WinFin: Analytics Problem Statement and Leadership Requirements

By: Hari Priya Ramamoorthy

Date: Aug 17, 2024

Table of Contents

Introduction

I propose to start a business in the FinTech industry, due to its rapidly expanding customer base and collaboration opportunities with traditional financial institutions to drive technological innovations. Real-time payments are expected to surpass 70.3 billion transactions by 2025, according to ACI Worldwide and Global Data Report. Despite this digital shift, key issues remain: 60% of consumers struggle with effective budgeting, and 70% are unsure of optimal investment strategies. Additionally, a survey by Business Insider Intelligence reveals that the average user relies on over five financial apps, leading to inefficiency and complexity. **WinFin,** short for "Win in Finances," addresses these challenges by providing a unified platform that offers payment solutions, budgeting assistance, investment guidance, and strategies to boost savings. Targeted at diverse consumers, from tech-savvy professionals to unbanked populations, WinFin aims to simplify financial management and provide accessible, comprehensive financial services. CRISP-DM, **CR**oss-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining, is a robust and well-proven methodology, used to plan the steps in the data mining process. It breaks data mining problem into 6 phases: Business Understanding, Data Understanding, Data

Preparation, Modeling, Evaluation and Deployment. Now, Let’s employ CRIP-DM to understand and plan on how WinFin’s can employ data mining for fraud detection.

Based on SWOT analysis, as an emerging company, WinFin's key strength lies in its comprehensive, user-friendly financial services platform, with a strong focus on user-centric innovations. Significant opportunities include forming strategic partnerships with traditional financial firms, expanding into under-served markets, and establishing itself as a trusted leader in the FinTech industry. However, to sustain growth and maintain a competitive edge, WinFin must continuously invest in technology, user education, and market research. Most importantly, the company must address the critical challenges of navigating regulatory complexities and gain user trust. Based on SWOT implications, WinFin's leadership focus is on staying ahead of the competition by enhancing user trust and experience.

Building trust is the crucial objective for WinFin as it ensures the security and trustworthiness of the platform, which is crucial for user retention. As per Juniper Research, global ecommerce fraud losses were $41 million in 2022 and are expected to surpass $48 billion in 2023. In the U.S., 34% of consumers have been victims of fraud, with losses exceeding $3.3 billion in 2020. With these alarming statistics on fraudulent transactions, the need for advanced analytics in WinFin is paramount. Analytics can help identify patterns of fraud, predict potential threats, and proactively secure user transactions. Moreover, analytics will enable continuous improvement of the user experience, ensuring that WinFin remains competitive and responsive to the evolving needs of its customers.

Analytics as a Change Agent: Fraud Detection System

To tackle the challenge of fraudulent transactions, WinFin need a more sophisticated fraud detection system leveraging analytics that detects and predicts fraudulent activities by identifying patterns associated with fraudsters. This proactive approach to fraud detection will help prevent financial losses for customers and reduce the time spent resolving fraud-related issues, ultimately building trust and encouraging users to engage more with WinFin's services.

To identify and respond to potential fraud more effectively, the detection system requires a comprehensive dataset from internal and external sources that captures various aspects of user transaction behavior. By integrating diverse data sources, we provide the system with a holistic view of user behavior, allowing it to learn from comprehensive activity patterns and detect anomalies more accurately. This approach enhances the security and reliability of the WinFin platform, ensuring a robust defense against fraudulent activities. Let’s dive into data needed, modeling requirements, and leadership role in the analytics

Data Understanding: Transaction, User, Behavior, Financial Status, Blacklist

In building a fraud detection system for WinFin, selecting the right data is crucial for effectively identifying and managing fraudulent transactions. Key data categories include transaction data, historical fraud data, user behavior data, and contextual financial external data. Transaction data, sourced from internal and payment processors, bank APIs, and internal systems, provides essential details like transaction amounts, timestamps, and merchant information, which are critical for detecting anomalies and suspicious patterns. Historical fraud data, gathered from internal records and public datasets, offers insights into past fraudulent activities, enabling the training of models to recognize similar patterns in future transactions.

User behavior data, collected through user accounts and web analytics tools, helps identify deviations from normal activity, such as changes in login frequency, device usage, or geographical location, which can indicate fraudulent behavior. Additionally, financial dataset that encompasses users' financial status, including account balances, transaction histories, and credit card usage and external data, including public blacklists sourced from credit bureaus and fraud detection services, provide a broader perspective on transaction risks and help validate the legitimacy of transactions.

Key elements such as transaction amount and frequency, timestamps, merchant details, user behavior patterns, historical fraud trends, and external contextual data are essential for addressing potential fraud. By integrating these data sources and elements, WinFin can develop a robust fraud detection system, enhancing platform security and protecting users from fraudulent activities.

Given the data are collected from diverse sources and the FinTech industry's regulatory environment, focusing on the following key components of governance will ensure the reliability, integrity, security, and proper management of data for fraud detection is crucial. Focus on ensuring data quality by validating accuracy and consistency, securing data with encryption and access controls, and maintaining privacy through clear user consent and compliance with regulations like GDPR and CCPA. This approach builds trust and enhances fraud detection.

ML Modeling: Neural Network for Fraud Detection System

With the different sources identified, the objective is to illustrate key patterns by highlighting feature importance in fraud detection, identifying anomalous transactions, monitoring significant changes in user behavior, and flagging high-risk transactions with elevated fraud probability for classifying the fraud transactions right. For the classification problem, **Statistical Methods**, traditional mathematical approaches like

Logistic regression and Boosting Models and **Computational Methods**, that encompasses Modern intelligence techniques like Neural Network can be leveraged. My choice is to use neural networks.

The reason for choosing neural network is that it provides enhanced pattern recognition in the diverse dataset identified as it can effectively identify non-linear, complex and subtle patterns in the diverse dataset identified. Based on research papers [1], Neural Network's ability to recognize and adapt to new fraud patterns is highly effective to the evolving tactics of fraudsters than statistical methods. This ensures security and thereby enhances trust. However, as the downside, Neural networks require high Computational Resources (GPUs, cost, training time), black box nature make it difficult to explain and interpret decisions, and can easily be overfit to the training data, especially if it's not representative of real-world scenarios which can be addressed by dropout and Regularization methods.

Model Implementation & Evaluation Strategy

Building a ML Model requires collaboration from cross functional Teams. In this case, the successful implementation of Neural model for fraud detection relies on collaboration among data scientists, data engineers, IT specialists, fraud analysts, and regulatory officers, ensuring comprehensive development, integration, and compliance. With many Teams coming in, it's important to follow a software development methodology to have a plan with defined objective and timeline for each of the Team member to align and meet the data, model, business requirements. Let's discuss the steps involved in the implementation and the role of each of the Team member and software methodology to follow.

Step 1: Data Ingestion / ETL

Objective: Acquire, ensure data quality like missing values, accuracy, clean the dataset and establish a data pipeline for accessing data from all systems.

Who: Data Engineer

Tools/Techniques: Use Big Query for storage, and Alteryx for data manipulation. Build a robust data model, ensure regular data refreshes, and maintain data quality.

Software Methodology: Agile, as the requirements are defined.

Step 2: Sampling & Feature Selection

Objective: Identify key variables and prevent multicollinearity to ensure accurate model learning. Balance the dataset by oversampling fraud and legitimate transactions.

Who: Data Scientist

Tools/Techniques: Utilize t-Test for identifying top predictors and regularization methods to handle multicollinearity.

Step 3: Training

Objective: Train the model on diverse datasets to enhance learning.

Who: Data Scientist

Tools/Techniques: Use Vertex AI in GCP with Python for modeling. Apply probabilistic neural network architectures and GCP Vertex AI's resources (GPU, TPU) for model building. Perform hyper-parameter tuning, including adjustments to learning rate, optimizer, batch size, and network layers.

Step 4: Evaluation

Objective: Assess model performance to ensure to detect as much as fraud transactions proactively as possible.

Who: Data Scientist and Fraud Analyst

Key Metrics: Focus on **sensitivity**, measuring the proportion of correctly predicted fraudulent transactions. Ensure high true positive rates to effectively identify fraud. This Ensures fraudulent transactions are correctly identified and maximize the true positives.

Software Methodology (Step-2 to Step 4): Iterative methodology for modeling phase, as it allows for modeling phase, as it allows for ongoing experimentation, adjustments, and optimizations, accommodating the unpredictable nature of training time and evolving fraud patterns.

Step 5: Deployment

Objective: Once model is finalized, Integrate the model into the cloud and connect it to the fraud detection system.

Who: Data Engineer, Data Scientist and IT Specialist

Actions: Ensure smooth integration and monitor real-time performance for prompt fraud detection and mitigation.

Step 6: Model Monitoring

Objective: Review and retrain the deployed model every 3-6 months.

Who: Data Scientist

Tools/Techniques: Automate monitoring based on performance thresholds or schedules to maintain consistent model performance.

Software Methodology (Step 5-6): Agile, as the requirements are defined

Let's dive on steps to integrate the built ML model with the transaction system.

Deployment: Real-Time Integration of the Fraud Detection Model

The end goal of fraud detection model is to integrate the with the transaction processing system to check the transactions in real-time if it's fraud or not. Once the model is finalized based on "sensitivity", below are the steps that requires collaboration and plan for real time deployment:

1. Plan on Integration with Existing On-Prem Systems: Data scientists & IT Analyst

API Development: Create APIs to integrate the model with existing transaction processing system which is on-prem for better security.

System Compatibility: Ensure the model's output is compatible with the components of IT transaction platform.

2. Plan on model Environment for scaling and Real-Time Processing: IT Analysts

Infrastructure: Set up the necessary infrastructure for deploying the model, such as cloud services GCP Vertex AI.

Scaling: Plan for scaling the deployment to handle varying volumes of transactions and data.

Latency: Ensure the model can process transactions in real-time or near-real-time to minimize fraud detection delays.

Batch Processing: For non-real-time scenarios, I'd set up batch processing mechanisms.

3. Model Monitoring: Data Scientist, Fraud Analyst

Continuous Monitoring: Continuously monitor the model's performance, including accuracy, false positives/negatives, and processing times. Regularly check for model drift or changes in data patterns that might affect performance.

4. Documentation and Training: Data Scientist, IT

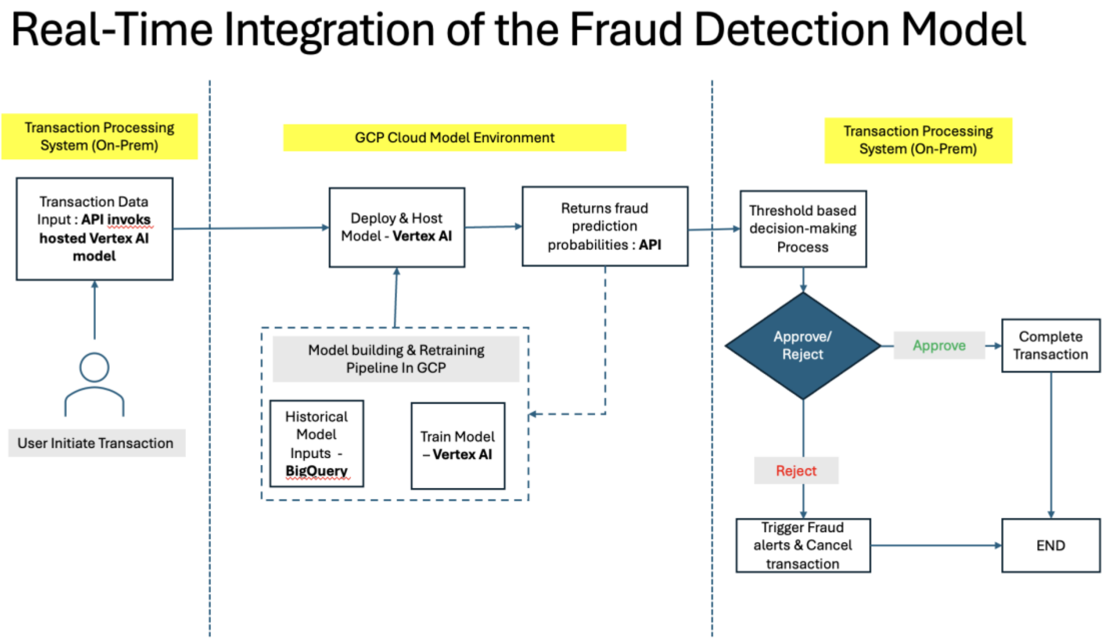
Documentation: Create comprehensive documentation for the model, including its functionality, integration points, and maintenance procedures. Train relevant personnel on how to use, monitor, and maintain the model.

5. Rollback Plan: Leaders, IT Team

Contingency: Develop a rollback plan in case the deployment encounters critical issues, including steps to revert to the previous system.

Below flowchart depicts the architecture of a Fraud Detection model integrated in WinFin transaction Platform.

Figure 1


Real-Time Integration of the Fraud Detection Model
Transaction Processing System (On-Prem)
GCP Cloud Model Environment
Transaction Processing System (On-Prem)
Transaction Data Input : API invoks hosted Vertex AI model
Deploy & Host Model - Vertex AI
Returns fraud prediction probabilities : API
Threshold based decision-making Process
Model building & Retraining Pipeline In GCP
Historical Model Inputs - BigQuery
Train Model – Vertex AI
User Initiate Transaction
Approve/ Reject
Approve
Complete Transaction
Reject
Trigger Fraud alerts & Cancel transaction
END

1. **User initiate Transaction:** It begins with a user-initiated transaction in an on-premises transaction processing system, where data is sent via an API to a fraud detection model hosted on Vertex AI in GCP.
2. **Model Training & Prediction:** The model, trained using historical data from Big Query, returns fraud prediction probabilities through the API.

3. **Decision Making System:** A threshold-based decision process then either approves or rejects the transaction. If rejected, fraud alerts are triggered, and the transaction is canceled.
4. **Continuous Retraining:** The system continuously retrains the model using new data to improve its accuracy and effectiveness.

Hence, the deployment process involves ingesting transaction data through API, processing it using a model deployed on Vertex AI, using the fraud prediction probability through API for decision making, storing results in Big Query for model retraining, and monitoring the system for continuous improvement.

Leadership Requirements

It's the Fraud detection system's analytics leader responsibility to drive innovation, collaboration across the Teams to align with their business objectives and data requirements, drive the data collection process from different system, setup timely check-ins and get inputs and align with different stakeholders. Below are the leadership requirements:

1. **Drive Innovation & Collaboration:**

- Align teams with business objectives and data requirements.
- Oversee data collection, set up timely check-ins, and gather input.
- Clearly communicate how analytics projects align with business goals to increase buy-in, reduce resistance, and foster collaboration.

2. **Guide & Involve Experts:**

- Assist in decision-making when challenges arise, ensuring the right experts are involved.
- Collaborate with data scientists, fraud analysts, and IT analysts to build thorough test cases, including stress tests for latency, prediction discrepancies, and real-time processing.

3. **Use Appropriate Software Methodologies:**

- Adopt Agile for well-defined phases like ETL and testing.
- Implement an iterative approach for complex modeling phases, allowing ongoing experimentation, adjustments, and optimizations. This method accommodates the unpredictable nature of training time and evolving fraud patterns.

4. **Promote Data-Driven Decision-Making:**

- Encourage continuous learning in data literacy.
- Ensure decisions are based on insights, leveraging diverse perspectives for comprehensive, data-driven solutions.

WinFin Fraud Detection: CRISP-DM Summary

Concluding the discussion so far,

- **Business & Data Understanding:** WinFin aims to establish itself as a trusted leader in the FinTech industry by focusing on staying ahead of the competition, building user trust, and enhancing user experience. By focusing on security and user-centric innovations, WinFin can establish itself as a trusted leader in the fintech industry.
- **Data Preparation:** To effectively identify and predict fraudulent activities, the model requires a comprehensive dataset including —transactional, user, financial, behavioral, and external sources
- **Modeling:** Neural networks are chosen for their superior ability to recognize complex, non-linear patterns and adapt to evolving fraud tactics, despite requiring significant computational resources.
- **Evaluation:** The objective of WinFin's detection system is to detect as much as fraud transactions proactively as possible. Sensitivity is the key that ensures fraudulent transactions are correctly identified and maximize the true positives.
- **Deployment:** Deployment integrates the cloud-based model to includes API development and system compatibility. Ensure real-time processing and scalability. Monitor, retrain, document, train, and prepare rollback plans.

By leveraging advanced analytics, effective leadership, collaboration among Data engineer, data scientists, IT Team, Fraud analysts and the right methodologies, WinFin can successfully integrate a sophisticated fraud detection model into its transaction system, enhancing security, user trust, and overall business impact.

References

- https://fintechmagazine.com/articles/fintech-in-2024-the-big-questions-answered
- https://www.mckinsey.com/industries/financial-services/our-insights/fintechs-a-new-paradigm-of-growth
- https://plaid.com/resources/fintech/fintech-trends/
- National Foundation for Credit Counseling (NFCC) Survey
- ACI Worldwide and Global Data
- https://www.yodlee.com/6-payments-trends-2021
- https://www.sciencedirect.com/science/article/pii/S0167404815001261#s0145
- https://www.sciencedirect.com/science/article/pii/S0167923610001879?ref=cra_js_challenge&fr=RR-1#s0060
- https://kpmg.com/pl/en/home/insights/2021/02/comprehensive-fraud-detection-and-verification-process.html#:~:text=KPMG's%20approach%20includes%20descriptive%2C%20customer,data%20from%20a%20wider%20context.

- https://www2.deloitte.com/content/dam/Deloitte/xe/Documents/risk/Fraud-analytics.pdf
- https://www.sciencedirect.com/science/article/pii/S0167404815001261#s0020
- https://www.sciencedirect.com/science/article/pii/S0167923610001326
- https://www.kaggle.com/discussions/general/329792
- https://b2b.mastercard.com/news-and-insights/blog/ecommerce-fraud-trends-and-statistics-merchants-need-to-know-in-2024/
- https://www.cnbc.com/video/2021/01/27/why-credit-card-fraud-hasnt-stopped-in-the-us.html
- https://consumer.ftc.gov/all-scams/debt-credit-scams
- https://www.sciencedirect.com/science/article/pii/S2772662223000036
- https://cloud.google.com/blog/products/data-analytics/how-to-build-a-fraud-detection-solution