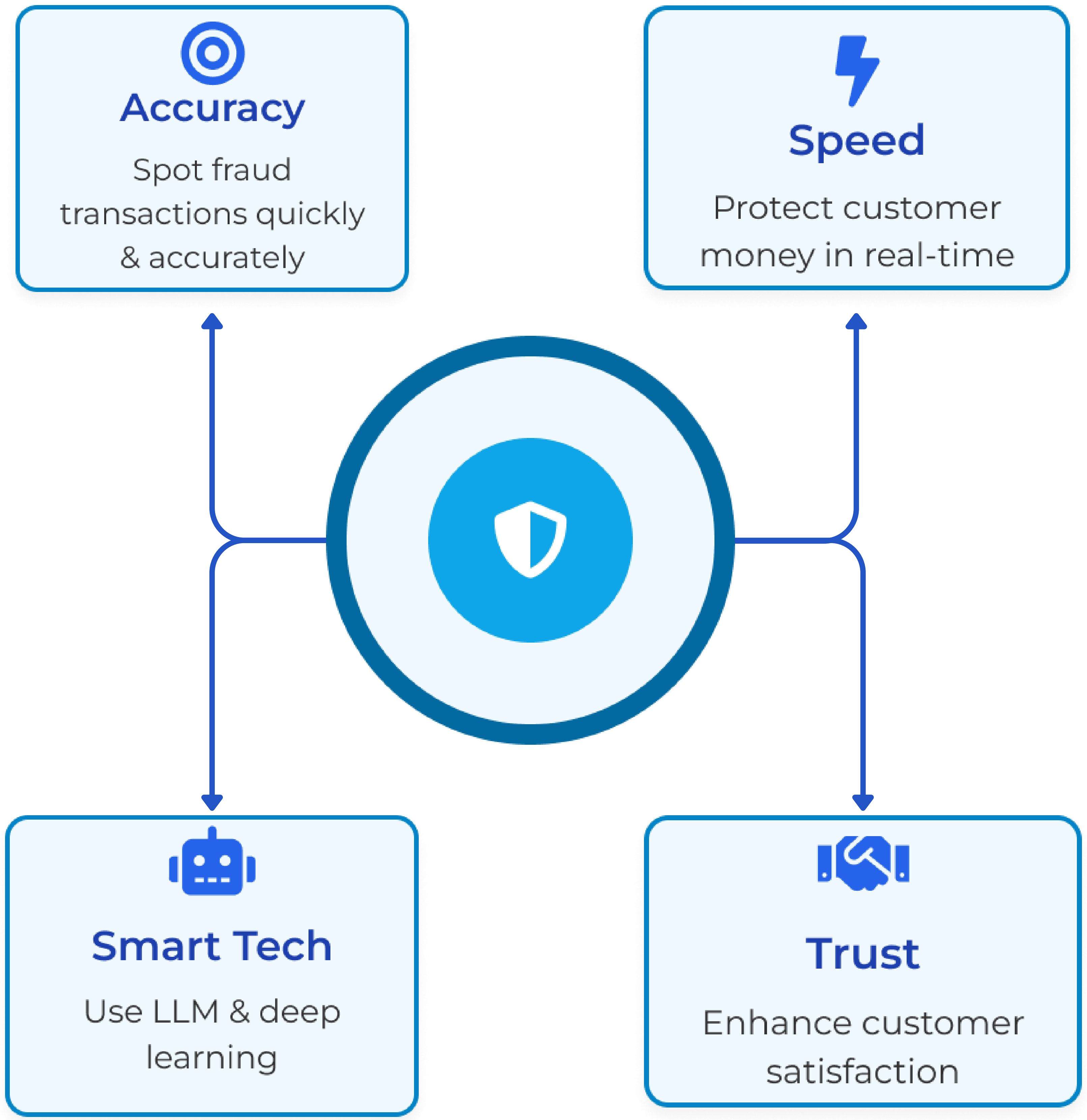
- Poor handling of missing data
- Underuse of metaheuristic optimization
- Lack of real-time, user-friendly platforms
- No use of LLMs for fraud detection

4. Objectives

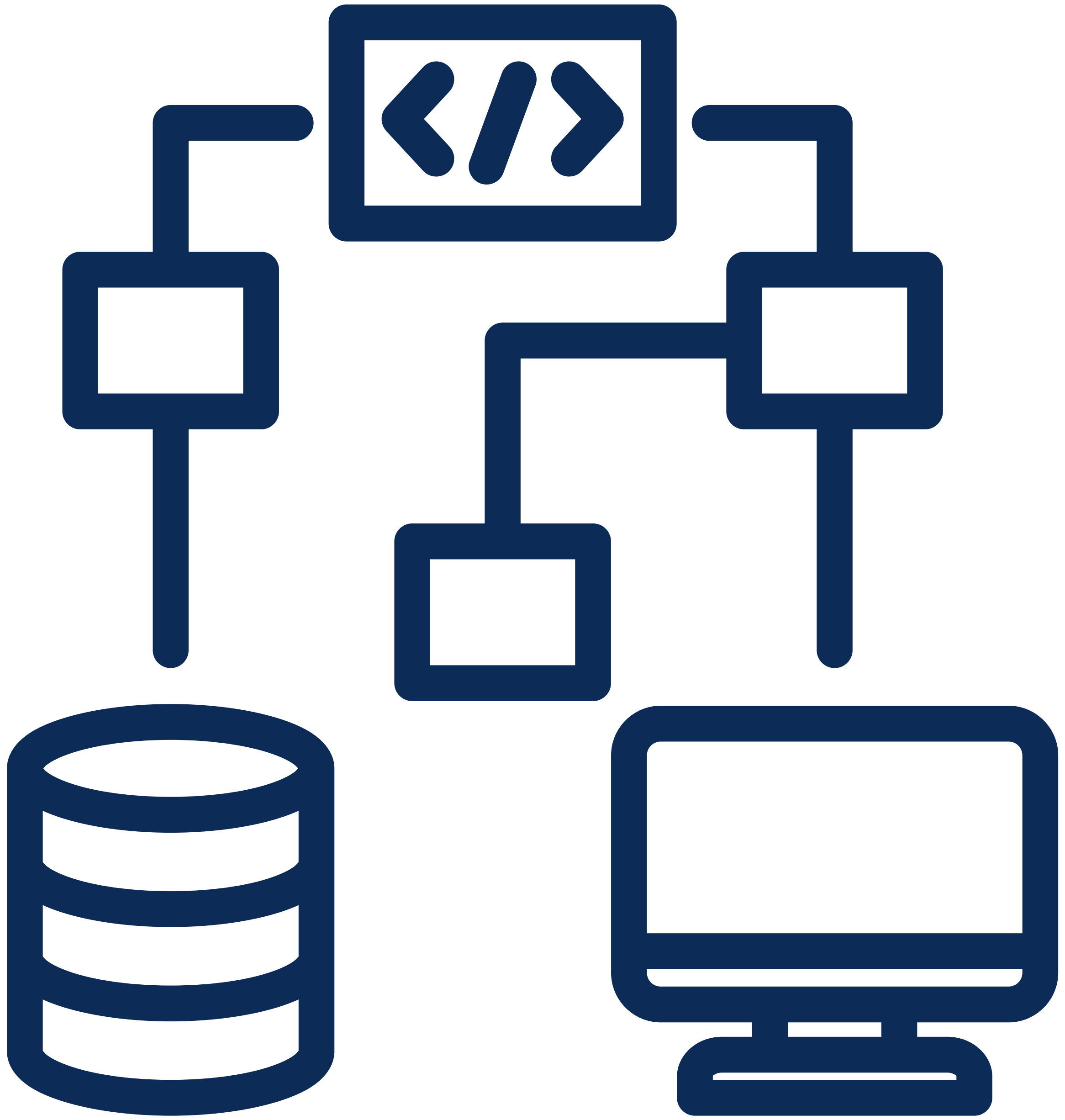


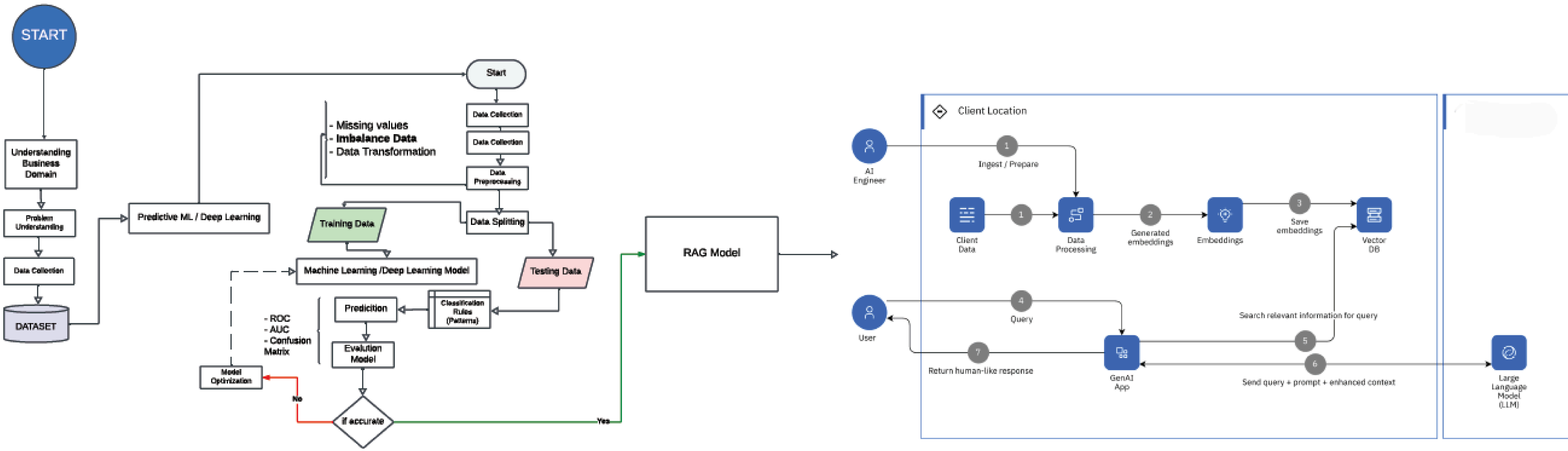
Objectives

- Spot fraud transactions quickly & accurately
- Protect customer money
- Use smart technologies like LLM & deep learning
- Enhance customer trust & satisfaction

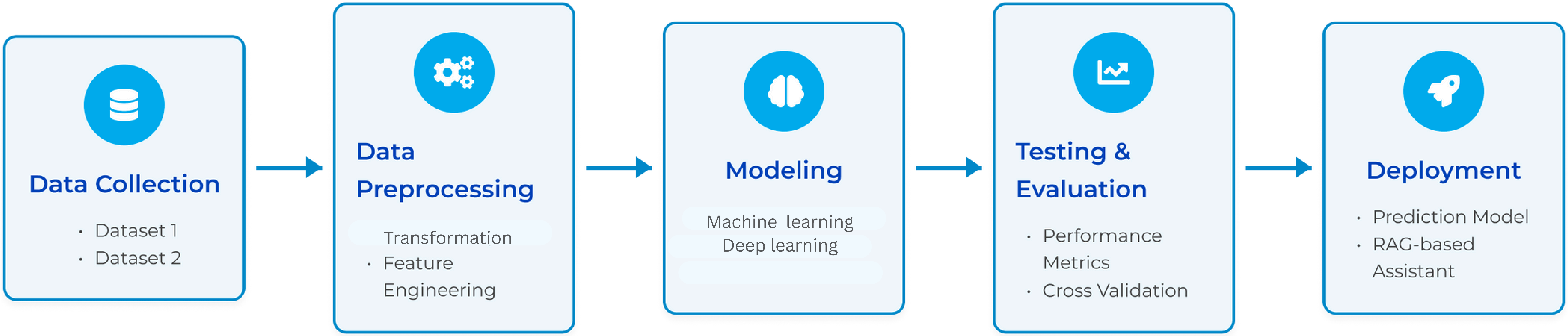


Framework




START
Understanding Business Domain
Problem Understanding
Data Collection
DATASET
Predictive ML / Deep Learning
Start
Data Collection
Data Collection
Data Preprocessing
- Missing values
- Imbalance Data
- Data Transformation
Data Splitting
Training Data
Testing Data
Machine Learning /Deep Learning Model
Classification Rules (Patterns)
Predicition
Evalution Model
- ROC
- AUC
- Confusion Matrix
Model Optimization
if accurate
No
Yes
RAG Model
Client Location
AI Engineer
Ingest / Prepare
Client Data
Data Processing
Generated embeddings
Embeddings
Save embeddings
Vector DB
User
Query
Search relevant information for query
Return human-like response
GenAI App
Send query + prompt + enhanced context
Large Language Model (LLM)

Framework



Testing & Evaluation

Train/Test Split
80% / 20%

Metrics
Precision, Recall, F1-score

Framework

Datasets Overview

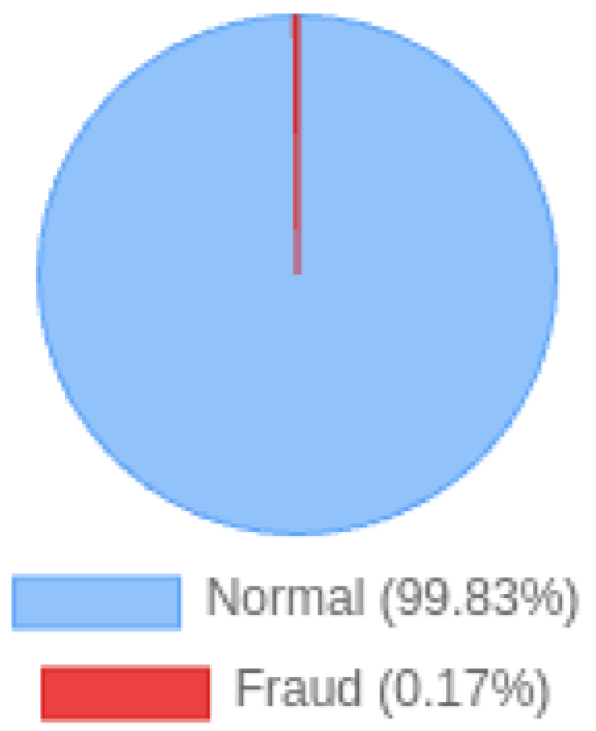
Dataset 1

- Records: 285,000
- Frauds: 498 (0.17%)
- Class ratio: 1:572
- Two Days of September 2013

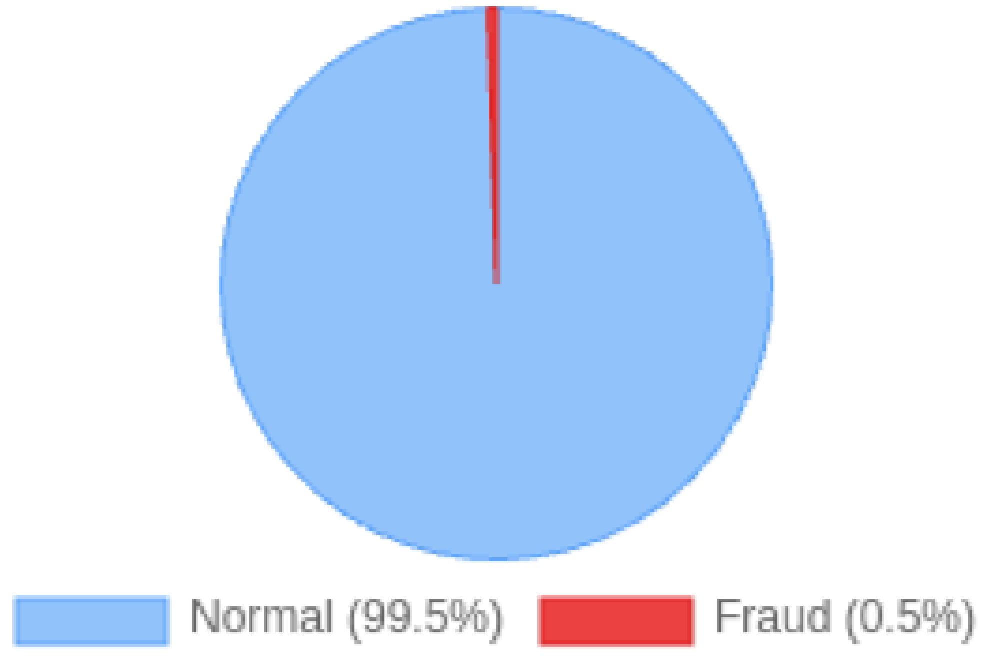
Dataset 2

- Records: 1,800,000
- Frauds: 9,000 (0.5%)
- Class ratio: 1:200
- From 1/2019 to 12/2020

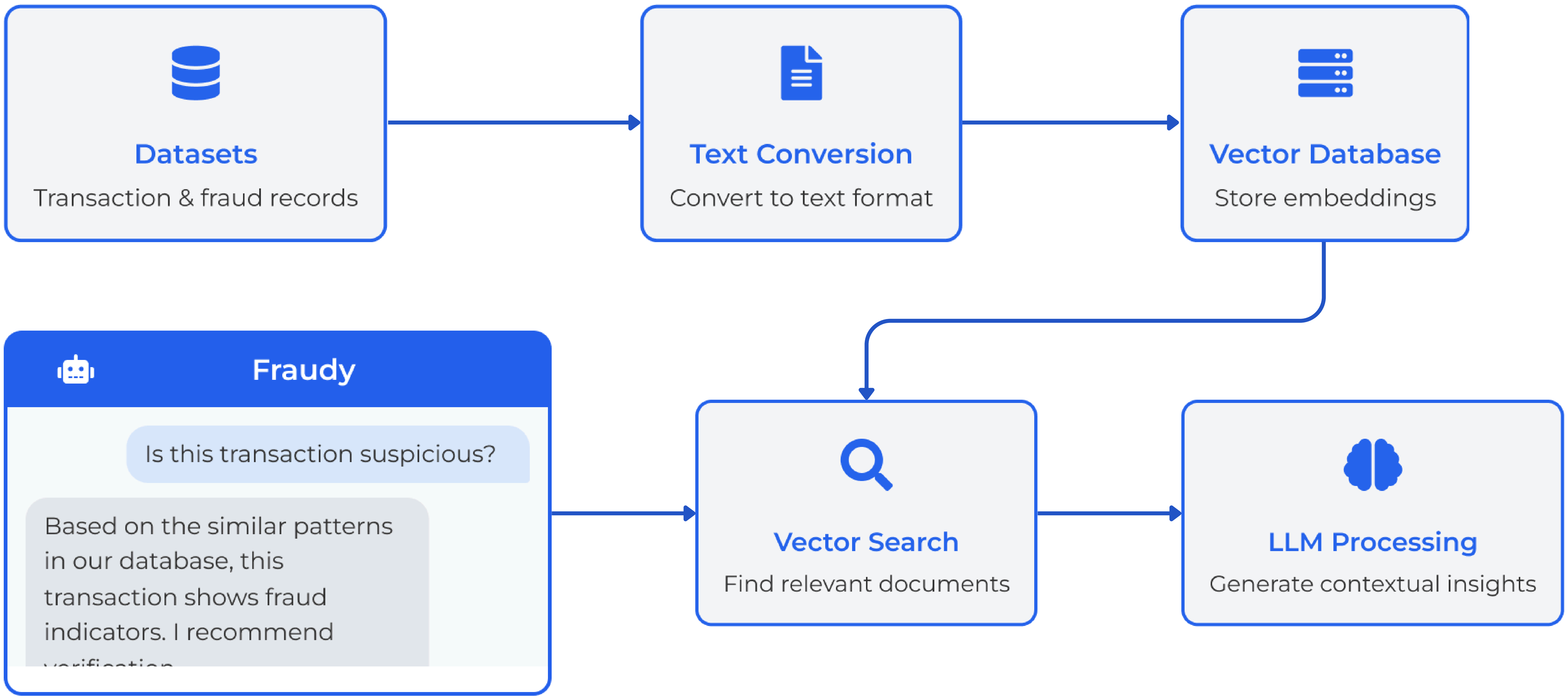
Dataset 1



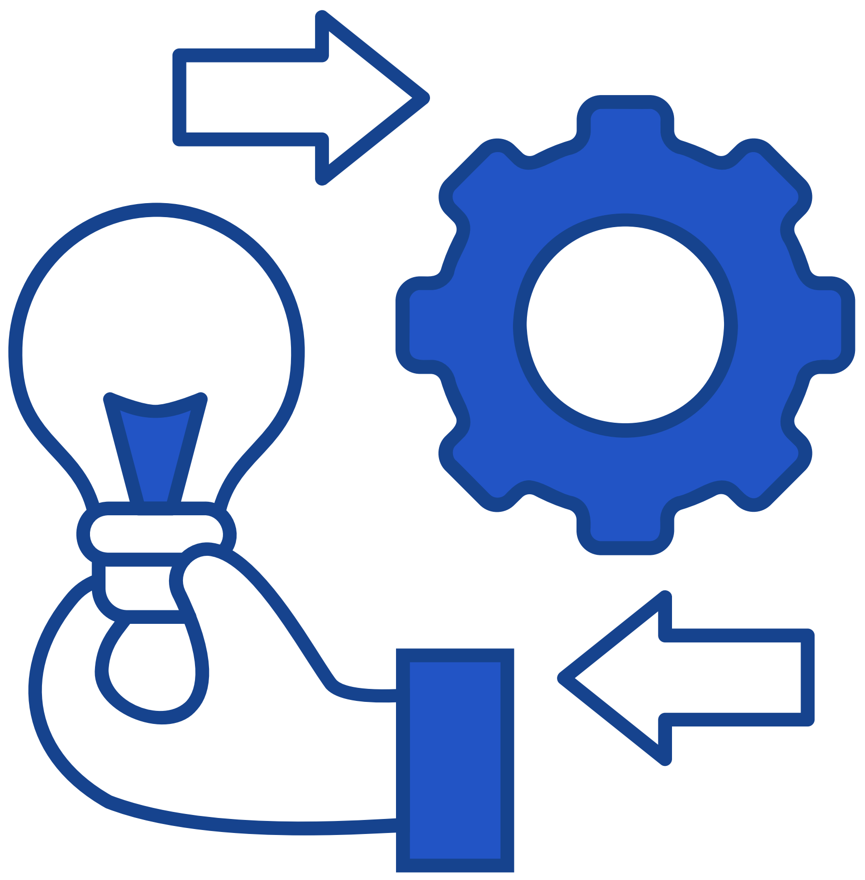
Dataset 2



Framework

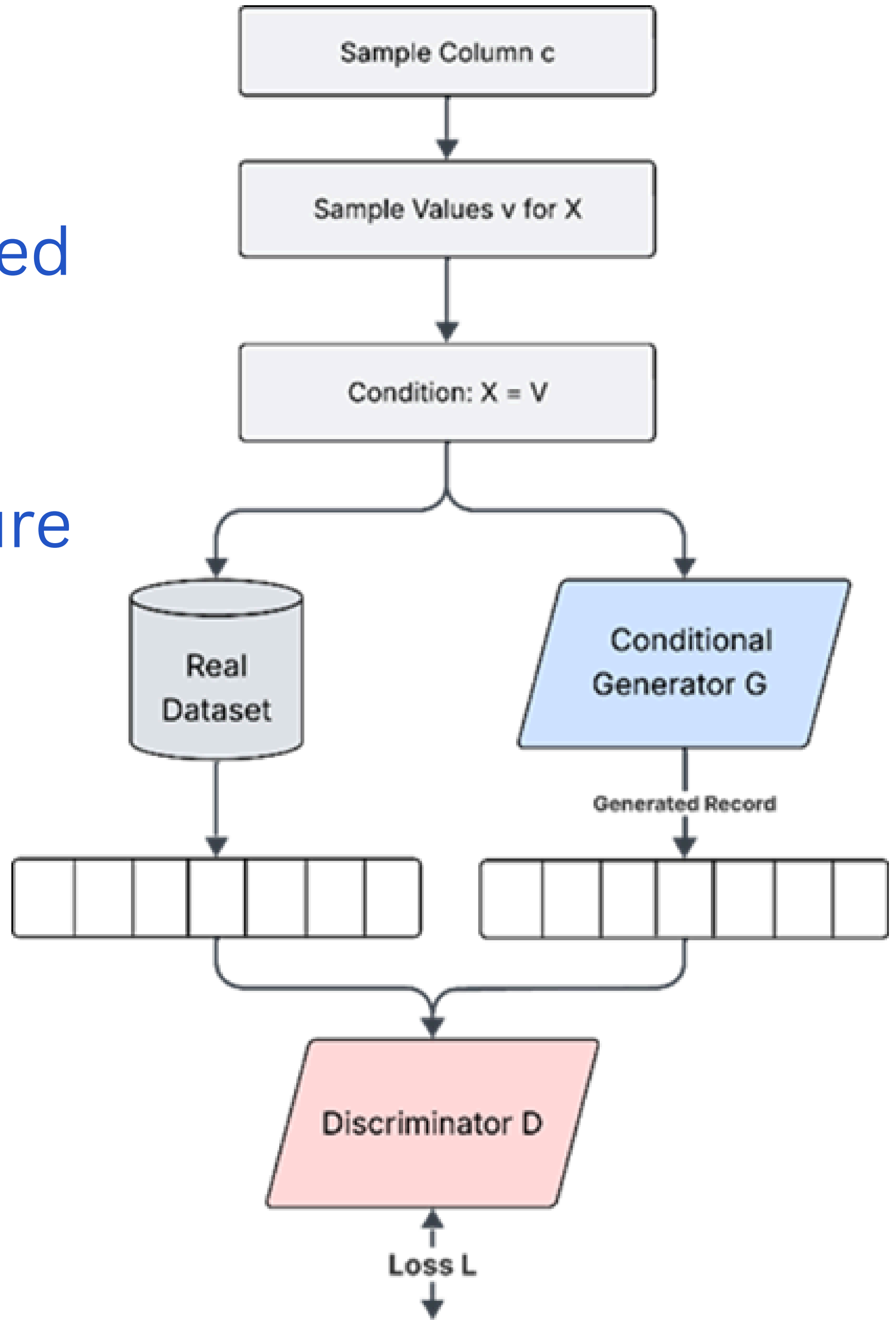

Datasets
Transaction & fraud records
Text Conversion
Convert to text format
Vector Database
Store embeddings
Fraudy
Is this transaction suspicious?
Based on the similar patterns in our database, this transaction shows fraud indicators. I recommend
Vector Search
Find relevant documents
LLM Processing
Generate contextual insights

Implementation



CTGANs - Conditional Tabular GANs

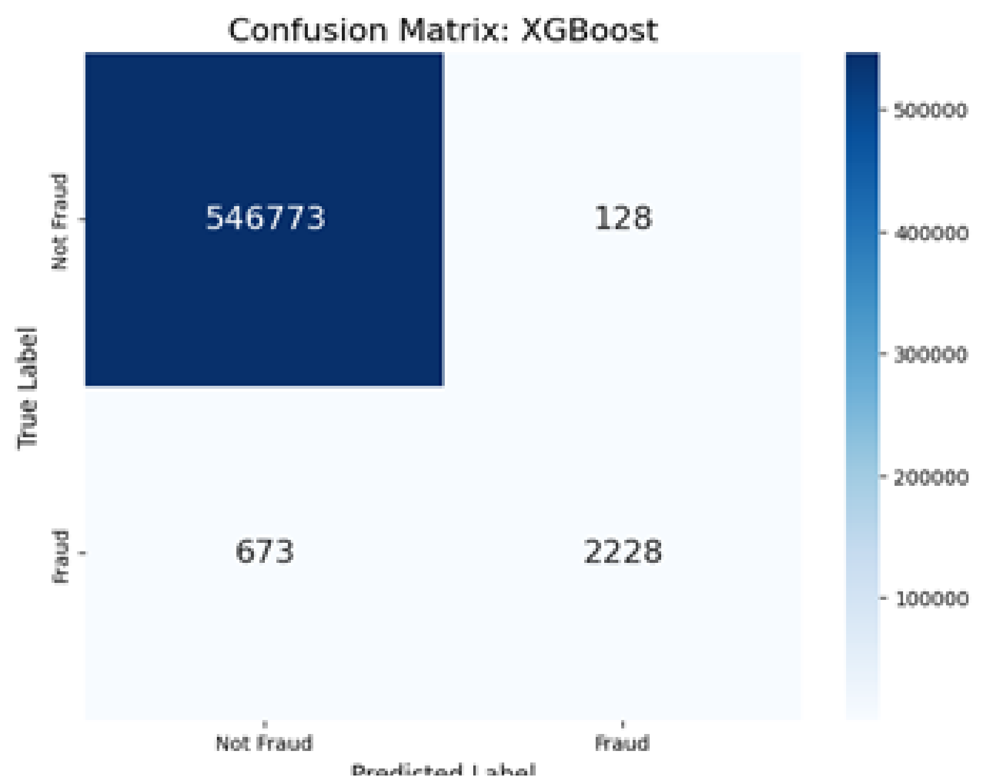
- GANs are built for images or continuous data so it struggle with mixed datatypes and rare classes
- A variant of GANs: Designed specifically for structured tabular data
- Based on Transformers: Uses attention to model complex feature dependencies
- Handles both categorical and numerical features
- Preserves relationships between columns
- Generated high-quality synthetic fraud data

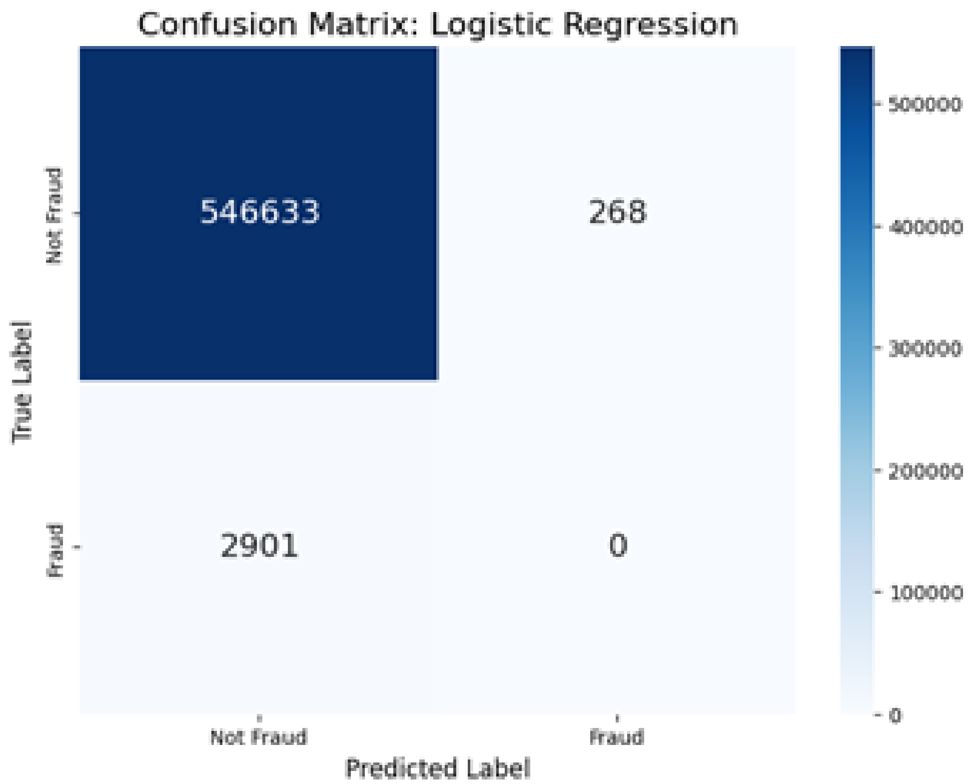


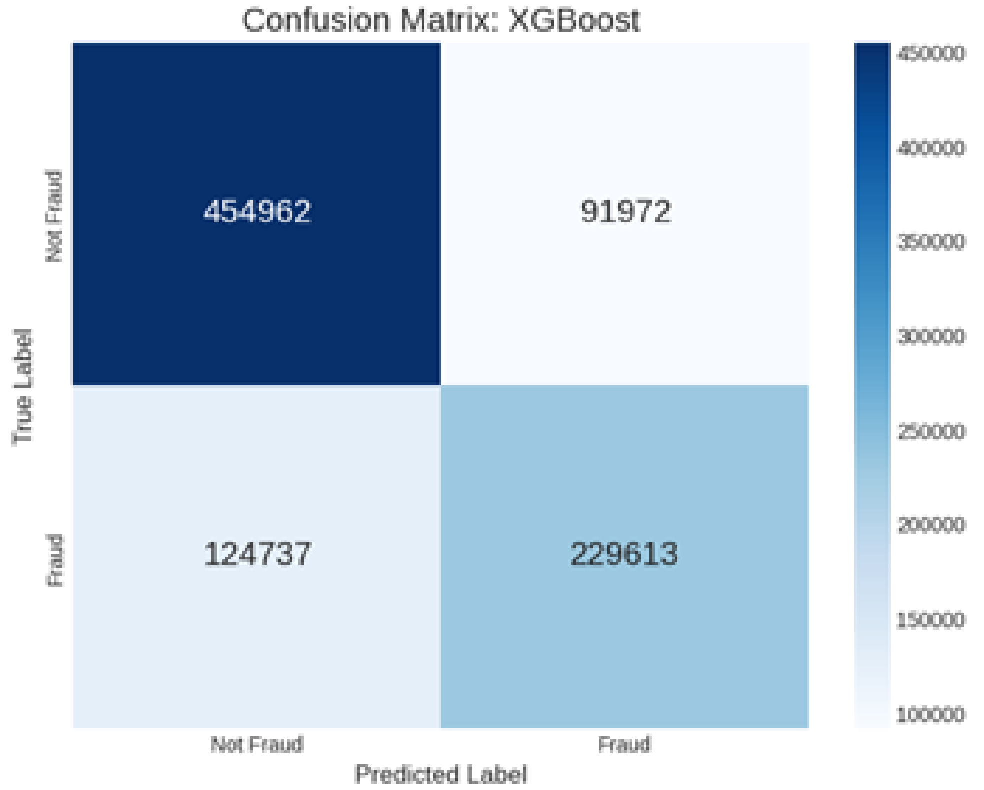
Model Implementation

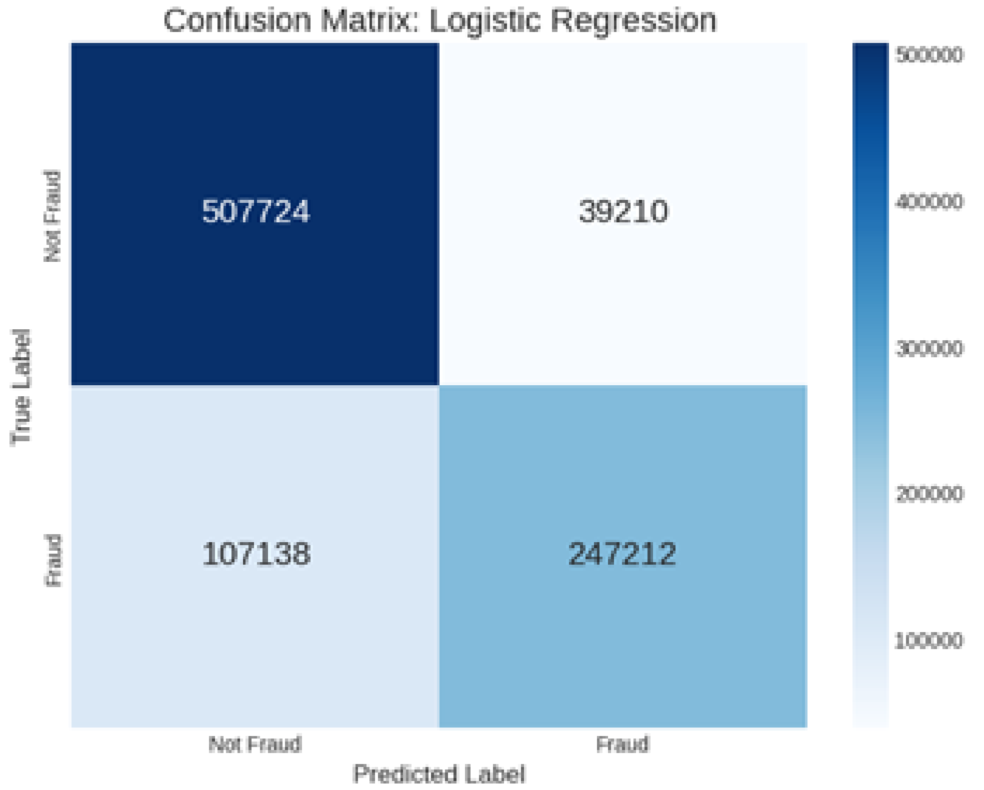
1. Imbalance and CTGANs
2. Logistic Regression
2. XGBoost
3. DNN
4. LSTM
5. AutoEncoder
6. AutoEncoder + LR

Machine Learning Models (CCFD)

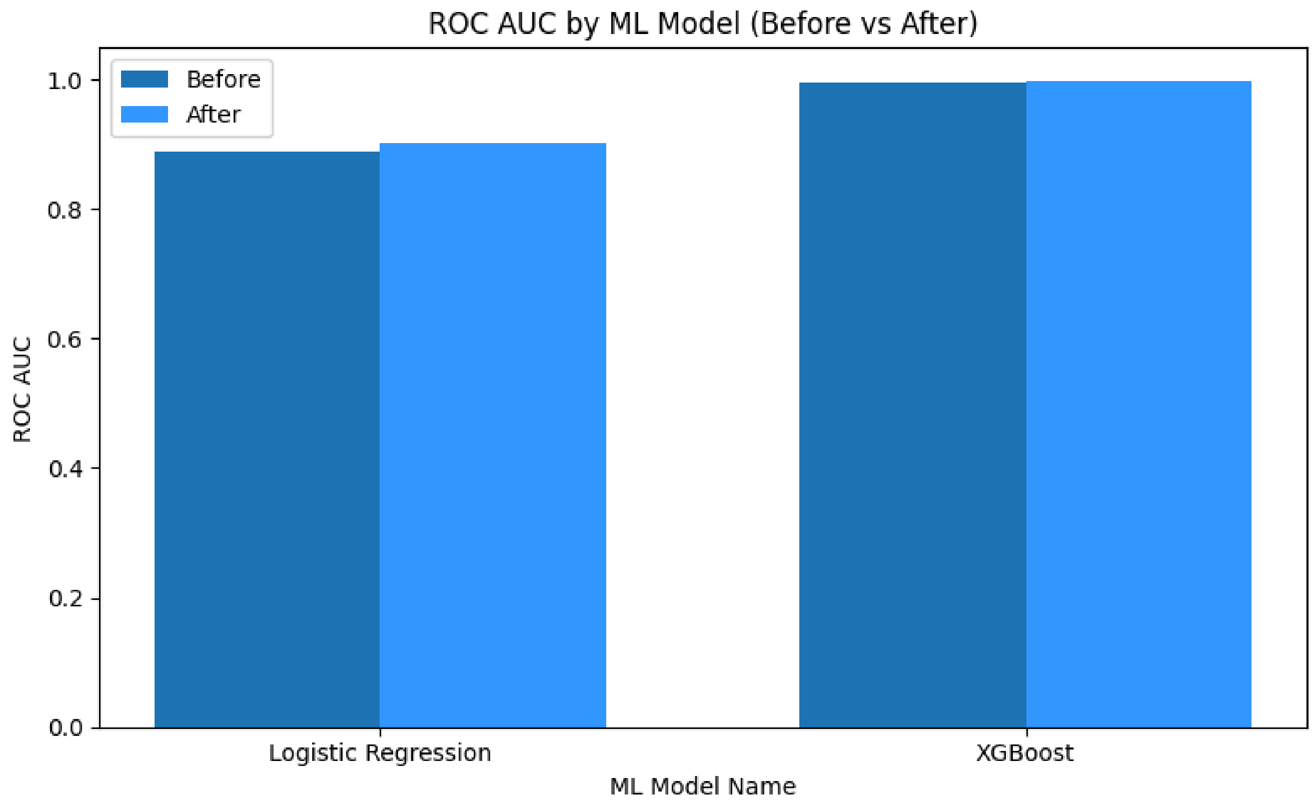


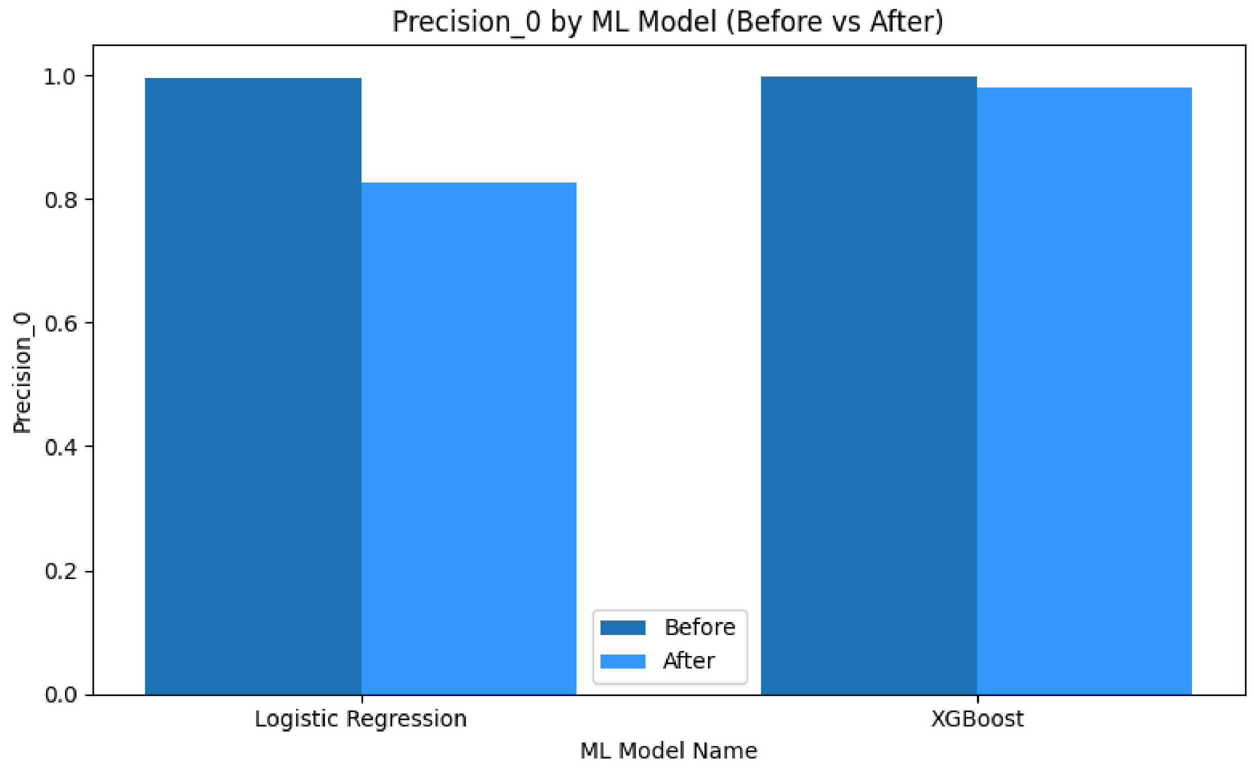


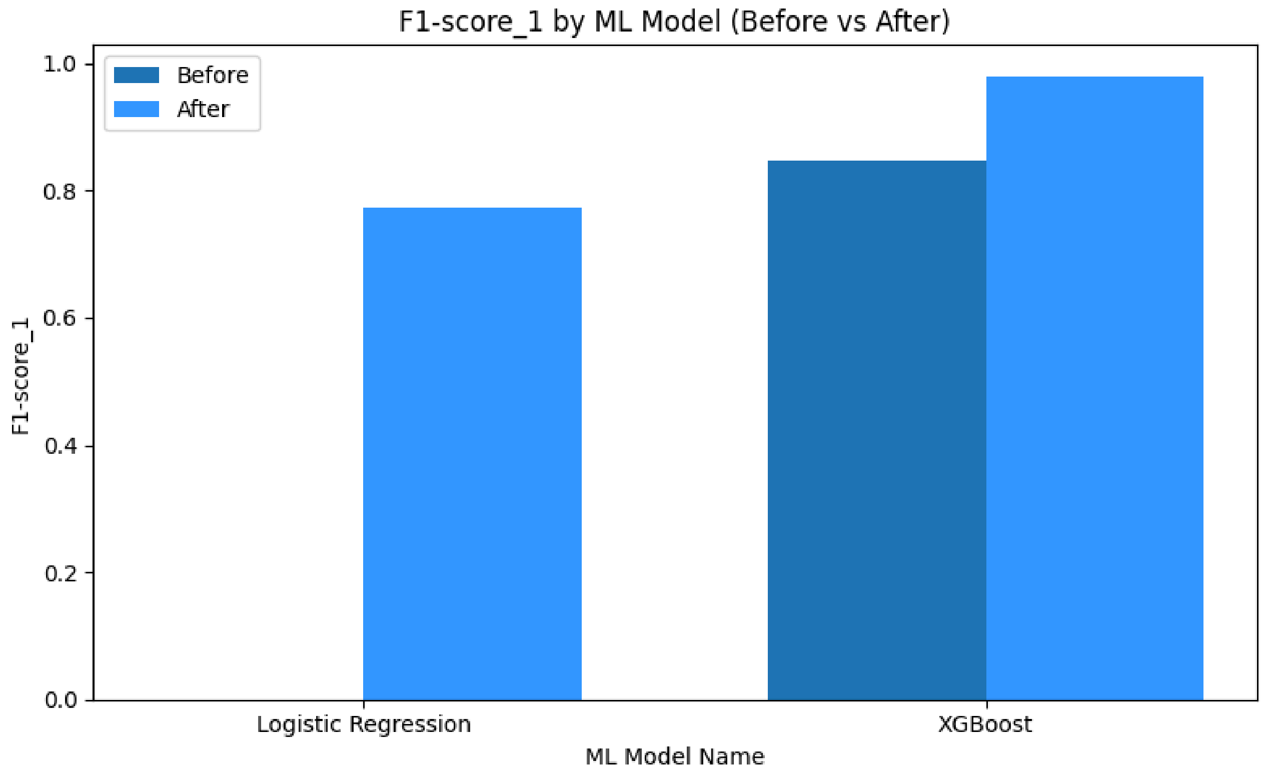


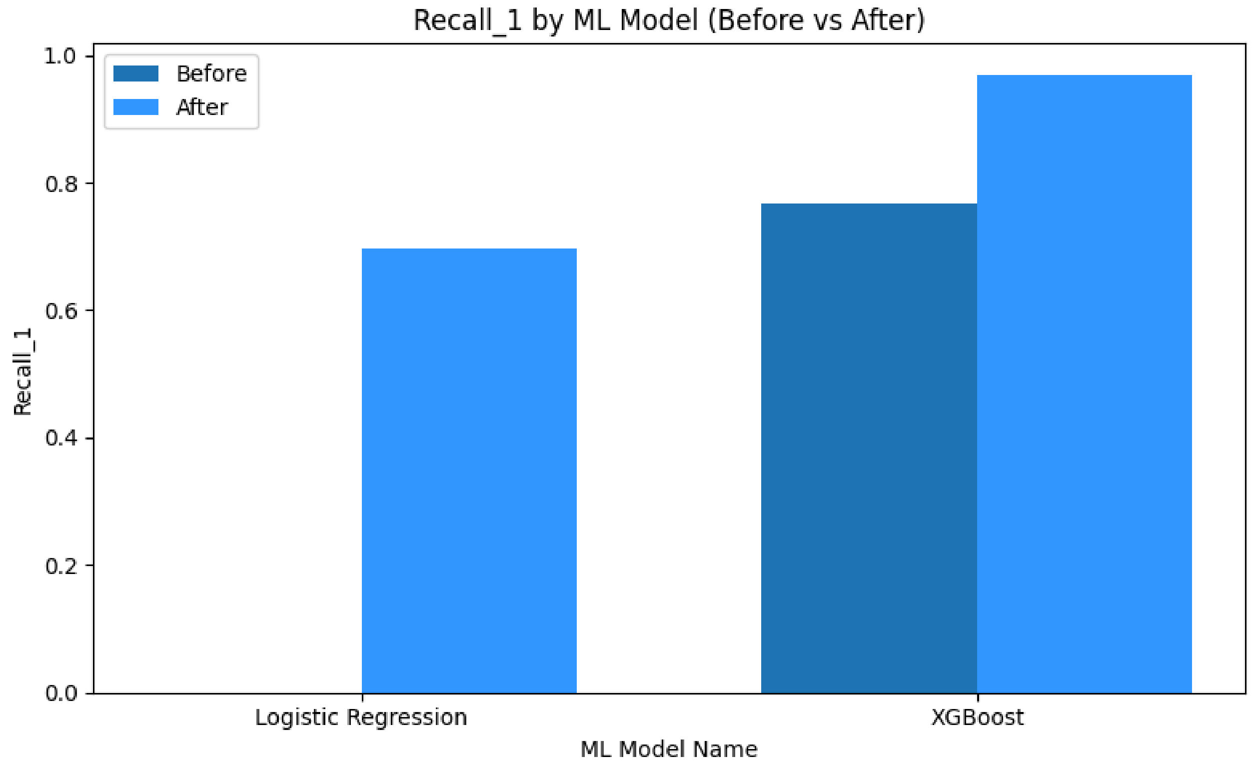


Machine Learning Models (CCFD)

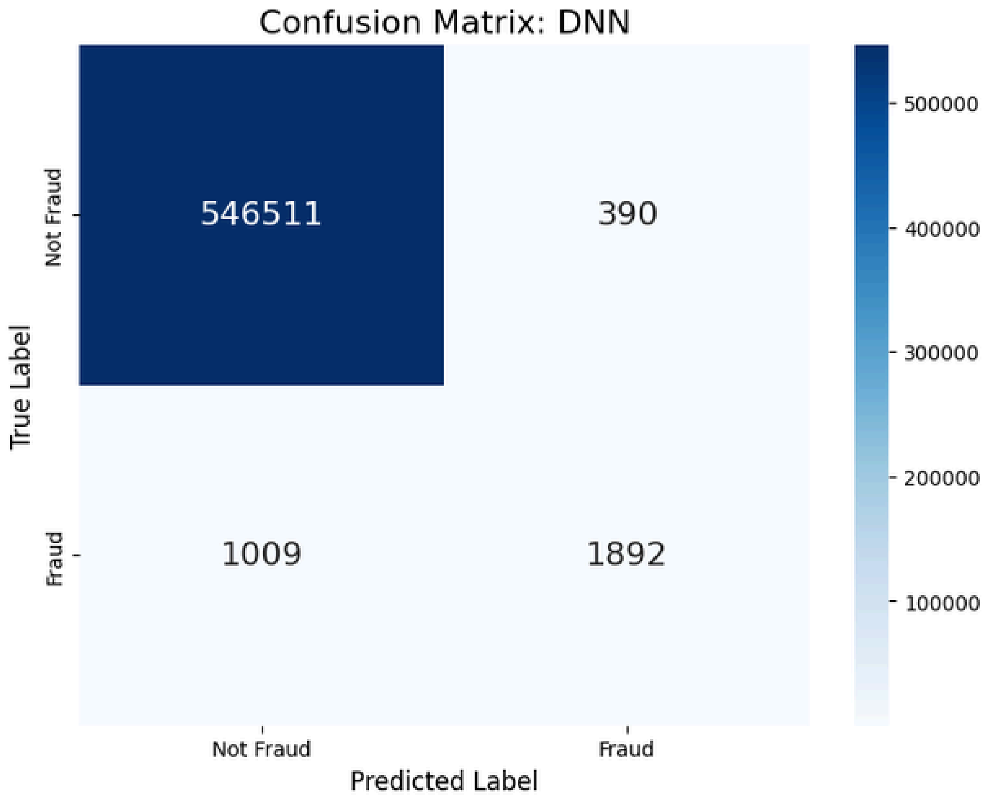


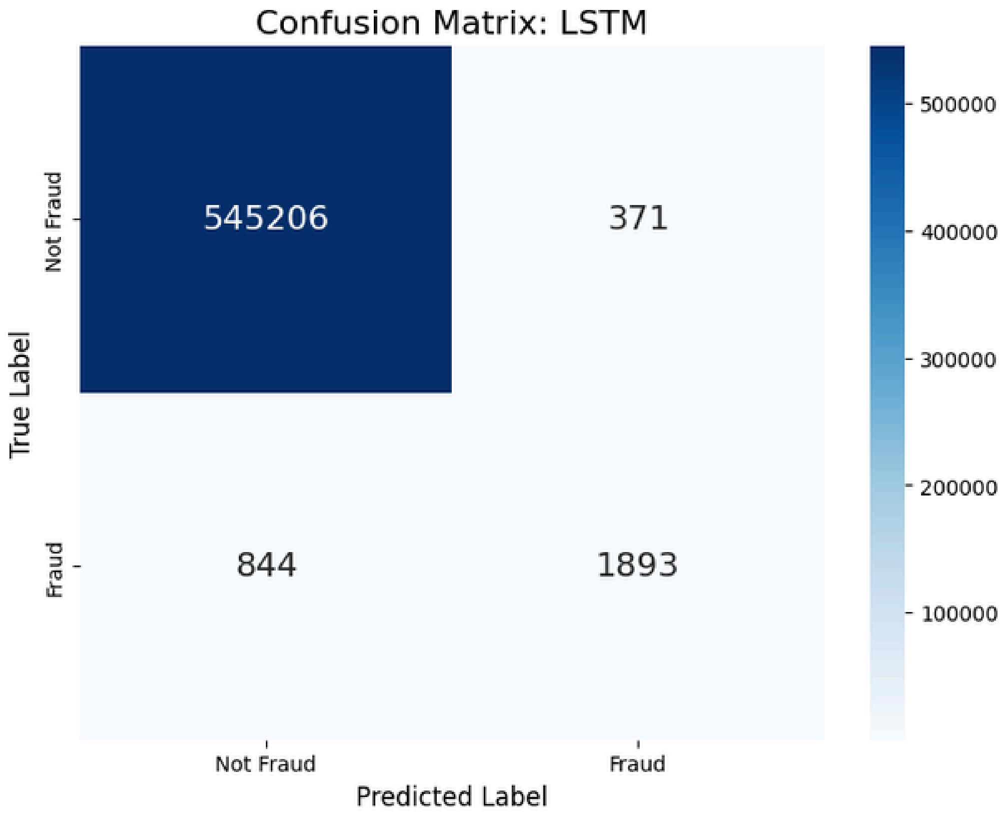


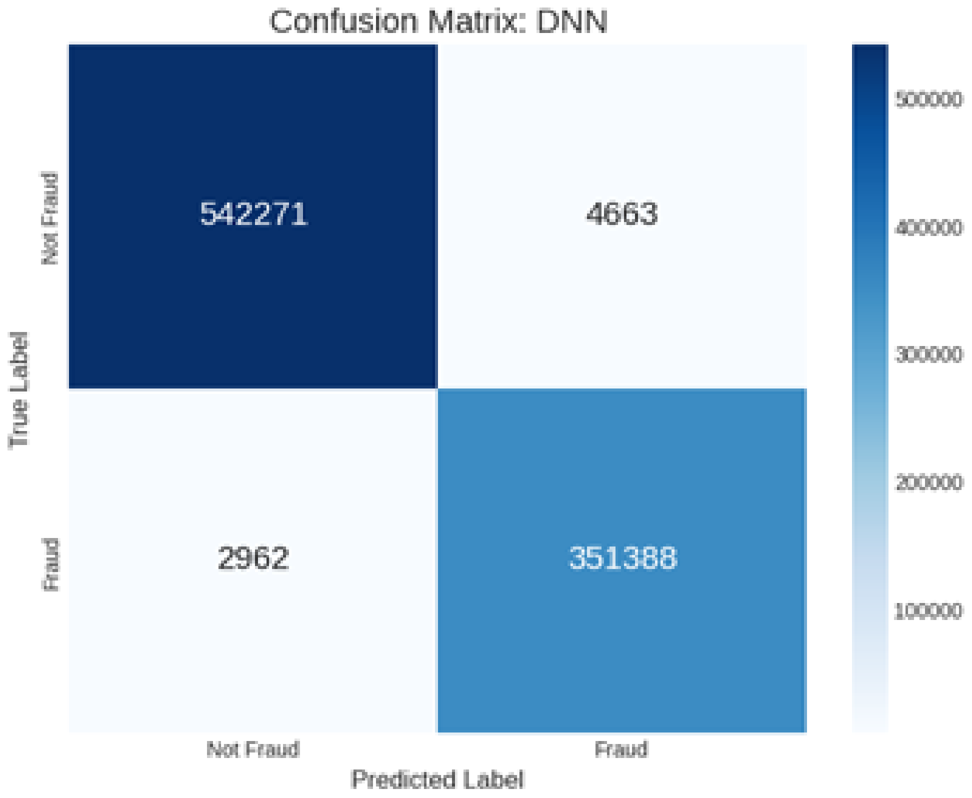


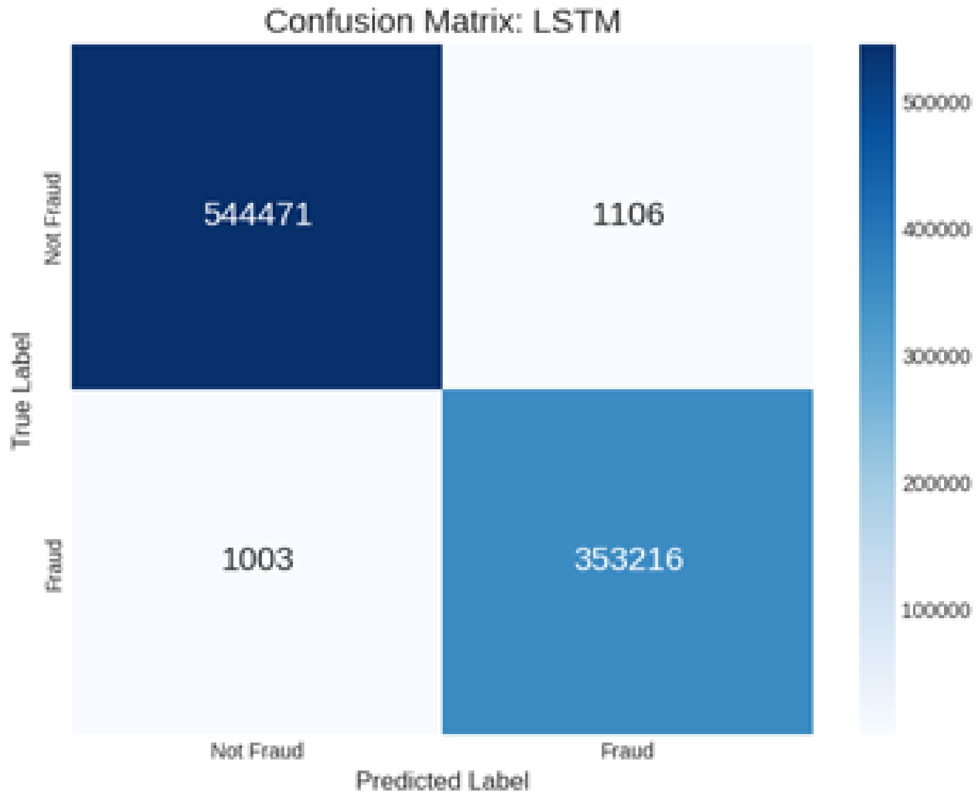


Deep Learning Models (CCFD)

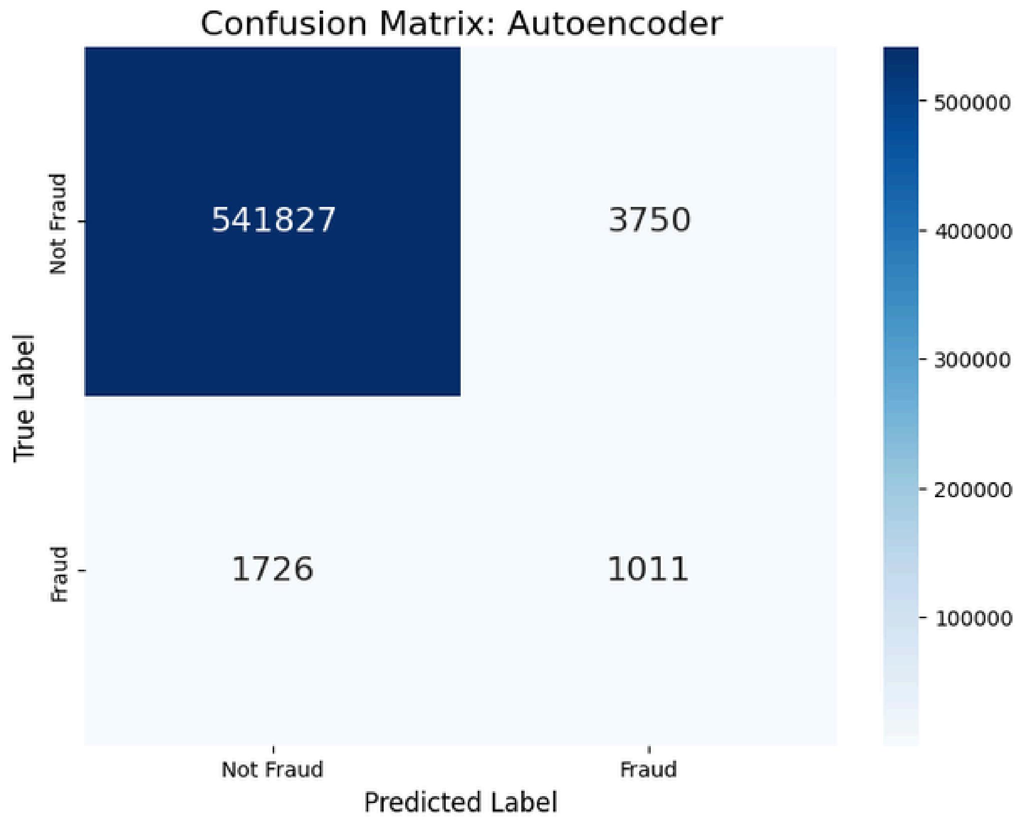


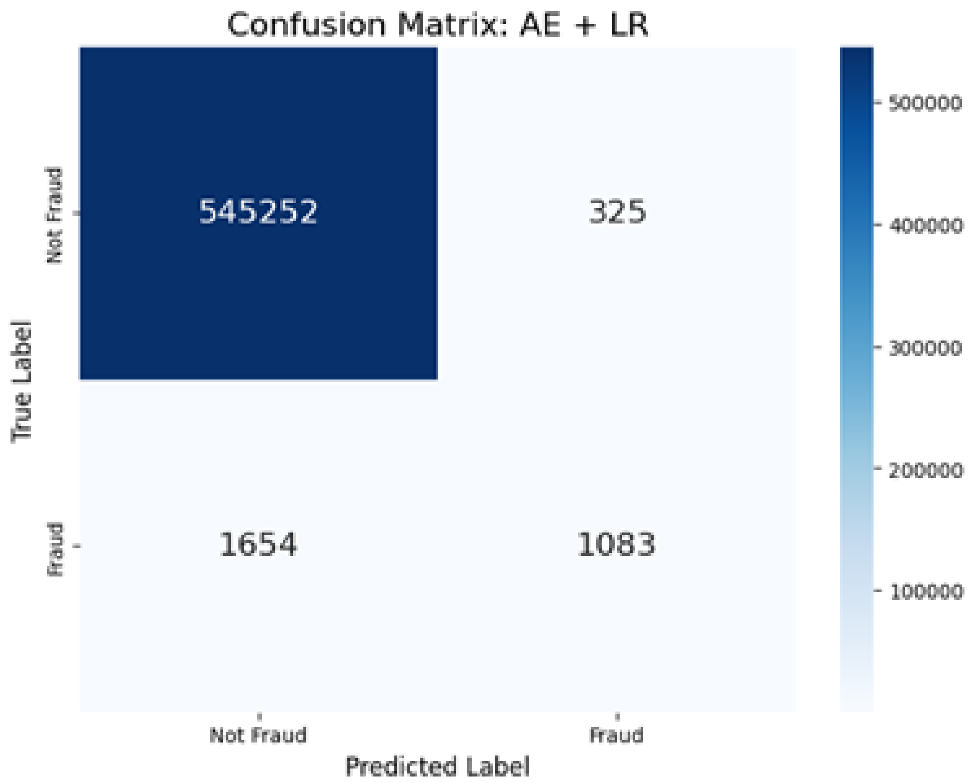


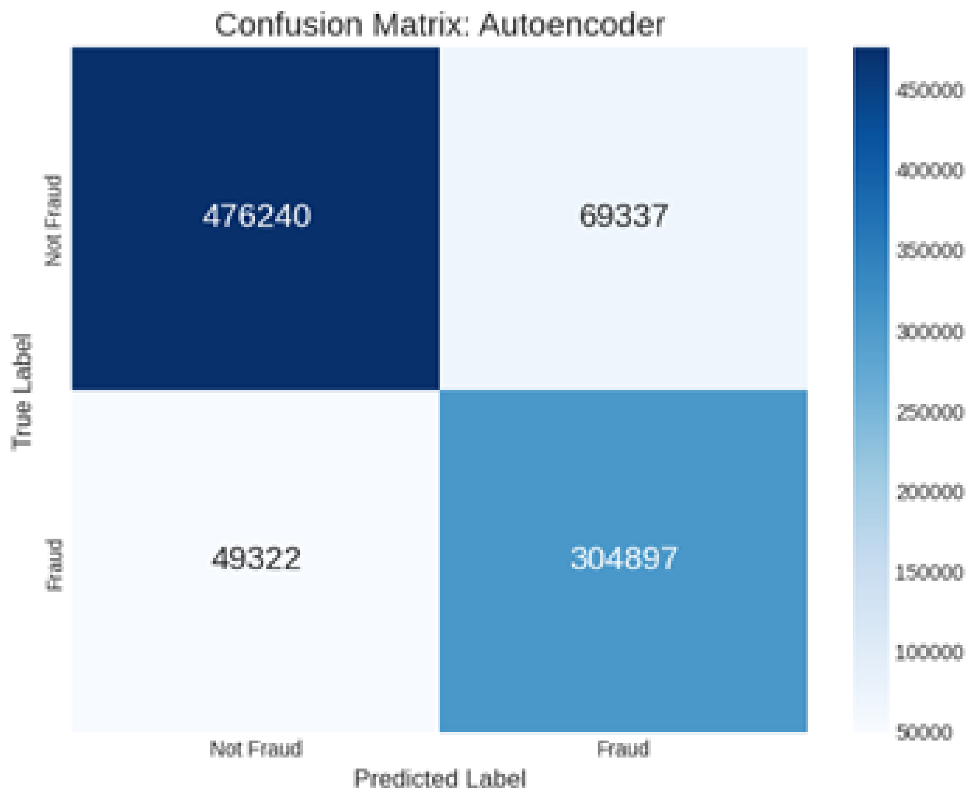


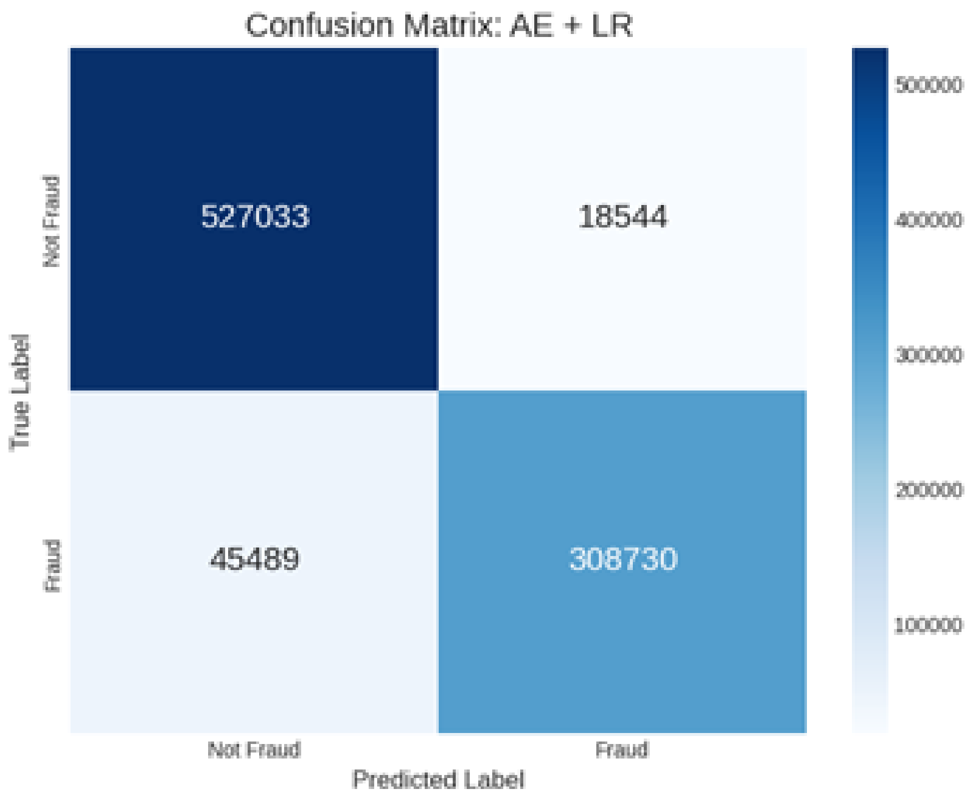


Deep Learning Models (CCFD)

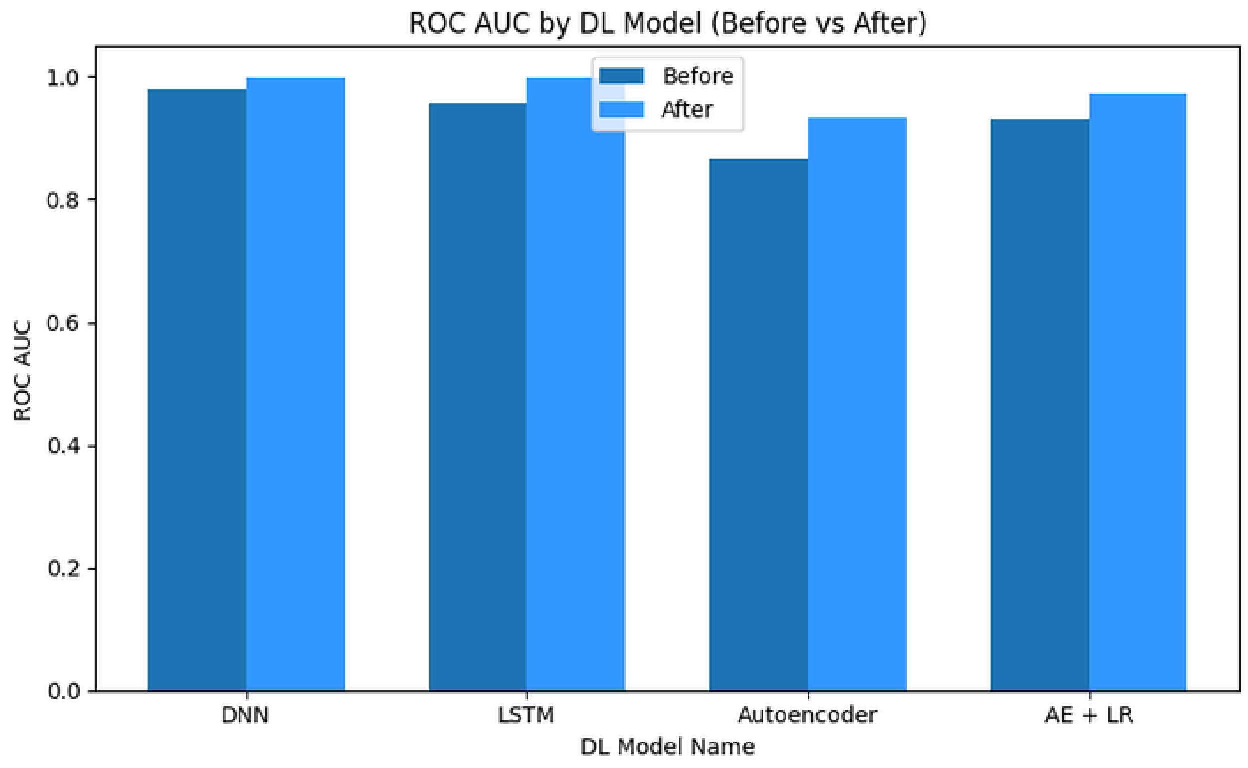


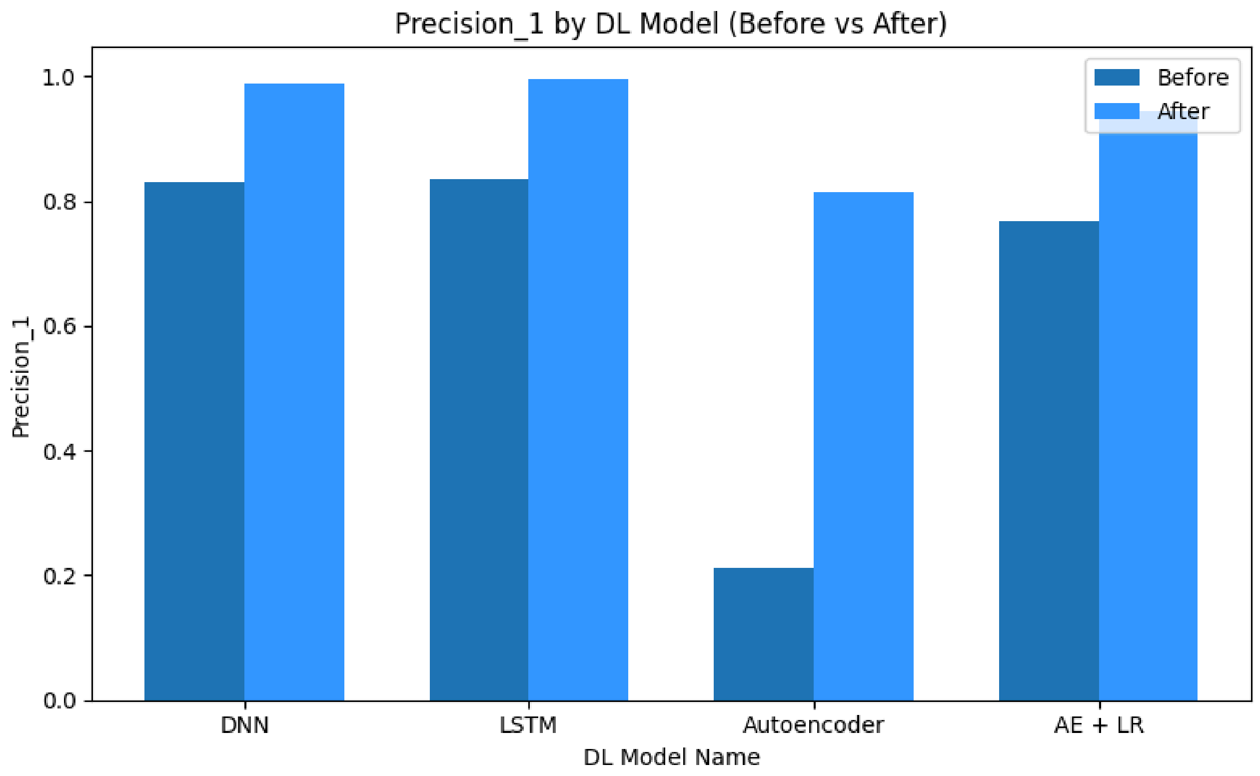


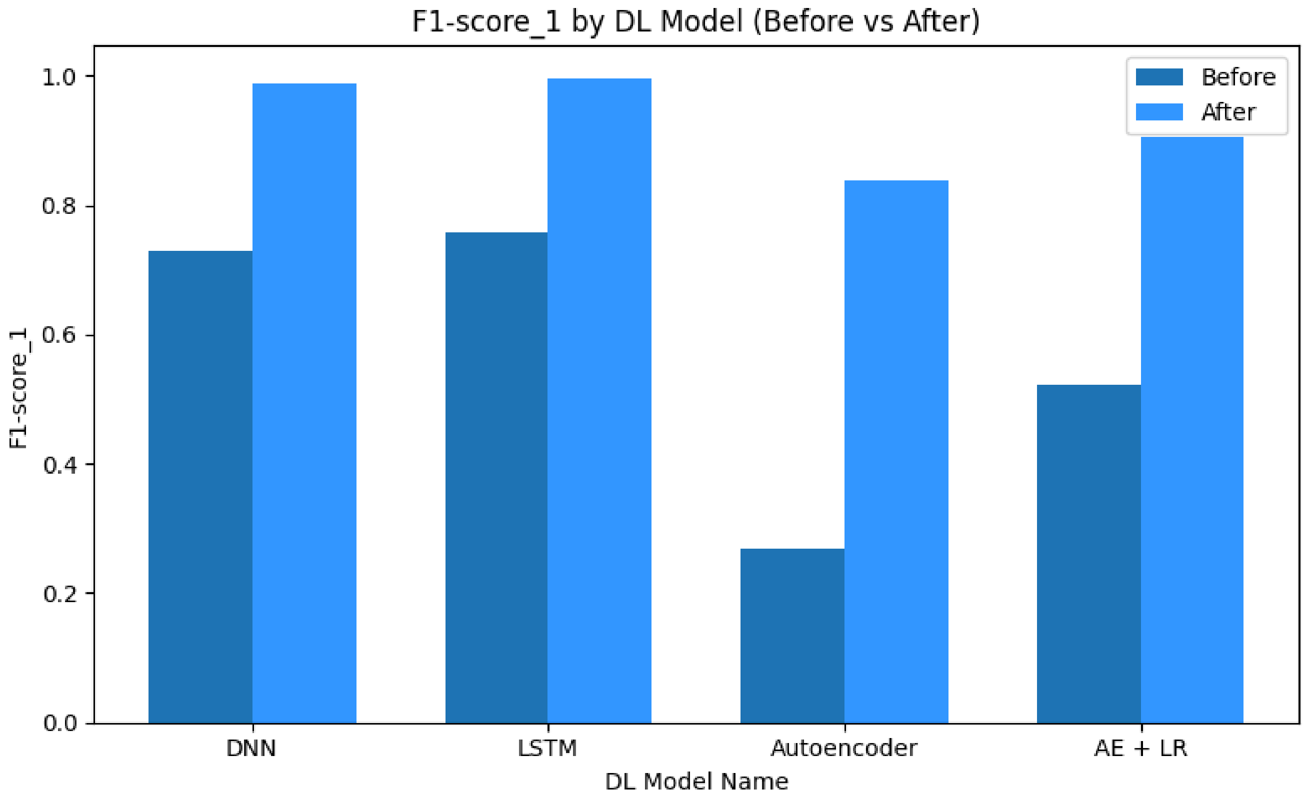


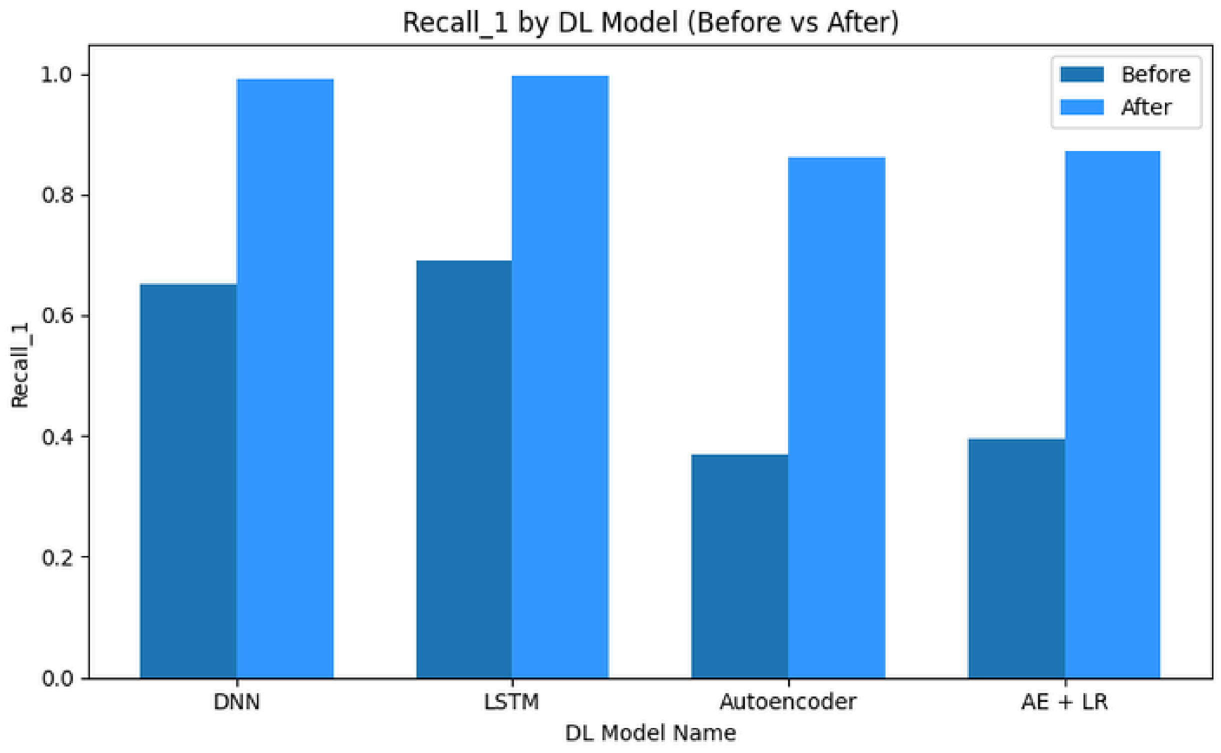


Deep Learning Models (CCFD)



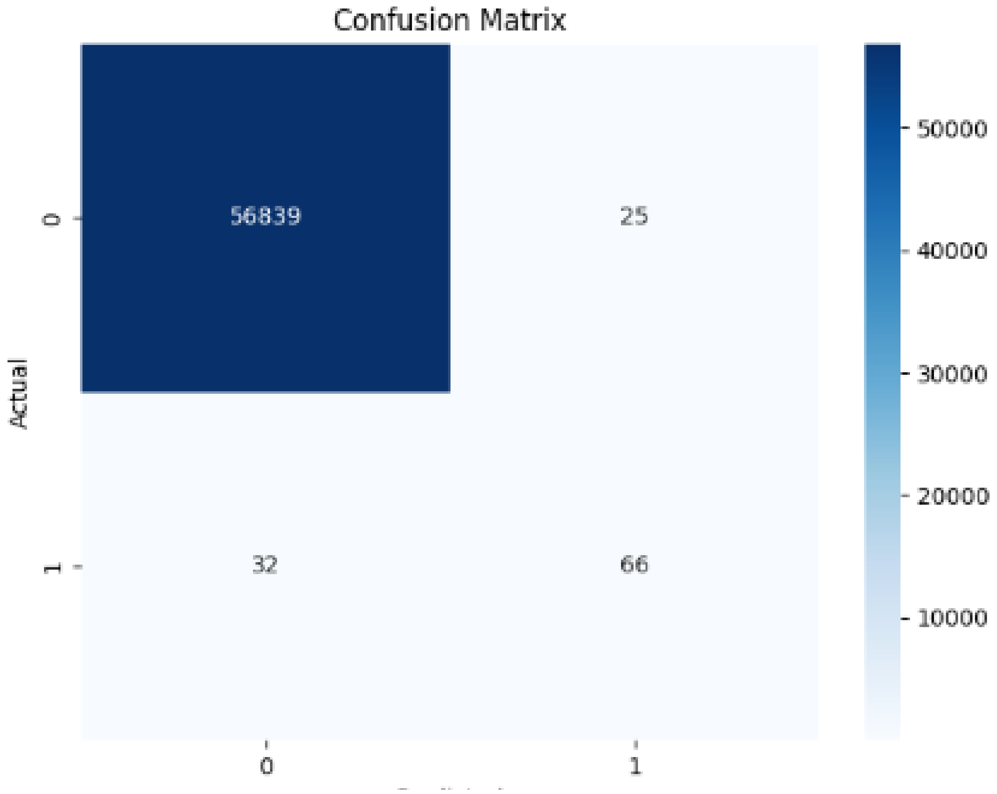




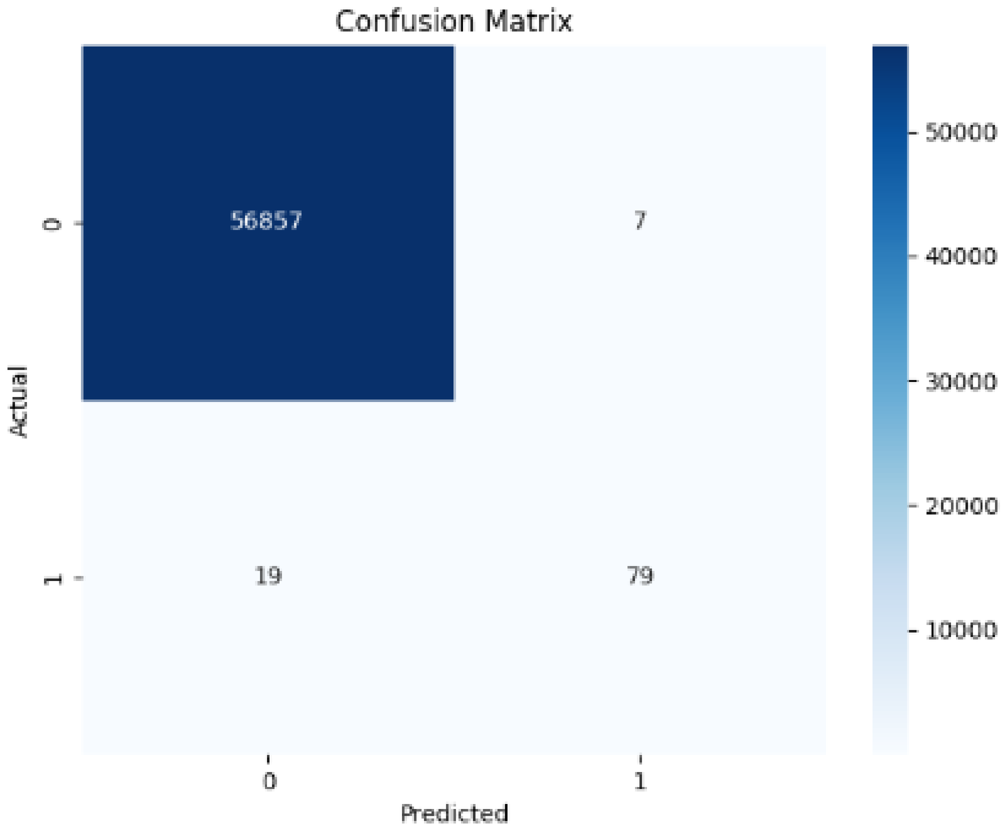


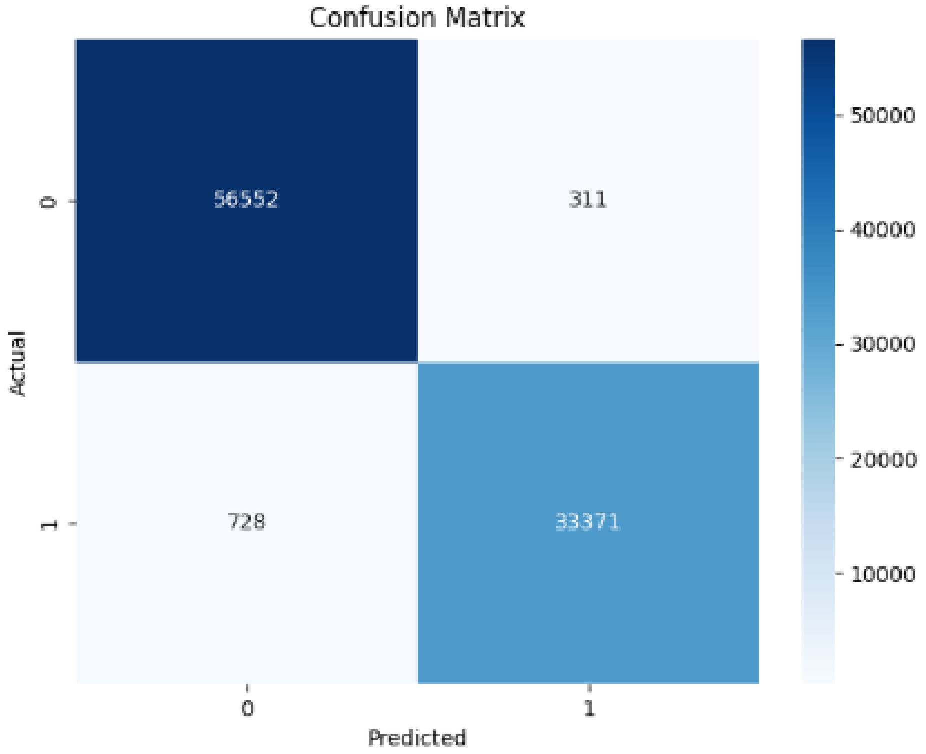
Machine Learning Models (ECCFD)

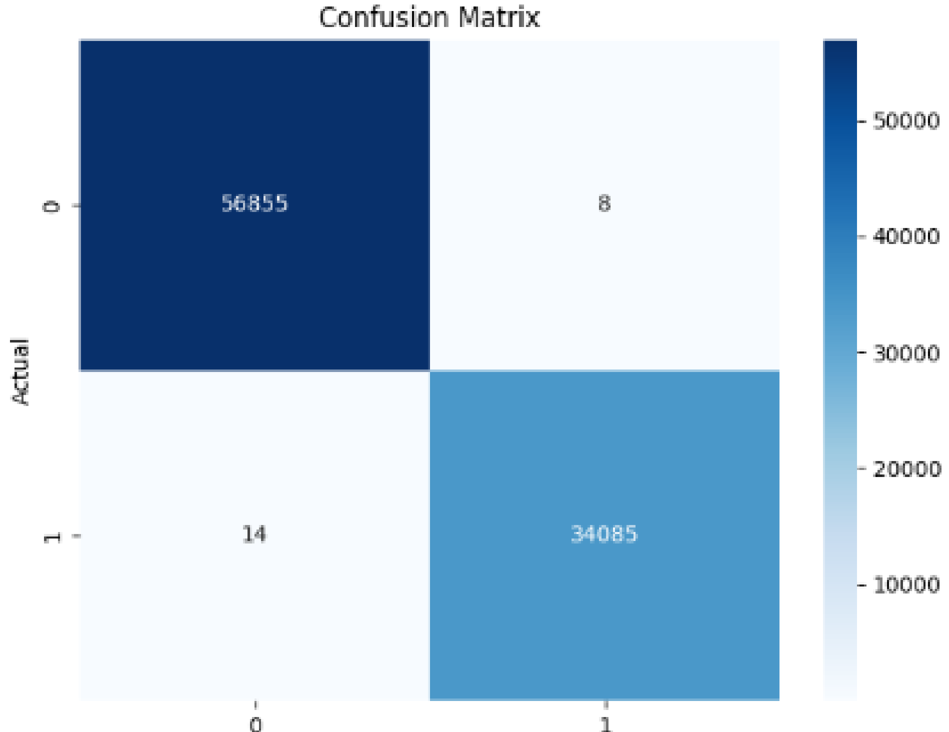
Logistic Regression



XGBoost

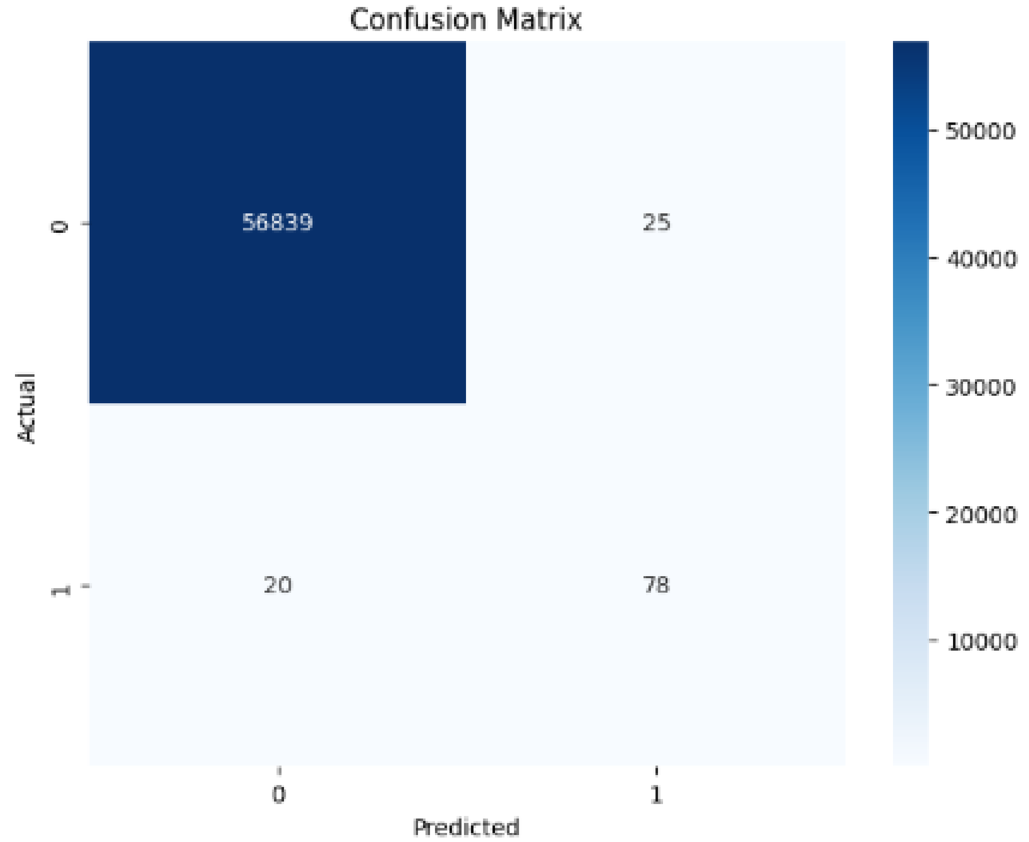




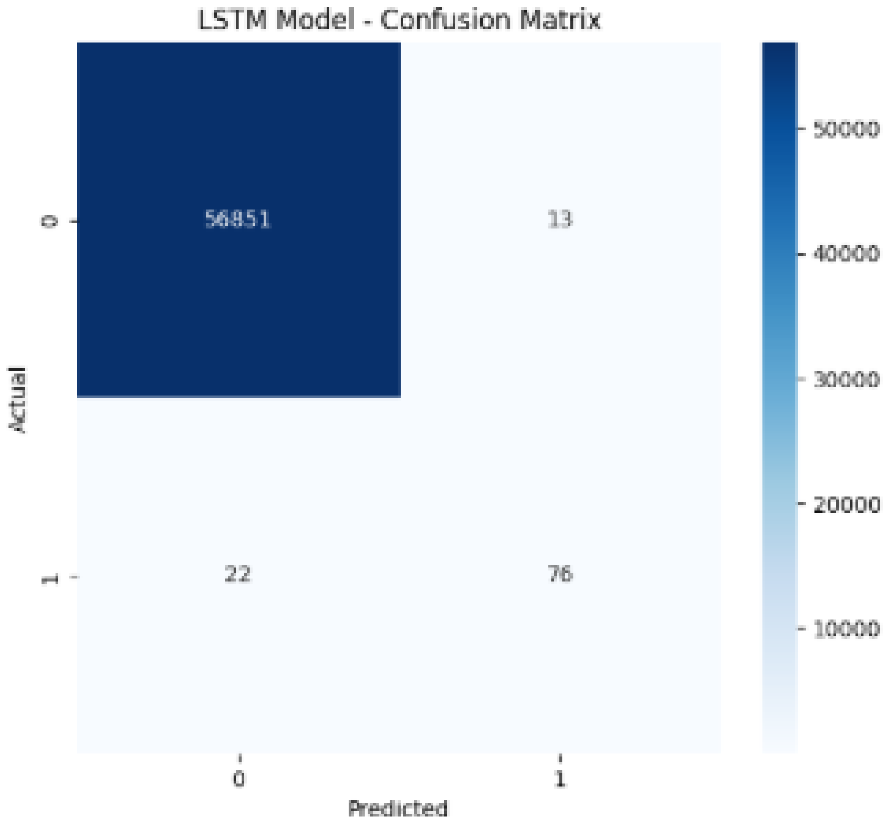


Deep Learning Models (ECCFD)

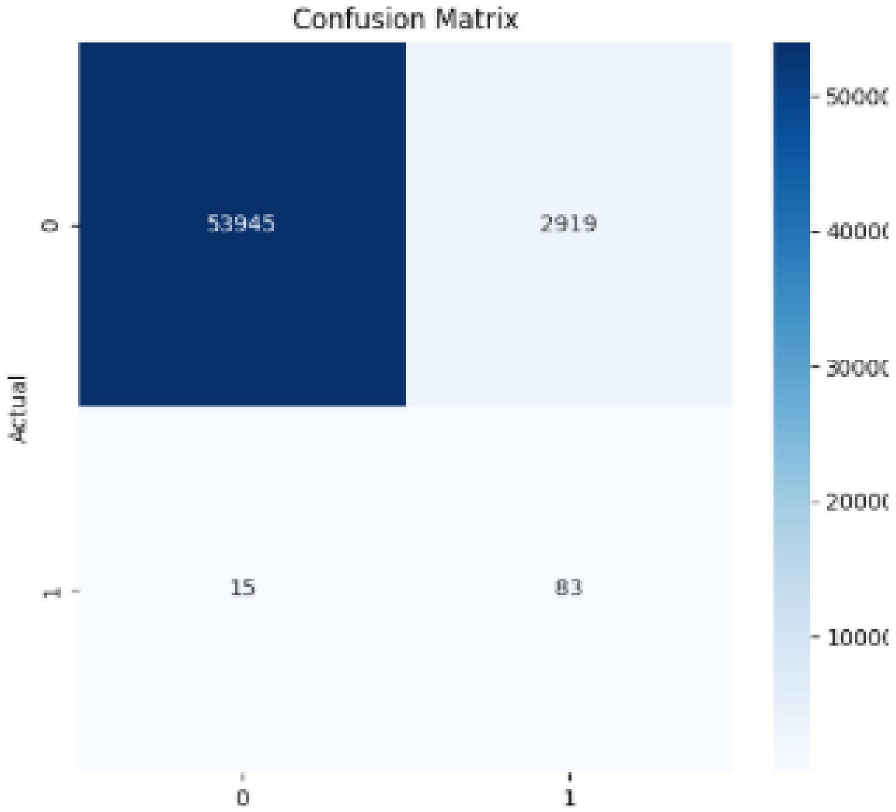
DNN

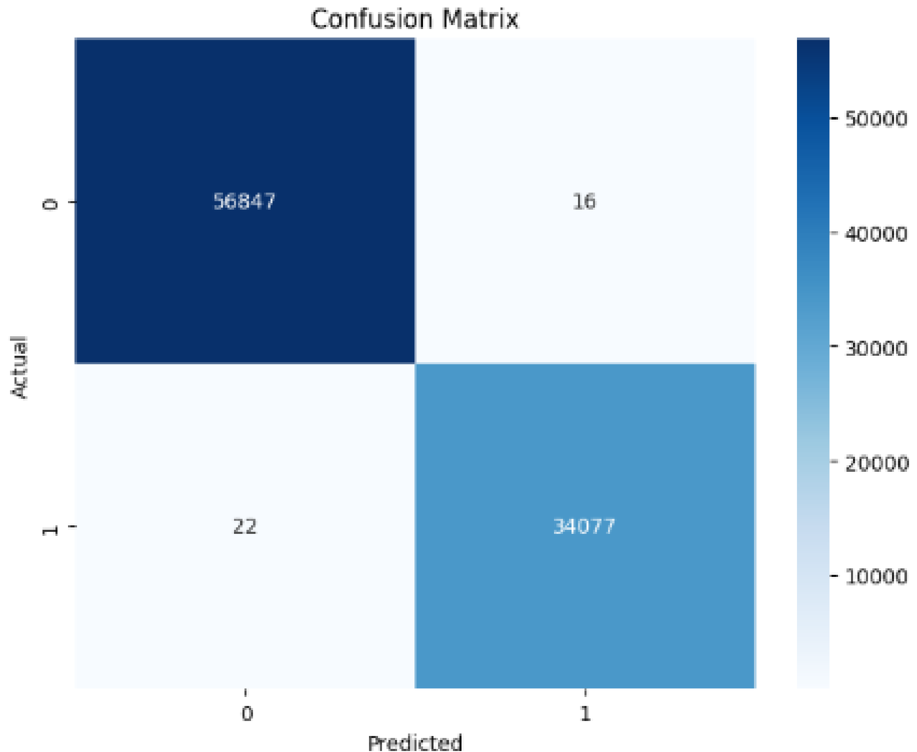


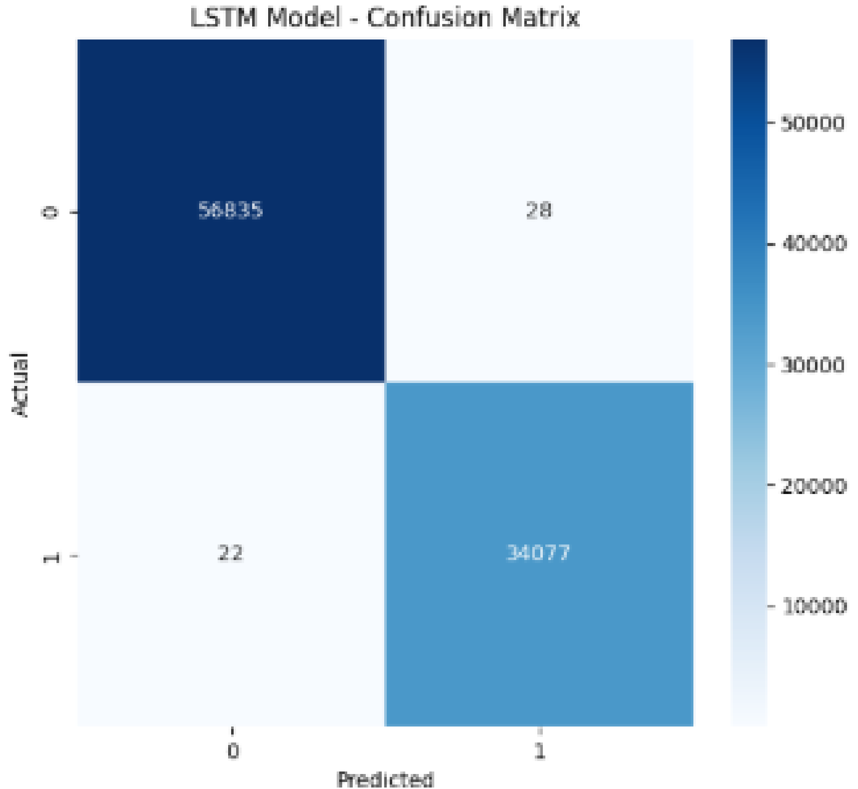
LSTM

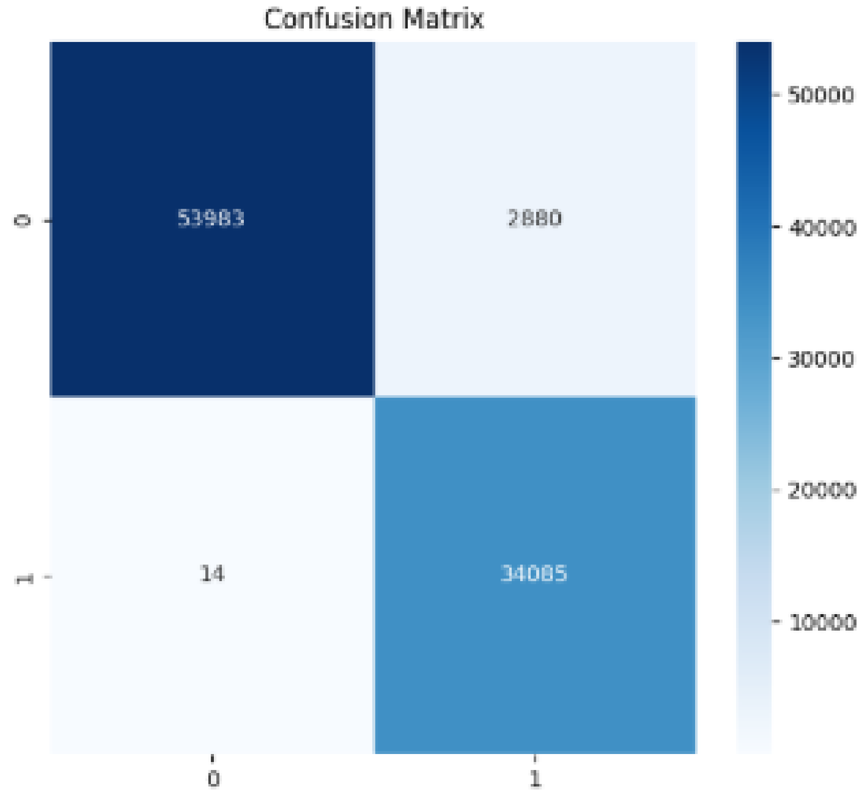


AE









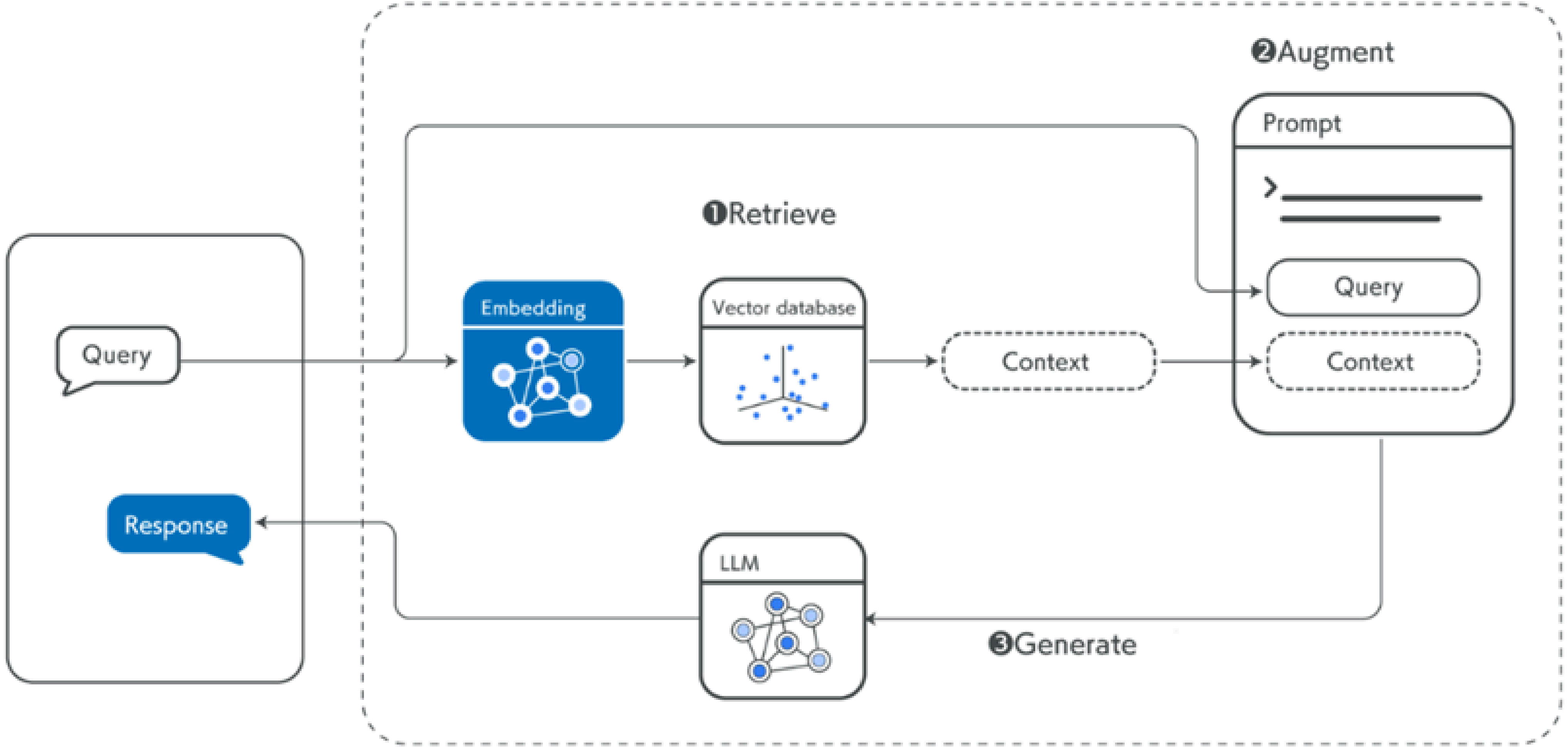
RAG Model

Flash 1.5 - 7B

FAISS VS

LangChain

Streamlit



LLM (Gemini 1.5 Flash - 7B)

- LLM by Google
- Parameters: 7 billion
- Max context length: 1 million tokens
- Latency: Low — optimized for real-time inference
- Works well with FAISS + LangChain in real-time pipelines
- Designed for long-context RAG, classification, and reasoning

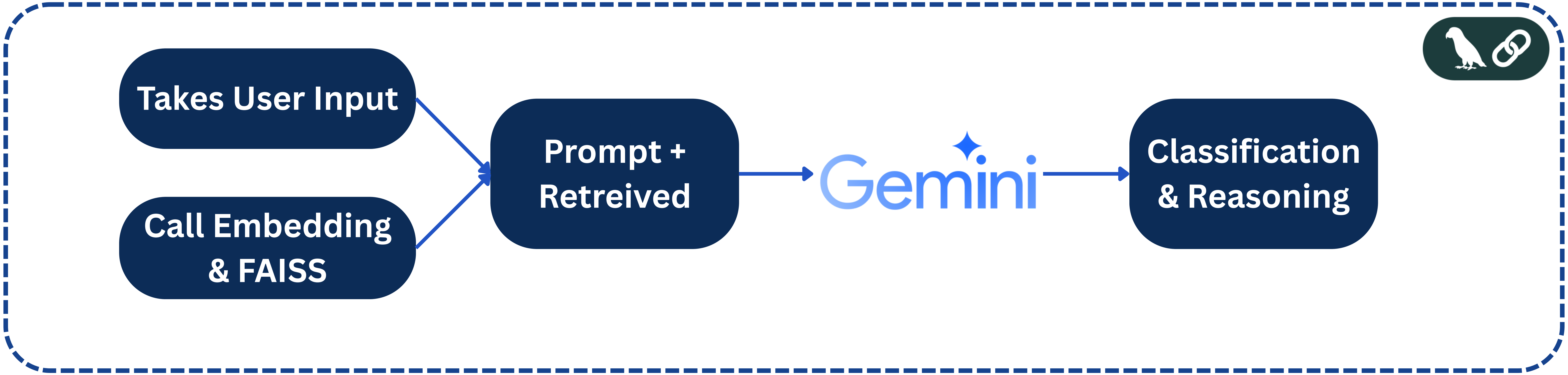
FAISS Vector Database

- We built a **custom function** to convert structured tabular data into plain text. Each transaction became a descriptive sentence combining fields like amount, merchant, user profile, and location.
- We used **Gemini embedding-001** to convert the text into high-dimensional vector representations that capture semantic meaning.
- These vectors were then **indexed using FAISS** to **enable fast similarity search** across millions of past transactions.
- The indexed vectors formed a vector database, **used during RAG to retrieve the most relevant past cases or contextual info for any new transaction.**

FAISS VDB

LangChain

- Acts as the orchestration layer that connects all components in the RAG pipeline
- Manages prompt templates, retrieval chains, and memory components
- Enables modular, maintainable, and scalable architecture for LLM applications
- Makes it easy to add logic, filters, or fallback steps without changing the core model

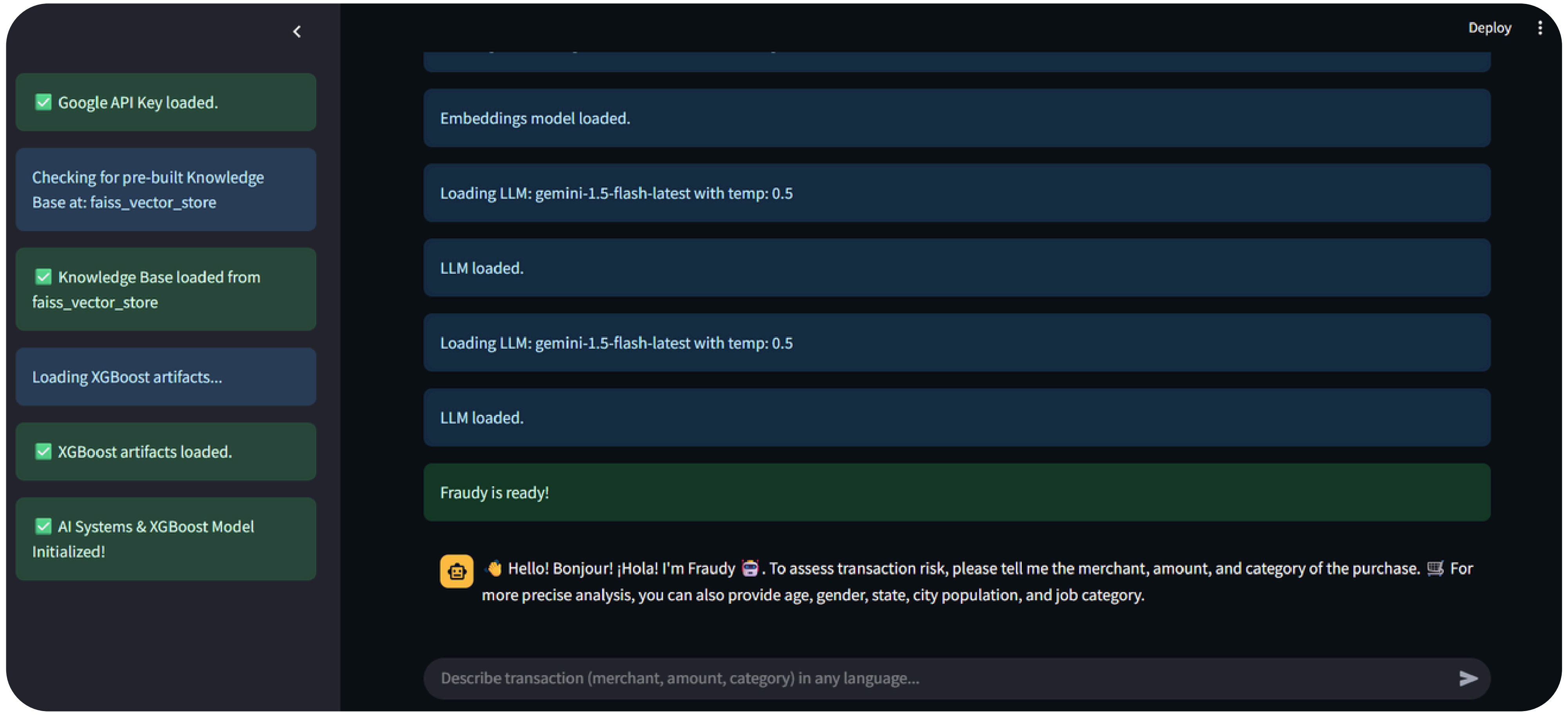


Streamlit

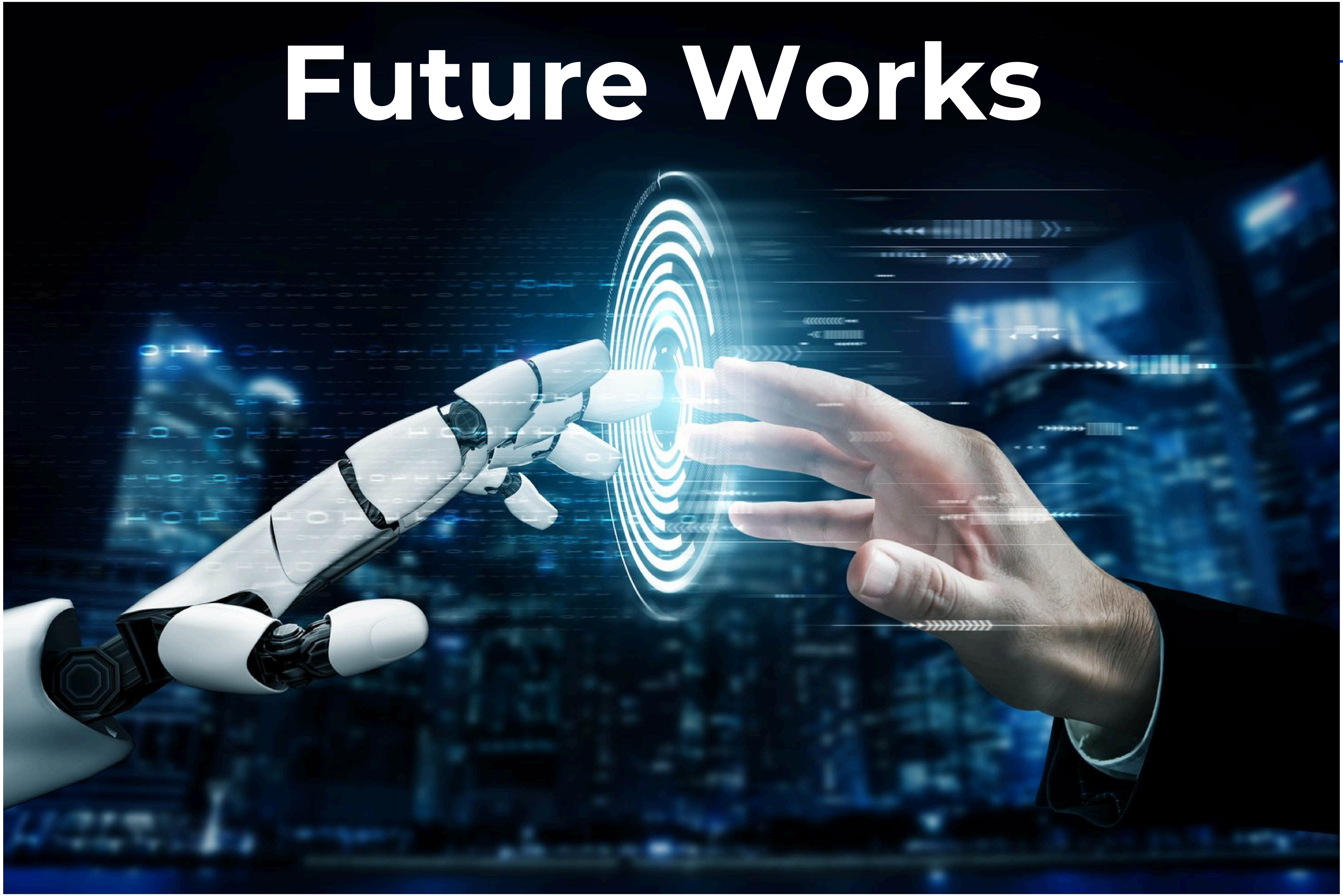
- Used to build the front-end interface of the fraud detection system
- Provided a clean, fast way to deploy the model as a web application

Main Features Used:

- Input Form for transaction details
- Real-Time Output showing fraud classification and explanation
- Live Interaction with LangChain pipeline and Gemini model



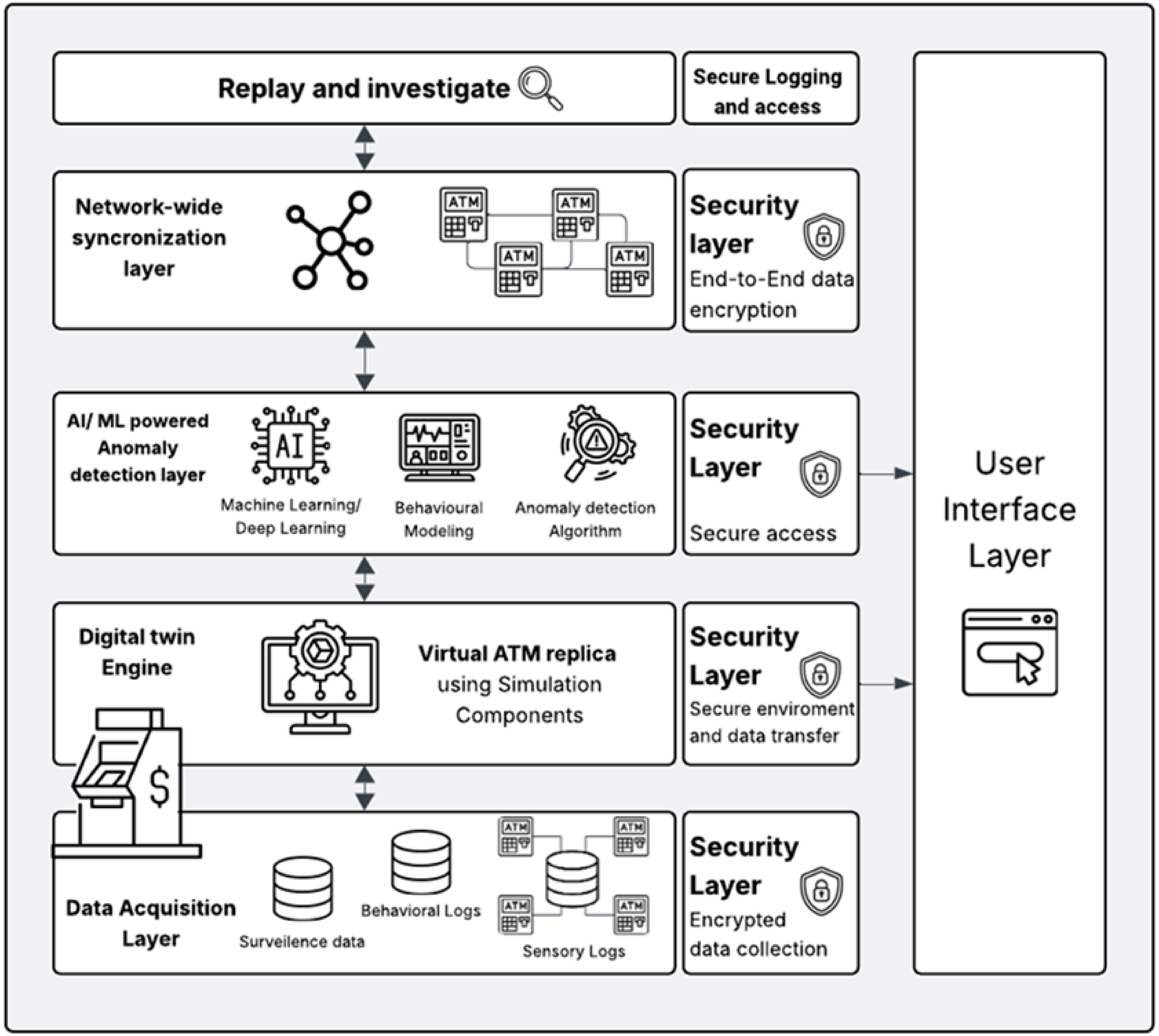
Future Works



Previous work

	Gasper vantage	NCR APTRA	Digital Twins
Monitoring	Monitor devices hardware	Manage tranactions	Simulates everything
Behavioural understanding	✗	✓	✓
Fraud detection intellegenice	Alerts for errors	Rule based fraud checks Such as: PIN failed attempets	AI-powered anomaly detection. Based on Posture, motion, timing, usage across Multiple ATMs
Enviromental awarness	✗	✗	✓
Network	ATMs are independent	ATMs are independent	Synchronizes twins to detect patterns
Real-time simulation	✗	✗	✓
Forensic capabilities	✗	✗	✓
Self Learning	Through static rules	Through software models	Continuous AI-driven learning

Digital Twins Frame work



Digital Twins in reality

Watch the video

Our Team


Gannatullah Gouda


Abdelrahman Mahmoud


Ahmed Khaled


Ali Mohamed Ali

Thank You