CASE STUDY

N.4 | MARCH 2025

INTELLIGENCE PLATFORM FOR FLAGGING FRAUDULENT FINTECH APPS

This case study outlines the development of an intelligent platform designed to identify fraudulent fintech applications in India under collaboration with the Reserve Bank of India and the Reserve Bank Innovation Hub, and the suptech solution provider Winnow Technologies. The solution integrates advanced fraud detection algorithms, behavioural analytics, anomaly detection, and other AI driven techniques to proactively flag suspicious applications, enhancing regulatory oversight and consumer protection in India's digital financial ecosystem.

Cambridge Centre for Alternative Finance
UNIVERSITY OF CAMBRIDGE Judge Business School

CAMBRIDGE SUPTECH LAB

www.cambridgesuptechlab.org

Project overview

With the advent of technology and the widespread penetration of smartphones globally, obtaining loans has become remarkably easy, often achievable within minutes. In India, the fintech revolution has experienced exponential growth, with online loan applications significantly advancing financial inclusion, particularly reaching unserved and underserved customers.

However, this rapid growth has also led to the emergence of numerous fraudulent fintech applications. Unscrupulous entities are exploiting these online platforms to defraud customers in many ways. Some applications impose predatory interest rates, while others access private customer information, resorting to threats and blackmail. Additionally, a few applications have been implicated in money laundering activities, identified in the RBI project charter for this project as posing significant risks to Indian national security.

In response, the Reserve Bank of India (RBI) initiated a crackdown on such fraudulent applications, collaborating with Google and Apple to identify and ban many of these apps. Despite these efforts, malicious actors continually flood the app stores with new fraudulent applications.

The RBI recognized that to fully address this persistent and large-scale issue requires an intelligent platform designed to flag fraudulent fintech applications as soon as they are launched on app stores. This ambitious project, conducted with RBI's Fintech Department and the Reserve Bank Innovation Hub (RBIH), leverages web scraping, data processing, artificial intelligence (AI), machine learning (ML), and natural language processing (NLP) to identify and flag suspected malicious applications. Additionally, it provides a dashboard for real-time monitoring with access controls.

Following preliminary specifications developed by members of RBI's Fintech Department through the Cambridge SupTech Lab's online SupTech Frontiers capacity building and education programme, the RBI and the RBIH further engaged the Cambridge SupTech Lab to assist in identifying and selecting a suitable vendor for the development of a prototype that will help achieve the project's objectives.

Project partners

- **RBI**: The apex bank in India, licences, controls, and regulates the financial and banking system in India, including banks, non-banking financial companies (NBFCs), and cooperative banks.
- **RBIH**: The Innovation Hub is a wholly owned subsidiary of the RBI that was set up to promote and facilitate an environment that accelerates innovation across the financial sector.
- **Cambridge SupTech Lab**: The Cambridge SupTech Lab accelerates the digital transformation of financial supervision and supervisory agencies. The Lab delivers world-class online leadership education, experiential training, ground-breaking research, market intelligence, new analytical frameworks, innovative digital tools, and cutting-edge suptech applications. The Lab's Application Incubation programme, which developed this prototype, is a suptech accelerator for financial authorities and technology vendors to co-create and deploy cutting-edge, scalable suptech prototypes and applications.
- **Winnow Technologies**: Winnow is a suptech vendor specialising in web-based data mining, natural language processing, and advanced analytics.

Challenges in the pre-existing frameworks

- **No controls on new apps**: No permissions or approvals are required to publish an app on the Google Play Store. Any developer can publish a financial lending app for consumers, even if not tied to a regulated entity.
- **Delayed response time for flagging apps**: The existing manual processes for investigating and flagging fraudulent apps result in those being removed after they have impacted many users. By this time, considerable damage has already been done.
- **Achieving real-time monitoring of apps is impossible without suptech**: There is no real-time monitoring or flagging of apps at this scale without digital transformation of the currently manual processes and incorporation of suptech into critical supervisory processes.

Key features

- **Advanced web scraping techniques** to extract and stage novel data from financial lending apps and their user reviews on the Google Play Store, processed and stored in a manner that allows for automated analysis to augment supervisory insights.
- **App discoverability tool** that monitors new and existing listings on the app store, automatically filtering to assess only financial lending apps in the Indian store, specifically.
- **Advanced analytics tools in AI/ML** to recognize signals across app store reviews and comments that raise flags of potential fraud.
- **An overall fraud probability score** that can be fine-tuned over time, with a ML model that adjusts with more data and aggregates across apps.

- **An automated report** that presents fraud probabilities along with disaggregated features and their contribution to the probability, for the sake of interpretability, further interrogation by supervisors, and extensibility and potential incorporation into other models and notification systems.

Benefits

- **Real-time Monitoring and Flagging**: Using real-time monitoring to flag potentially fraudulent apps provides financial supervisors novel insights to potentially prevent consumer harm before many people are defrauded with predatory interest rates, harassment or blackmail, or misuse of data.
- **Data Protection**: Detecting misuse of consumer data early informs a swift supervisory response to prevent further spread and scaling of the set of risks. It will help prevent fraud and cyber crime and compliance with data protection laws and regulations. Data breaches can lead to significant financial losses and undermine the financial system's stability.
- **Consumer Satisfaction**: By regulating and monitoring lending apps, RBI ensures consumers access innovative and user-friendly financial products. This promotes financial inclusion and helps meet diverse financial needs. It also promotes fair lending practices to prevent predatory lending. This includes capping interest rates and ensuring consumers are not harassed for repayments. Hence, this project will improve the consumer's trust in the financial sector and help navigate options to work with legitimate fintech apps.

1. BACKGROUND AND SUPERVISORY CHALLENGES

The RBI is India's central bank, established on April 1, 1935. It regulates and supervises the country's banking system, manages its currency, and ensures monetary stability. The RBI plays a crucial role in safeguarding the nation's and its citizens' interests by maintaining financial stability and promoting economic growth in India. The RBI also provides deposit insurance, promotes fair banking practices, and ensures access to financial services for all.

The RBI focuses on several key areas. Monetary policy is formulated and implemented to maintain price stability and ensure adequate credit availability. The RBI manages currency, overseeing the design, production, distribution, and circulation of Indian currency notes and coins. Banking regulation is another core function, with the RBI supervising and regulating banks to ensure their financial health and stability. The RBI also manages the country's foreign exchange reserves and regulates foreign exchange transactions. Furthermore, it promotes the development of efficient and stable financial markets. Protecting customers' interests is paramount, encompassing deposit insurance, consumer protection, fair banking practices, and financial inclusion.

The RBI has actively promoted fintechs and online lending apps because they can reach underserved populations in remote areas without access to traditional banking services. They leverage technology to provide financial services to the unbanked and underbanked, promoting financial inclusion. Digital platforms make financial services more accessible to a broader range of people, regardless of location or income level. This helps to bridge the gap in access to finance. They also bring innovative solutions and technologies to the financial sector, improving efficiency and customer experience. They offer services like digital payments, online lending, and personalized financial advice, often at lower costs than traditional banks. Fintechs introduce healthy competition in the financial market, encouraging traditional banks to innovate and improve their services. This benefits consumers.

But with the advent and advancement in technology and the proliferation of online lending apps, bad actors have emerged in the scene, and the occurrence of frauds has increased. Online scammers follow non-transparent methods, collect predatory interest rates, cause harassment through harsh recovery measures, and unauthorised use of personal data. This has affected the general trust in the financial system, especially on the fintech space.

The proliferation of fraudulent apps in the Indian market is of particular concern, not just because of the size of the market but also because of its social impact. With a population of nearly 1.5 billion people, the potential size of the financial market and the subsequent implications of fraud and other malicious practices should not be underestimated.

2. PROJECT CONCEPTUALISATION AND INCEPTION

RBI enrolled a team of four interdisciplinary supervisory experts from the Fintech Department as part of the Lab's Capacity Building and Education offerings. Through this course, the team developed a capstone entitled "An intelligence tool that can greenlight fintech apps in the Appstore as a preventative anti-fraud measure," which recognized the need to address several key novel and newly magnified risks arising in the Indian financial sector as well as globally.

The number of fintech apps was noted in the Project Charter notes provided by the RBI, which initiated this project to increase, with many customers affected by fake fintech applications. For example, online scammers have been observed to follow non-transparent methods, collect predatory interest rates, cause harassment through harsh recovery measures, and engage in unauthorised use of personal data. This was observed to risk breaking the public's trust in fintech apps in general and, in turn, harming inclusion. To address this issue, the team sought to collect metadata related to fintech apps from smartphone app stores, data from social media apps and other data sources to identify and flag malign actors for pre-emptive review by supervisors.

The Lab recognised the RBI team's capstone as having high potential for global impact on Consumer Protection and Market Conduct Supervision. Moreover, a strong base of past research relating to some components of such an approach (e.g. validation of modelling fraud on "slice-in-time" batched data by the University of Zurich and Innovations in Poverty Action served as an initial proof of value, exemplifying a clear opportunity for further investment in a suptech (supervisory technology) solution that provides this value and beyond on a live, ongoing basis. With financial support from the Bill and Melinda Gates Foundation and International Finance Corporation's India facility and additional partnership from the RBIH, the Lab selected the capstone for further support through the Lab's Application Incubation facility.

The subsequent prototype solution that was developed is an intelligent system to flag to supervisors any instances of potential fraud in fintech apps based on metadata from (i) purported lending apps from app stores, (ii) other concerned apps and (iii) other relevant sources within the system, to identify patterns and flag malign actors pre-emptively.

This robust suptech solution to detect fraud is expected to contribute to the essential protection of financial consumers, continued growth, and reliability of the financial system. The system has been built extensible to anticipate future required capabilities for RBI as well. For example, the solution can be integrated to serve as a basis for applications that inform the public (e.g., a website or smartphone-based service that can act as an anti-fraud layer when consumers consider installing any relevant fintech app on their device).

Finally, the technology partner/vendor was required to propose a solution for which the underlying technology and code could ultimately be transferred to RBI upon completion of the project, per the expressed intention of RBI to transition to in-house maintenance and development. To this end, the entirety of the solution was transferred to the RBI's in-house team at the end of the prototyping stage. The vendor maintained and shared up-to-date documentation with RBI and the Lab, reflecting the results of such conversations throughout the project.

3. LEAN VENDOR PROCUREMENT AND SELECTION

Upon collaborating to receive executive approval and co-creating the Request for Proposal (RFP) with the RBI and RBIH in July 2024, the Lab invited applications from competent suptech vendors through a publication of the RFP to a global audience, along with targeted dissemination via the Lab's suptech vendor database.

Responses to this globally competitive RFP passed through a rigorous selection procedure with Subject Matter Experts (SMEs), suptech specialists, and specially created scorecards with criteria with varying weights: relevant experience (60%), technical and managerial expertise (30%), and adequate resourcing (10%). Subsequently, the proposals were further assessed based on topic responsiveness (65%), execution plan (25%), and innovative approach (10%). Winnow Technologies was ultimately selected as the vendor to develop the prototype for the RBI.

Following a no-objection clearance from partner financial authorities, the Lab conducted the necessary due diligence. This included addressing legal issues related to data sharing and storage, intellectual property licensing, and public procurement. All these aspects were formalised in a project agreement between the University and Winnow, including a non-disclosure agreement, during contracting in September 2024. Once contracted, the Lab provided project management and specialised technical support throughout the development and testing phases of the working prototype.

4. WORKING PROTOTYPE AGILE DEVELOPMENT

During the procurement process, Winnow designed and produced specifications for the prototype in a manner that centred around the supervisor's needs. Following contracting, Winnow proceeded to iterate on these specifications, engaging in an agile manner to ensure they were building a tool tailored to augment existing regulatory frameworks and actions.

Through this iterative specification refinement and the agile development process described further below, Winnow has built several unique features, which are illustrated in Figure 1 and described further in the section below:

1. An app discoverability tool that filters out only apps in the Indian store, specifically financial lending apps. It ignores other apps in the finance category, such as ATM finders. This focus helps address the most pressing supervisory issue.
2. A scraper that gathers relevant reviews and metadata from the discovered apps.
3. Predictive models that are initially trained on large natural language data sets, then fine-tuned against labelled historical fraud data, to predict the likelihood of fraud for newly discovered apps based on the indicators gathered by the scrapers.
4. Reporting of granular indicators of fraud probability scores rather than a binary yes or no answer.

The fundamentally critical component of Winnow's system is the app discoverability tool, which operates on a predetermined schedule (e.g., weekly) to systematically identify and analyse newly uploaded apps on the Google Play Store. This feature of Winnow's tool continues monitoring these apps until they can be determined about their legitimacy.

The system's scraping process then captures the most current and relevant publicly available data associated with these apps, including app descriptions, user reviews, ratings, publisher details, and other accessible metadata. This data is then processed by Winnow's predictive models, which have been trained to identify fraud indicators, such as misleading app descriptions, artificially inflated user ratings, or suspicious patterns in user feedback.

Predictive models employed by Winnow are designed to continuously improve their detection capabilities by learning from new data as it becomes available. These models assess—and, importantly, allow for disaggregated reporting and interrogation of—numerous factors that contribute to fraudulent behaviour, including but not limited to the app's publisher history, user sentiment trends, review inconsistencies, and data safety practices. By integrating these models into the suptech solution, the RBI is ultimately equipped with an innovative solution that proactively identifies high-risk apps, thereby safeguarding consumers and maintaining the integrity of India's digital financial ecosystem.

The model's outputs are automatically reported as a fraud likelihood, along with the significance of various factors in predicting that likelihood. This approach is dynamic and can be fine-tuned over time, meaning it can adjust its scores as more data is aggregated across apps. Moreover, it empowers supervisors to address the most pressing issues first (e.g., apps with a 50% or higher probability of being fraudulent) while continuing to monitor those with a lower score. Finally, it accounts for natural language nuances and filters for general complaints (i.e., "This app sucks") versus reviews that are higher indicators of fraud.

Winnow has built two unique features:

An app discoverability tool that filters out only apps in the Indian store, specifically financial lending apps. It ignores other apps in the finance category, such as ATM finders. This focus helps address the most pressing supervisory issue.

Fraud probability scores rather than a binary yes or no answer. This approach has several different benefits:

- It is dynamic and can be fine-tuned over time, meaning it can adjust its scores as more data is aggregated across apps.
- It empowers users to address the most pressing issues first (e.g., apps with a 50% or higher probability of being fraudulent) while monitoring those with a lower score
- It more finely accounts for natural language nuances and filters for general complaints (i.e., "This app sucks") against reviews that are higher indicators of fraud.

Details of data handling across the suptech spectrum:

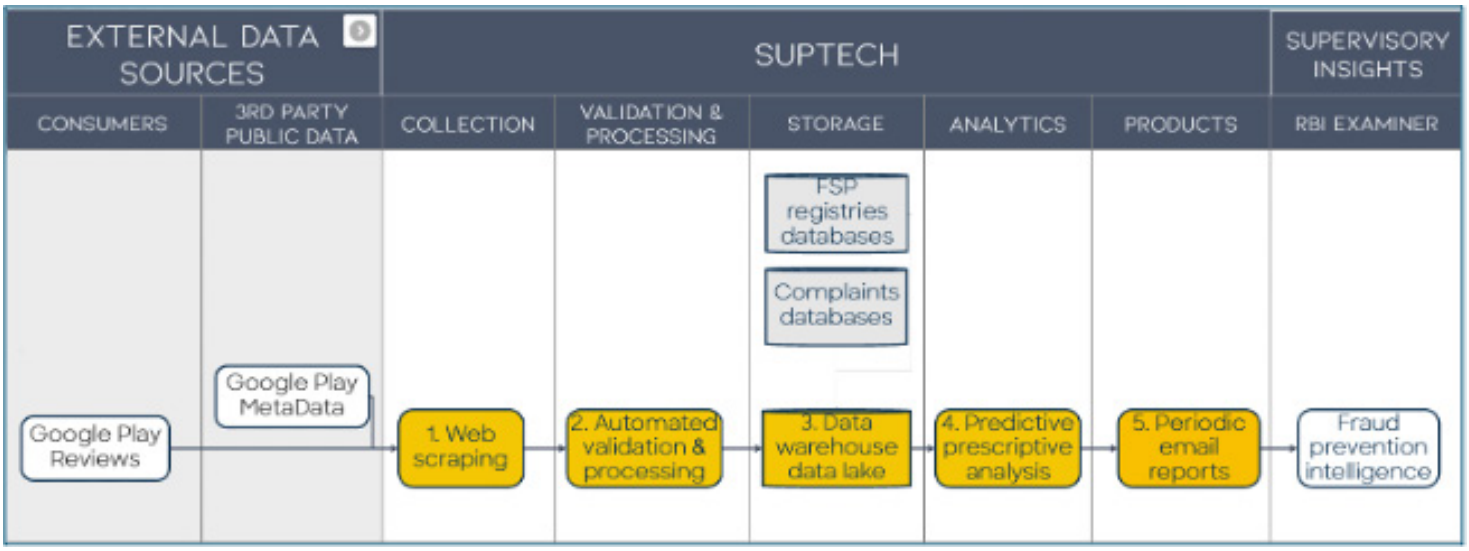
1. Web scraping includes a backend module and scrapers that initiate discovery and collect data.
2. Automated validation and processing include filtering based on keywords and phrases.
3. Data warehousing stores data from web scraping collection augmented with data from AI classification and topic modelling and peripheral data sources such as complaints data and agency register of regulated entities.
4. Analysis includes advanced AI techniques of topic modelling, categorisation, and fraud detection.
5. Data products include a notification module that produces a periodic report emailed to the agency's examiner.The working prototype's application development was divided into six sprints as detailed below:

Sprint 1: Data review

The data review sprint was designed to address the following project goals:

1. Identify training data sources
2. Check training data sources for viability
3. Gain an understanding of previous, manual fraud detection mechanisms, and which can help inform the ML models

FIGURE 1. DATA FLOW MODEL OF THE INTENDED PROTOTYPE



4. Obtain training data

The data review sprint was the most challenging part of the project. In any AI/ML system, training data is imperative for the model's success. The more closely the data mirrors that which will be collected and analysed, the more accurate the final models will be.

Unfortunately, a substantive batch of training data from previously flagged fraudulent apps was unavailable to Winnow, the Lab or RBI. While the stakeholders collectively tried to find alternative sources of data including other complaints websites, working documents from FACE India, lists of removed apps from the Google Play Store, and a few other minor links, few were appropriate for training the models.

There was an initial list of apps identified by FACE that were known to be fraudulent, however they were taken down from the app store. Later in the project, Winnow identified four apps that, upon second review, remained live on the app store. It remains unclear if Google reinstated them from when they were first considered or if there were other reasons for this inconsistency. The project team used these apps for the validation and testing discussed in more detail below.

Sprint 2: Tailoring the collection tool

For the tool to be useful to RBI into the future, not just during the prototype, it must be able to identify the apps and collect data. This is called the relevance engine, and the filters used to determine relevance are a pillar component of the tool.

To comply with the filtering requirements the apps must:

1. Be available for download in the Indian market
2. Come from the "finance" category on the Google Play Store
3. Be in the "financial lending" subcategory

Once the potential apps have been identified through the API search, each result is enriched with additional supporting data. This enrichment process involves retrieving the full app listing from the App Store, which includes detailed metadata such as the app's description, developer information, user ratings, and download statistics. This additional data provides crucial context for the subsequent stages of analysis.

The success of this collaboration and iterative process with RBI was particularly evident during this sprint. Like many tech problems, this success proved to be more challenging than initially thought. Despite robust filters, the first pulled list of newly uploaded apps to the Indian Google Play Store had an accuracy of just 54%.

The second pull of newly uploaded apps improved accuracy to 75% but was still insufficient. The reason for this was server location: While many apps are visible in a particular country's store, it does not mean it is downloadable. As Winnow's servers are based in the US, this was not apparent, until locally reviewed in India.

Finally, addressing that issue, Winnow improved the accuracy of new app pulls to 100%. This robust result should remain equally accurate once the tool is deployed in-country.

Sprint 3: Data collection

Once confirmed that the collection tool identifies the relevant apps, scrapers are deployed to gather all relevant data from the apps to be monitored. The scraper focuses primarily on user-generated content, pulling reviews left by customers on the app store and any responses from the app's publisher to those reviews. This review data is crucial for

detecting patterns of fraudulent behaviour, as it provides direct insight into user experiences and potential warning signs of misconduct or deceptive practices. These reviews and responses are then analysed and classified in subsequent stages to determine the app's overall risk profile. This is a standard part of Winnow's proprietary offering.
To ensure that data collection is current and reflects the latest developments, the search of apps is limited to those updated or published within a specific period. Typically, this time range spans from the present back to the time of the last data scrape, usually within the past 24 hours. This approach guarantees that analysis is based on the most up-to-date data available, capturing new and recently added comments and reviews that may exhibit emerging patterns of fraudulent behaviour.

Winnow's tool is currently able to collect a million comments, reviews, and other data points for analysis per 24-hour period for the prototype; this is, however, scalable, and a production version of this tool should be able to increase the volume of data collected (as the list of monitored apps grows). 608,345 data points were analysed for the Proof-of-Concept report to arrive at probability scores. Naturally, there is a limit based on how many new comments and reviews have been added.

These results are consistent with Winnow's capabilities and believed sufficiently fast to identify potential fraud and empower supervisors to regulatory action. It is important to note that comments and reviews are the key ways in which the tool identifies potential fraud. It may take some time after the app is published until sufficient data is available to offer an accurate probability score.

Sprint 4: Data management validation

Winnow used a collection tool to identify relevant apps, and scrapers gathered data primarily focusing on user-generated content, such as customer reviews and publisher responses. This data helped detect fraudulent behaviour patterns by providing insights into user experiences. Reviews and responses were analysed and classified to assess the app's risk profile, forming a key part of Winnow's proprietary pipeline.
To ensure up-to-date data collection, the tool limited its search to apps updated or published within a specific period, typically the past 24 hours. This approach captured recent comments and reviews, identifying emerging fraudulent patterns.

The tool's speed and capability were sufficient for identifying potential fraud and supporting regulatory actions. It relied on user comments and reviews to detect fraud, improving accuracy as more data became available.

Sprint 5: Data management validation

In this phase, Winnow enriched raw, unstructured data to provide actionable insights for fraud detection. The classification engine used categorisation techniques like sentiment analysis and product classification to organise the data, enabling the system to identify patterns and anomalies. NLP helped evaluate reviews' tone and emotional context, highlighting potential fraud indicators.

Winnow collaborated with the RBI to develop a list of keywords for identifying fraudulent apps, available in English, Hindi, and Hindi written in English. The tool analysed linguistic patterns, sentiments, and app data, such as privacy statements and publication dates, to flag suspicious activities and high-risk features. Winnow's models were refined with training data to achieve higher accuracy, accounting for local languages, slang, and other language factors. The system improved with more data, enhancing fraud detection capabilities.

Sprint 6: Development of AI analytics

The fraud detection engine represented the final and most critical component of the overarching detection system, synthesising all data collected from an application and its associated content to predict the likelihood of fraud. This predictive process was driven by machine learning algorithms trained on historical data from fraudulent and non-fraudulent applications. These models identified patterns and detected anomalies within the data, such as misleading marketing practices, manipulative user reviews, or high-risk keywords.

The engine evaluated the provided data by applying a series of tests, each assessing the likelihood of fraud based on specific, predefined criteria. Examples included:

- Reports or indicators of fraudulent activity within the app's reviews
- Atypical high volume of positive reviews over a brief time span
- Ambiguous or unfavourable terms and conditions
- Incomplete or suspiciously limited information about the app or its publisher

Each test contributed a score reflecting the degree of fraud likelihood, enabling the engine to identify potentially fraudulent applications and provide a breakdown of the specific factors leading to this assessment. Fraud detection models were continually refined and updated to incorporate new data and evolving fraud patterns, ensuring the system remained robust and adaptable to emerging threats.

The fraud probability scores were positive but limited by the project's scope, resources, and available training data. The system was designed for long-term analysis, continuously re-scraping previously analysed app listings to identify emerging patterns. Final data processing efficiency depended on further investigation of analysed apps. Given the limited test data, validation apps scored at or near the top of analytics as expected.

Regarding latency, the system could scrape approximately 700,000 comments every 24 hours, and the analysis could process about 3,600 comments per hour. Techniques worth exploring for improved processing included larger Graphics Processing Unit machines or parallel processing across multiple servers to enhance model throughput.

Sprint 7: Development of custom queries

At RBI's request, Winnow designed a system to publish the result outputs in a customized report, allowing RBI's supervisors to analyse and query the data.

User testing design

Throughout the project, Winnow Technologies and RBI/RBIH engaged in an iterative process of developing user stories for the tool. The stories were categorised under "Data Acquisition." "Classifications." "Alerts," and "Reports."

Data Acquisition and Classifications refer to the underlying functionality of the tool and cannot be tested in the traditional sense by the end user. They either work or they do not.

However, user stories classified as Alerts and Reports can be tested by the end user, providing an opportunity for a brief, iterative process with the client. Once the tool is ready, Winnow will provide reports via email using an attached Excel spreadsheet.

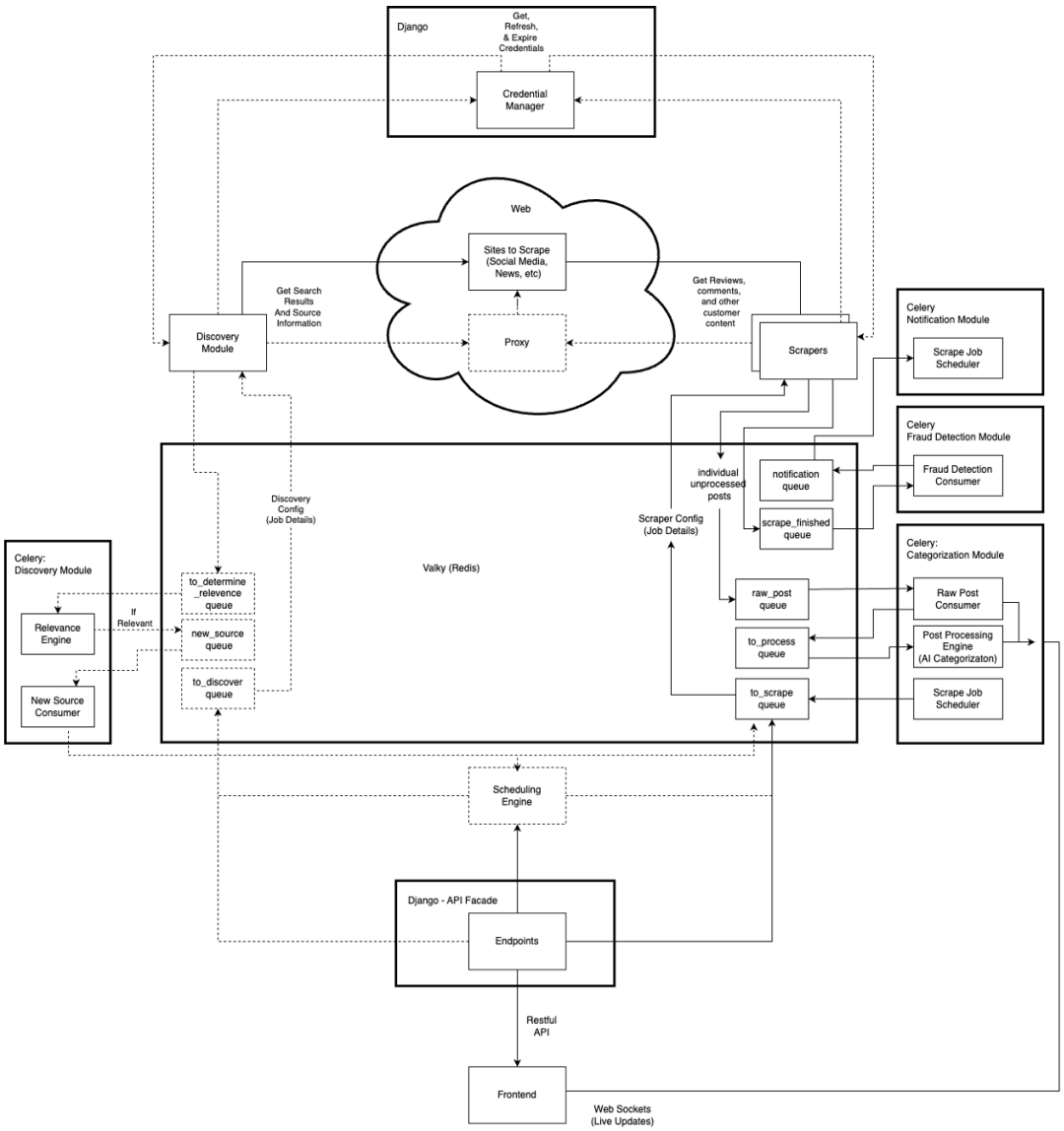
5. THE APPLICATION

The architecture of the solution, illustrated in Figure 2, is best viewed through the lens of the data stack framework, introduced in the Lab's State of SupTech Report 2023, which consists of technologies to facilitate data collection, processing and validation, storage, analysis, and presentation via data products..

Data schema

To tailor the collection engine to RBI's specific supervisory needs, the Lab first facilitated the development of a data schema. The schema is crucial not only for standardising the data to be collected but also to ensure the downstream storage and analytics are working with clean, tidy, and normalised data sets. The fields below are thus collected and stored for every app.

FIGURE 2. THE TECHNICAL ARCHITECTURE OF WINNOW TECHNOLOGIES' SOLUTION FOR RBI AND RBIH



Name of the app

The official name of the app as listed on the app store. The app's name can provide initial insights into its legitimacy, as misleading or deceptive names may indicate an attempt to confuse users or imitate more reputable apps.

App publication date

The date on which the app was originally published in the app store. Recently published apps with little user history may be more prone to fraudulent behaviour, especially if they gain popularity rapidly without a clear record of accomplishment.

Publisher information

- *Name:* The name of the app publisher or developer. Fraudulent apps may be associated with publishers with a history of releasing multiple low-quality or malicious apps.
- *Support number:* The contact phone number provided by the publisher. A lack of valid contact information may be a red flag for potential fraud.
- *Support email:* The publisher's support email address. Fraudulent publishers may provide non-functioning or generic email addresses to evade user inquiries.
- *Official publisher website:* Reviewing the website can offer insights into the company's legitimacy behind the app. A poorly designed or uninformative website may suggest fraud. Please note that Winnow's tool does not scrape or review the website but collects the website name for manual review.

Number of downloads

The total number of times the app has been downloaded. Apps with unusually high download numbers but poor reviews or low-quality content could indicate manipulated download metrics, a common tactic in fraudulent operations.

Categories assigned

The categories under which the app is listed in the app store. Fraudulent apps may be miscategorised or belong to high-risk categories such as finance, where users' sensitive information may be at risk.

Version number

The current version of the app. Frequently updated apps may respond to security concerns or improve functionality. However, apps with many version changes and no apparent improvement in functionality may be engaging in deceptive practices.

Number of app updates

The total number of times the app has been updated. A high frequency of updates without meaningful changes can indicate an attempt to avoid detection or improve user ratings artificially.

App description text

The publisher describes the app. The content and language of the description can be scrutinized for misleading claims, grammatical errors, or overly vague promises, which may indicate fraud.

Price

Whether the app is free or paid. Fraudulent apps may lure users with free downloads but later include hidden charges or in-app purchases that are not adequately disclosed.

Age appropriateness

The age rating assigned to the app. Fraudulent apps may incorrectly categorise themselves as appropriate for younger users to broaden their reach and appear less harmful.

Content warnings

Warnings regarding explicit content or other potentially harmful material. A lack of proper content warnings for apps with sensitive material could indicate deceptive practices or non-compliance with app store policies.

Data safety warnings & data usage details

Information on how the app collects and uses data. Apps that do not disclose data usage practices or warnings related to unsafe data handling may pose security risks and exploit user information, a common characteristic of fraudulent apps.

Average review score

The overall user rating for the app. Low ratings and reports of poor user experience or deceptive behaviour can suggest fraudulent intent. Conversely, apps with artificially inflated ratings (through bots or manipulated reviews) are also suspect.

Individual reviews (from scraper)

These user-generated reviews provide a wealth of information for detecting fraud. Specifically:

Reviewer

- *Username:* Repeated reviews from the same usernames across different apps can indicate fake or paid reviews.
- *Profile picture:* A profile picture's presence (or absence) may help identify potentially automated or bot accounts.
- *Review title:* The title of the review, when available. Short, vague titles may indicate low-effort or fake reviews.
- *Review body:* The full text of the review. A closer look at the content can reveal patterns of fake or repetitive reviews, such as generic praise or repeated complaints that suggest manipulation.
- *Review rating:* The rating assigned by the reviewer. A skewed ratings distribution (all 5 stars or all 1 star) may point to manipulated feedback.
- *Review publication date:* The date the review was posted. A surge of reviews in a brief time span could indicate an orchestrated effort to manipulate the app's rating.

Publisher Responses

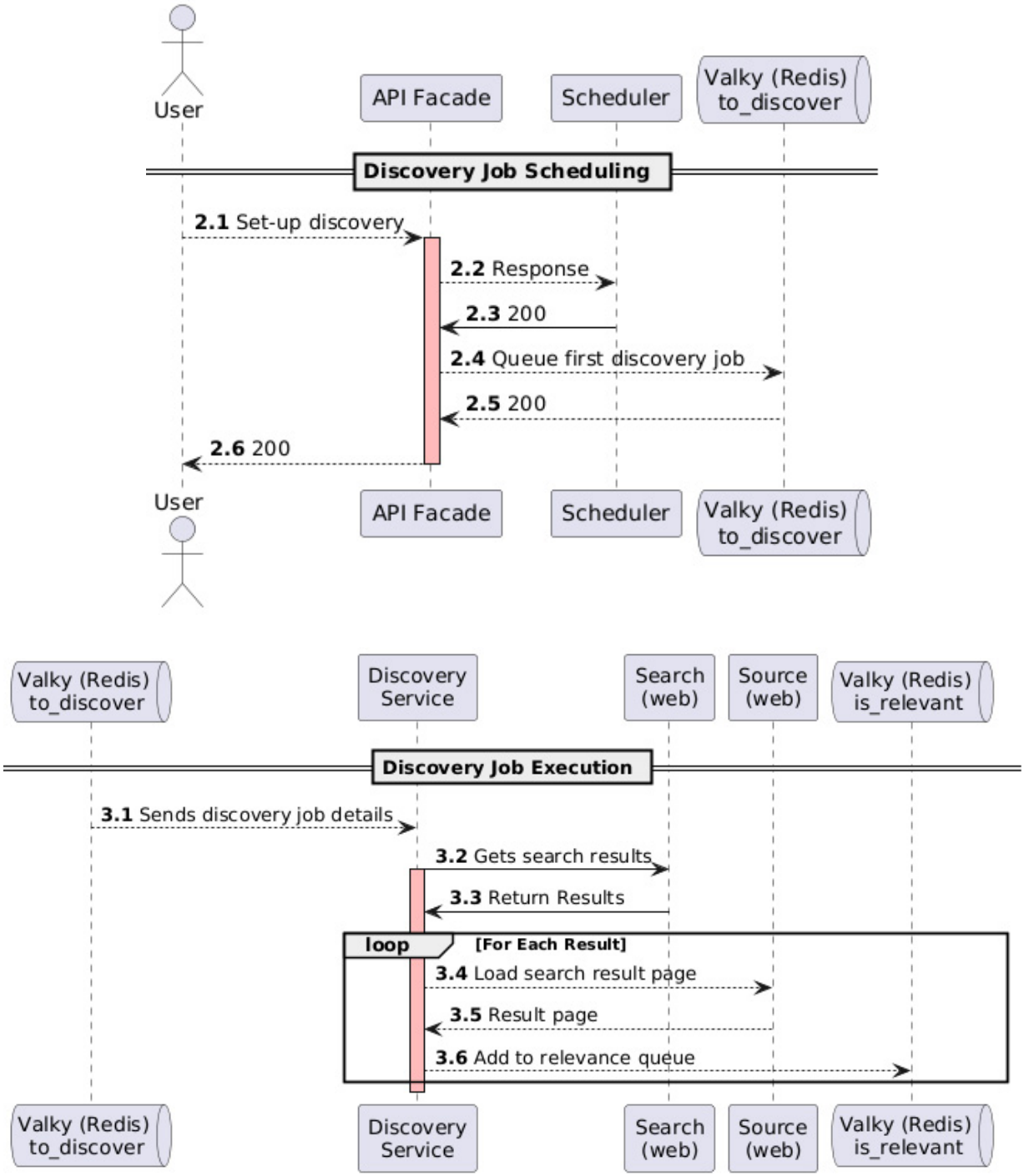
- *Response Body:* The publisher's response to user reviews. Fraudulent apps may ignore complaints or provide generic responses to hide bad practices.
- *Response Date:* The date the publisher responded. A lack of timely responses to user concerns can indicate poor customer service, a potential red flag for fraudulent behaviour.

Links

- *Privacy Policy:* The privacy policy of the app. Fraudulent apps may have vague or non-existent privacy policies, which could indicate improper handling of user data.
- *Terms of Service:* The app's terms of service. Like the privacy policy, the terms of service provide critical legal information that may highlight questionable or exploitative practices.
- *Publisher Website:* The official website of the publisher. A thorough review of the publisher's website can reveal signs of illegitimacy, such as lack of contact information, minimal content, or inconsistencies with the app's listing.

Each data point plays a role in assessing whether an app may be fraudulent. By analysing the metadata (app name, publisher details, download count), user-generated content (reviews, ratings), and additional context (privacy policy, app description), patterns of behaviour that align with fraudulent practices can be identified. For instance, an

FIGURE 3. WINNOW'S DISCOVERY ENGINE SEQUENCE DIAGRAM



app with many downloads but overwhelmingly negative reviews could indicate an attempt to deceive users about the app's popularity or effectiveness. Similarly, multiple low-effort reviews or publisher responses may suggest that the app is engaging in reputation management or review manipulation to mask its true nature.

Data Collection

The data collection process, scheduling and execution, illustrated in Figure 3, begins with identifying potential apps using a search results API. This API allowed Winnow to retrieve a comprehensive list of apps based on predefined filtering criteria, ensuring that only the most relevant and timely results were considered for further analysis. The primary filters applied included.**Source type:** Focused on apps from the Google Play Store, ensuring relevance for fraud detection and app behaviour analysis. Without Google Data, generic categorisation and manual refinement were used to tailor models for India's cultural and political landscapes.

Category

Narrowed the search scope to specific categories, particularly "Finance," to monitor apps involving financial transactions or activities prone to fraudulent behaviour.

Date range

Limited the search to apps updated or published within a specific period, usually within the past 24 hours, to ensure the most up-to-date data.

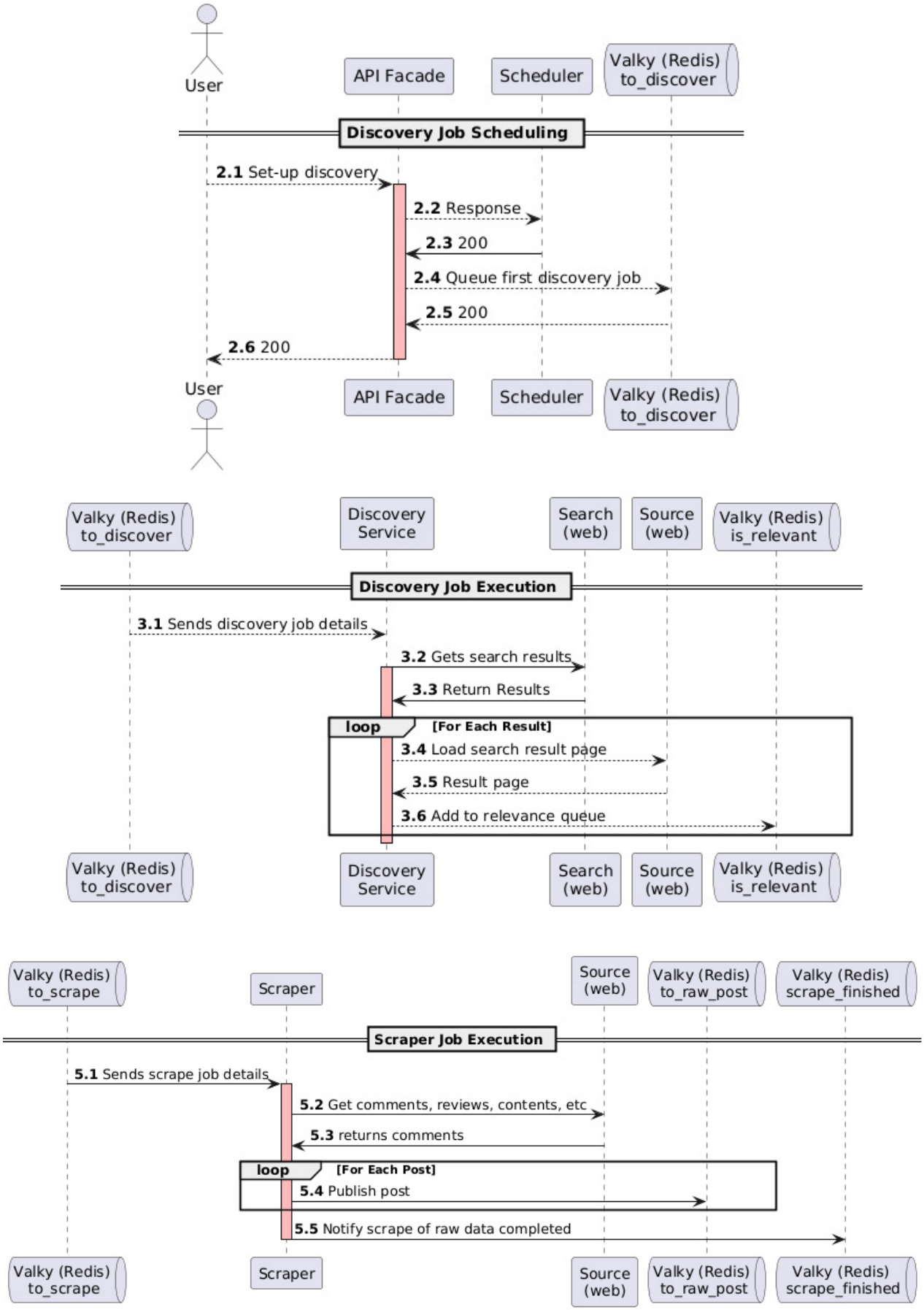
Once potential apps were identified, each result was enriched with additional supporting data from the Google Play Store, including app descriptions, developer information, user ratings, and download statistics. This enrichment provided crucial context for subsequent analysis stages.

After enrichment, each app was sent to the relevance engine for further evaluation. The relevance engine was pivotal in determining whether an app should undergo a full scrape and fraud detection analysis. Using predefined criteria and algorithms, it assessed whether the app met the thresholds for potentially fraudulent activity. Relevant apps were passed on to the next stage.

The scraper was then activated to collect additional data related to the relevant apps, focusing primarily on user-generated content such as customer reviews and publisher responses. This review data was crucial for detecting patterns of fraudulent behaviour, providing direct insights into user experiences and potential misconduct warning signs.

Reviews and responses were analysed and classified to determine the app's overall risk profile.

FIGURE 4. FINTECH APP DISCOVERY AND DATA COLLECTION WITHIN WINNOW'S SUPTECH SOLUTION



Data validation & processing

The validation and processing phase was critical to the system's data enrichment and fraud detection capabilities. Following the initial data collection stage, where apps were identified and relevant data scraped from the Google Play Store, the system proceeded to enrich this data through advanced processing techniques. The primary goal during this phase was to transform raw, unstructured data into actionable insights to inform subsequent fraud detection processes. This was achieved by applying categorisation techniques, such as sentiment analysis and product classification, to structure and organise the data for deeper analysis. The enriched data enabled the system to identify patterns, anomalies, and key indicators suggesting fraudulent behaviour, ensuring downstream decision-making processes were based on accurate and meaningful data.

A core aspect of data enrichment was **categorisation,** involving segmenting the scraped data into relevant categories, such as **sentiment, product, and issue.** Sentiment analysis, driven by NLP techniques, was essential in understanding user feedback by evaluating reviews' tone and emotional context. This process enabled the system to detect whether users expressed satisfaction, frustration, or other emotions, signalling potential issues with the app's functionality or ethical practices. Product categorisation also organised the data by associating each app or review with specific product types or services, allowing the system to differentiate between various financial services or products offered by the app. These categorisations provided a structured framework for analysing the data, enhancing the system's capacity to identify trends and make informed predictions about the app's potential for fraudulent activity.

The enriched data informed the fraud detection process and operated as the final stage of the processing pipeline. Utilizing the predictive models, the system applied ML algorithms to the categorised data to assess the likelihood of fraud. The system flagged apps exhibiting characteristics commonly associated with fraud by analysing patterns such as suspicious review behaviours, abnormal publisher activities, and high-risk app features (e.g., privacy violations and misleading descriptions). Fraud detection models were continually refined and updated to incorporate new data and evolving fraud patterns, ensuring the system remained robust and adaptable to emerging threats.

By validating and enriching the data before fraud detection, the system enhanced the accuracy of its predictions and reduced the occurrence of false positives. This increased the likelihood that apps with a high probability of fraudulent behaviour were flagged for human review and potential regulatory action. Thus, the validation and processing phase ensured the system operated precisely, effectively identifying fraudulent apps while minimizing unnecessary interventions. The integration of categorisation and ML-driven fraud detection exemplified the system's capacity to handle complex datasets and deliver reliable, actionable outcomes.

Data Storage

Winnow employed a PostgreSQL database as its primary storage solution to manage data collected and processed throughout the scraping, categorisation, and fraud detection pipeline. PostgreSQL, an advanced open-source relational database management system, was chosen for its robust features and support for complex datasets. This relational database stored large volumes of structured data while maintaining integrity, reliability, and accessibility.

At each stage of the data lifecycle—initial scraping, enrichment and categorisation, or final fraud detection analysis—all relevant information was systematically stored within the corresponding smartphone app's record

in the PostgreSQL database. This approach ensured a cohesive repository where data was properly indexed, enabling efficient querying and retrieval. Data integrity rules, enforced through schemas, constraints, and relationships, maintained accuracy and prevented anomalies.

Core advantages of leveraging PostgreSQL in this context include:

- **Structured data in a table-based format.**

  Each table corresponds to specific entities, allowing for clear relationships between data points. These relational links enable powerful queries, providing deeper insights into app behaviours and potential fraud indicators.

- **ACID compliance.**

  This technical specification ensures reliable database transactions, maintaining data accuracy and reliability. The system handles large data volumes through horizontal scaling, extensive indexing capabilities, optimising performance, and query efficiency.

- **Robust security mechanisms.**

  Role-based access control, encryption at rest, and encryption in transit ensure data confidentiality and integrity at the storage layer. Audit logging tracks access to sensitive data, ensuring traceability and accountability.

- **Support for JSON and NoSQL-like features.**

  These allow the storage of unstructured data, accommodating various data formats. This flexibility simplifies architecture and reduces complexity.

- **Performance tuning.**

  Tools for database optimisation allow for maintaining system performance even under high loads. By centralizing all data processing stages within a single, reliable database, Winnow efficiently and accurately detects fraud while maintaining scalability and flexibility.

Data Analysis

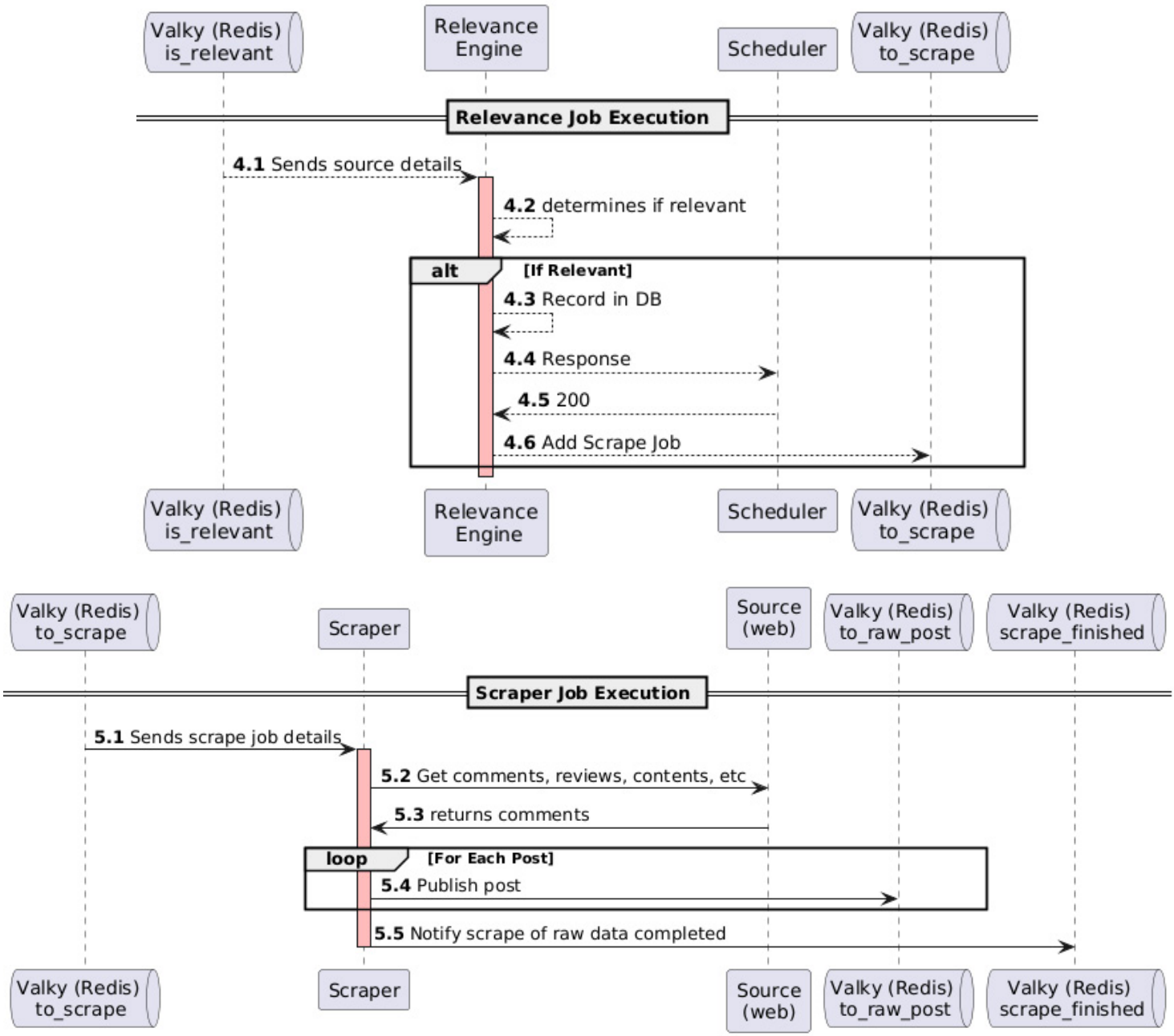
Winnow's approach to modelling consisted of a three-phase pipeline designed to process, filter, and analyse app-related data systematically. The first step involved app recognition, where the system leveraged various algorithms to identify and classify applications based on predefined parameters. This phase was critical for ensuring that only relevant applications were considered for further scrutiny. Once recognition was complete, the second step focused on isolating pertinent data from the broader set of available information. This involved extracting key data points such as metadata, user reviews, and app behaviours that could inform further analysis. ML models and statistical methods were employed to make an informed decision regarding the nature and intent of the app, particularly in the context of fraud detection or potential misuse. This structured pipeline allowed for efficient processing of large datasets while ensuring accuracy and relevance in the final assessment.

Winnow's data models worked with a pipeline, allowing customisation of each step to maximise efficiency and cost. These steps allowed each point to be replaced without redesigning the entire system, keeping up with the latest technologies and research at each touch point.

The pipeline consisted of the following three steps:

1. **Relevance engine:** Winnow detected if an app was relevant before moving forward by automatically analysing its title, description, and category as assigned by the Google Play Store. Existing categorisation models were fine-tuned using data from manual

FIGURE 5. WINNOW'S RELEVANCE ENGINE



searching and previous enforcement actions to narrow the focus to detecting loan and financial service apps. No conclusions on the presence of fraud were formed at this stage.

2. **Classification engine**

   Winnow scraped all available customer comments for relevant apps and used a second fine-tuned model to examine the overall sentiment of all comments and individual comments. This established a baseline of general feelings towards the app and plotted changes over time as policies shifted.

4. **Fraud detection engine**

   The final model categorises an app based on sentiment and app metadata such as title, date of publication, number of downloads, and similarity of comments (indicating possible synthetic comments). This model, refined using previous enforcement data, statistically determined the likelihood of fraud. The likelihood was determined by establishing a historical baseline and comparing recent data points to previously computed standards.

Winnow's approach to modelling consisted of a three-phase pipeline designed to systematically process, filter, and analyse app-related data. The first step involved app recognition, where the system leveraged various algorithms to identify and classify applications based on predefined parameters. This phase was critical for ensuring that only relevant applications were considered for further scrutiny. Once recognition was complete, the second step focused on isolating pertinent data from the broader set of available information. This involved extracting key data points such as metadata, user reviews, and app behaviours that could inform further analysis. ML models and statistical methods were employed to make an informed decision regarding the nature and intent of the app, particularly in the context of fraud detection or potential misuse. This structured pipeline allowed for efficient processing of large datasets while ensuring accuracy and relevance in the final assessment.

Analysis: Relevance engine

The relevance engine, the internal processes of which are illustrated in Figure 4, was crucial in determining whether a discovered application warranted deeper analysis or should be dismissed as irrelevant. Once the discoverability service identified a potential app, it was sent through the relevance engine, which evaluated the app against a set of predefined criteria. These criteria were designed to distinguish between irrelevant applications, such as personal finance tools, ATM locators, or budget education apps, and those that might exhibit predatory or fraudulent behaviour, such as unregulated loan applications or apps associated with non-financial institutions. By implementing this filtering mechanism, the relevance engine acted as an initial gatekeeper, ensuring that only apps with potential fraud indicators proceeded to the next stage of the pipeline. This step was essential for optimising the efficiency of the overall process by reducing the number of non-relevant apps subjected to further analysis. The filtering was based on an intricate algorithm that cross-referenced app metadata with known fraud patterns, keywords, and user feedback, allowing for a refined, evidence-based approach to app categorisation.

Data Analysis: Classification engine

The classification engine, the processes for which are illustrated in Figure 5, adds a deeper layer of analytical understanding by systematically categorising the content of scraped reviews, comments, and other user-generated data associated with the app. This engine leverages ML, particularly large language models (LLMs), to classify multi-dimensional data. Each post, review, or comment is analysed for its overall sentiment and for more granular segments such as issue type, product type, or user experience factors. The classification process involves feeding the scraped data into the LLM and a predefined set of potential categories and related keywords. The model then assigns each piece of data to one or more categories based on the patterns it detects.

This highly flexible classification process can be expanded to accommodate additional categories. For instance, should the need arise to classify data based on new product features or emerging issue types, the system can be dynamically updated to include these new segments. A forthcoming enhancement to this system includes the capability to detect multiple metrics for the same segment. This would involve iteratively prompting the model on a single segment while instructing it to blacklist previously identified metrics. Once the model exhausts all metrics for that segment, it proceeds to the next. This iterative approach enables a comprehensive, multi-layered analysis of the content, ensuring that all relevant aspects of the data are captured and classified for further use in the fraud detection process.

FIGURE 6. WINNOW TECHNOLOGIES' CLASSIFICATION ENGINE PROCESS

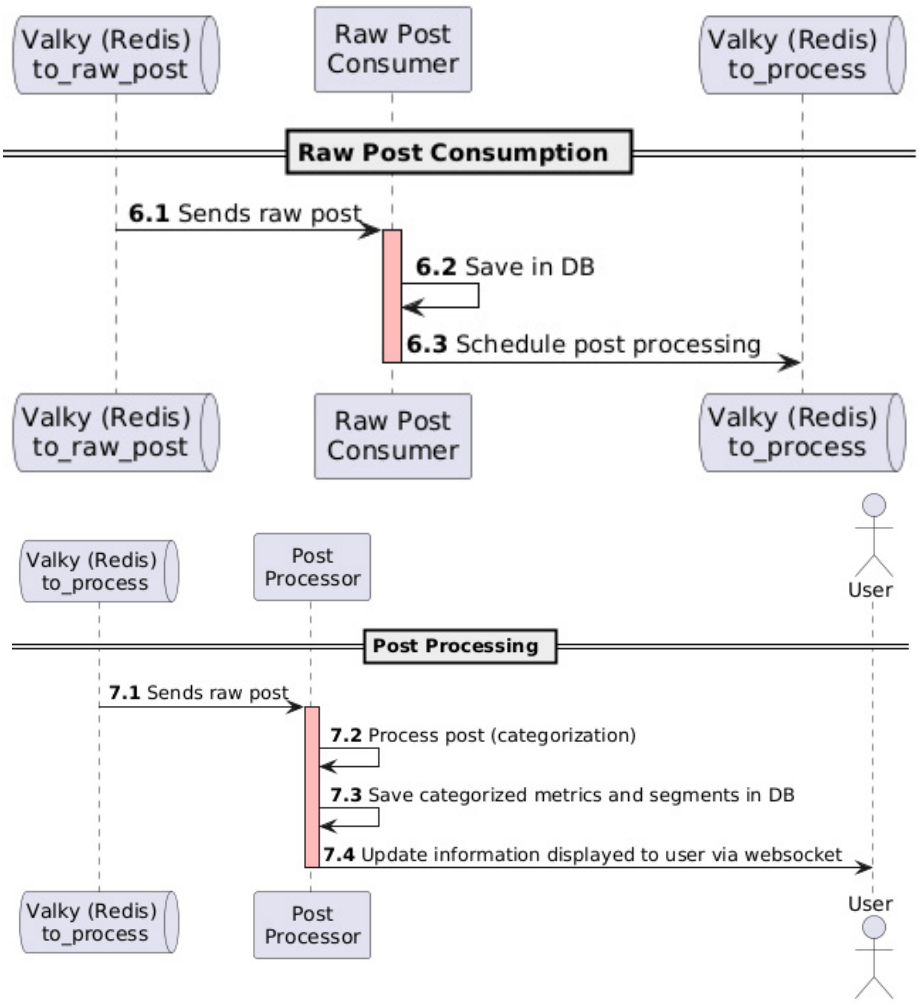
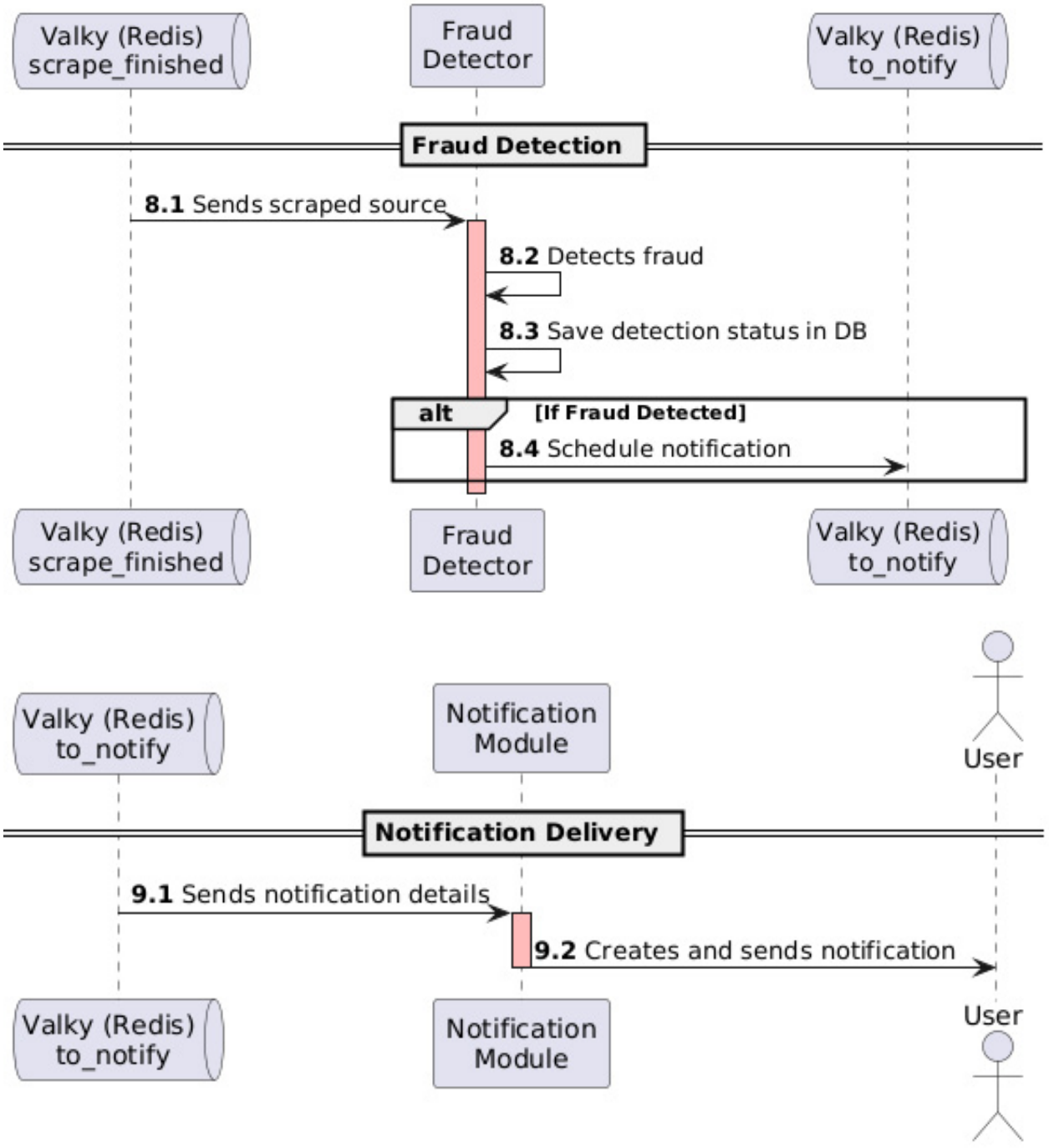


FIGURE 7. WINNOW'S FRAUD DETECTION ENGINE

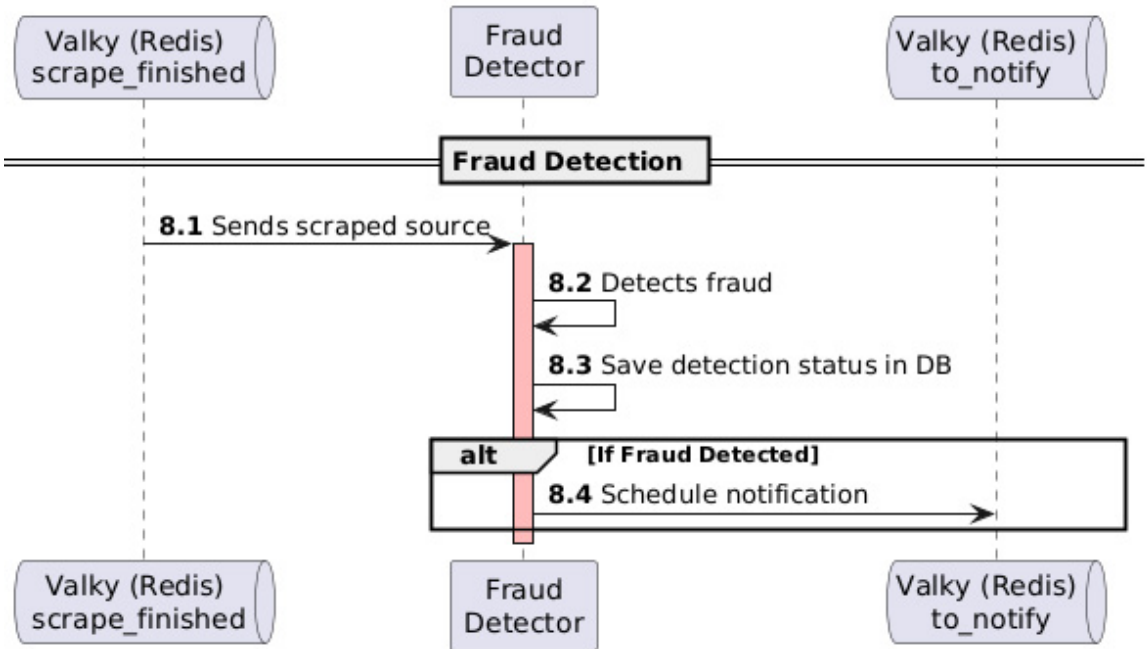


Data Analysis: Fraud detection Engine

The fraud detection engine, the processes for which are illustrated in Figure 6, serves as the final and most critical component of the overall system. It synthesises all the data gathered from an app and its associated content to predict the likelihood that the app is fraudulent. This engine integrates inputs from the relevance and classification engines to form a holistic picture of the app's behaviour and user interactions. By combining metadata, user sentiment, issue classifications, and behavioural patterns, the fraud detection engine applies advanced predictive models to assess the risk associated with the app. The prediction process is driven by ML algorithms trained on historical data related to fraudulent and non-fraudulent apps. These models identify patterns and anomalies in the data that may indicate fraudulent intent, such as deceptive marketing practices, manipulative user reviews, or the presence of high-risk keywords.

The fraud detection engine is designed to be highly adaptable and can update its predictive models as new data and patterns emerge. This ensures that the system remains effective even as the landscape of fraudulent apps evolves. Moreover, the engine can be scaled to handle large datasets, making it suitable for monitoring app stores with millions of apps and user reviews. The fraud detection engine provides a probabilistic assessment, allowing investigators or automated systems to prioritise which apps should be further scrutinized or flagged for removal from the platform.

FIGURE 6. WINNOW TECHNOLOGIES' FRAUD DETECTION ENGINE PROCESS



Winnow's data models worked in conjunction with a pipeline, allowing the customisation of each step to maximise efficiency and cost. These steps allowed each point to be replaced without redesigning the entire system, keeping up with the latest technologies and research at each touch point.

The pipeline consisted of the following three steps:

4. **Detection:** Winnow detected if an app was relevant before moving forward by automatically analysing its title, description, and category as the Google Play Store assigned. Existing categorisation models were fine-tuned using data from manual searching and previous enforcement actions to narrow the focus to detecting loan and financial service apps. No conclusions on the presence of fraud were formed at this stage.
5. **Sentiment analysis:** Winnow scraped all available customer comments for relevant apps and used a second fine-tuned model to examine the overall sentiment of all comments and individual comments. This established a baseline of general feelings towards the app and plotted changes over time as policies shifted.
6. **Categorisation:** The final model categorises an app based on sentiment and app metadata such as title, date of publication, number of downloads, and similarity of comments (indicating possible synthetic comments). This model, refined using previous enforcement data, statistically determined the likelihood of fraud. The likelihood was determined by establishing a historical baseline and comparing recent data points to previously computed standards.

Data products

The pipeline culminates with outputs for supervisors that comprise several data products, detailed below, plus a notifications engine illustrated in Figure 7.

Weekly reports: As each source (app) finishes going through the fraud detection engine, a list of all the apps will be generated with a fraud probability score (percentage, up to 100).

Upon completing the fraud detection process for each source (app), those identified as potentially fraudulent are subject to a fraud probability score. This mechanism ensures that relevant individuals or teams can take swift disciplinary action and prioritise apps accordingly. This means that if a report returns a list of apps, one of which has a fraud probability score of 80%, against the rest having a probability score in the range of 20%-45%, it allows the supervisors to prioritise supervisory action against the app with the higher fraud probability score. This is integral to the broader fraud detection framework, as it enables rapid dissemination of critical information to decision-makers, thereby allowing for immediate action, such as removing fraudulent apps from digital platforms or initiating further investigations.

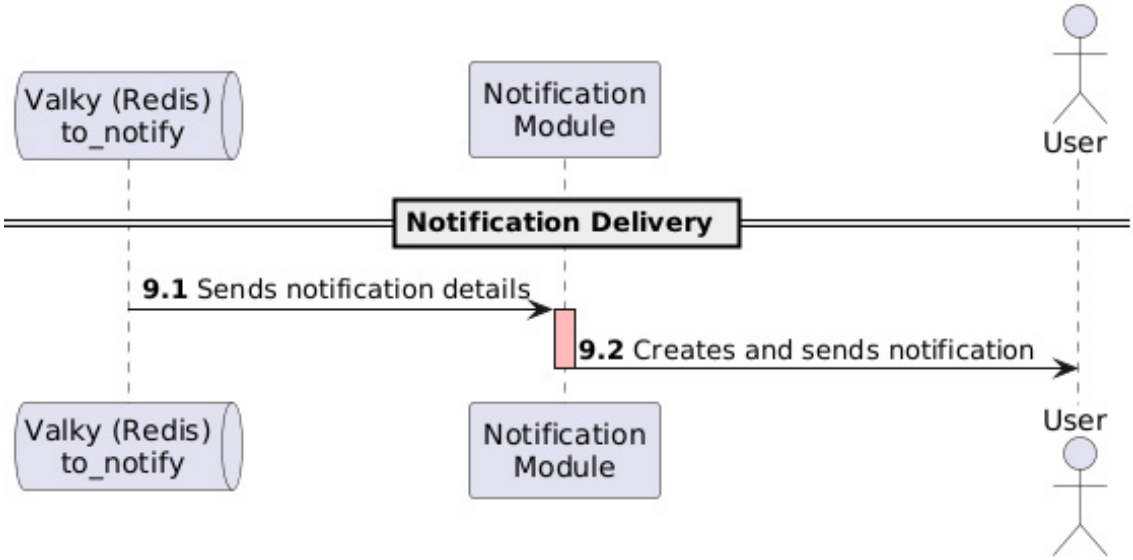
Periodic reports: At predefined regular intervals, such as weekly, comprehensive reports can be produced that present key analytics along with a detailed data dump of the apps that have been processed through scraping and fraud detection. These reports are designed to provide stakeholders with raw and synthesized insights into the system's performance, the behaviour of the apps under review, and the likelihood of fraudulent activity. The data dump accompanying each report will include all relevant details for each app, such as its fraud probability score, and the metrics and segments classified from the textual content of those reviews. This allows

stakeholders to assess the results of the fraud detection process, fostering transparency and accountability.

Detailed Analytics: As a core component of each periodic report or the interactive dashboard, critical statistics offer insights into the performance and outcomes of the fraud detection process. These key statistics are essential for assessing the scope and effectiveness of the system, as well as for tracking the prevalence and nature of potentially fraudulent activity across the app ecosystem. The following metrics will be highlighted:

1. **List of reviewed apps:** The report contains an enumerated list of apps that have been scraped and analysed, listed in the order of most likely to feature fraud indicators. This allows for an easily readable overview of potential apps to investigate further. This also includes columns such as the number of installations, the number of posts reviewed, when the app was last updated, and when the app was initially released.
2. **Comments indicated fraud:** This metric provides a percentage of comments analysed that indicate potential fraud given metrics such as sentiment, content and categories as processed by Winnow algorithms.
3. **Comment sentiment:** This number and metric indicates a percentage of positive sentiment from analysed comments, ranging from 0% positive to 100% positive.
4. **Fake comments (time clustered):** This is the percentage of potentially fake comments that cluster in each period. This shows how many fake comments were posted simultaneously, possibly indicating inauthentic bot behaviour. For example, if were 100 comments posted in a ten-minute period, but 40 were seemingly inauthentic according to our internal algorithms, the percentage would be 40%. This number will usually be exceptionally low or 0%, so if there is a spike, it can be an indicator of automated posting.
5. **Many comments from the same commenter:** This metric is a basic percentage of the comments posted by the same account in the observed period. Given that the prototype is scraping reviews, most commenters are expected to post only once or a few times. If this is not the case, that will be a positive indicator of potential bot activity.
6. **Author on the blacklist:** When a blacklist is supplied to the project this column lists the percentage of comments on or about or by entities on the blacklist, offering another major indicator of potential fraud. Please note that in the prototype, this list was not provided to the project, so currently, the number is always 0%. However, the algorithms to populate this are available, just awaiting data.

FIGURE 7. WINNOW TECHNOLOGIES' NOTIFICATION ENGINE PROCESS



User testing design

Throughout the project's life, Winnow Tech and RBI/RBIH engaged in an iterative process of developing user stories for the tool. The stories were categorised under "Data Acquisition." "Classifications." "Alerts," and "Reports."

Data Acquisition and Classifications refer to the underlying functionality of the tool and cannot be tested in the traditional sense by the end user. They either work or they do not.

However, user stories classified as Alerts and Reports can be tested by the end user, providing an opportunity for a brief, iterative process with the client. Once the tool is ready, Winnow will provide reports via email using an attached Excel spreadsheet.

6. IMPACT

The tool was able to estimate the probability of fraud among fintech apps using both metadata and complaints data. The tool used validation mechanisms to filter relevant apps and ran ML-based analytics using NLP.

The tool's relevance engine thoroughly passed each app scraped from the Google Play Store and probabilistically ensured they were financial lending apps in the Indian region.

The tools allowed for the automation of the following:

1. Discovering apps for monitoring
2. Scraping and analysing metadata and comments
3. Using predictive methods to give a fraudulent probability score

By giving supervisors a probability score rather than just a yes/no on potential fraudulence, the model allowed them to be proactive not just in addressing potential fraud but also by allowing them to prioritize and take any necessary action for those apps with higher fraud likelihood scores before those with lower scores.

This process removes a large chunk of manual checks and validation and is expected to save hundreds of person-hours and quickly trigger regulatory compliance actions.

Beyond fraud detection, the model's built-in probability indicators mean RBI can continuously measure trustworthiness. For example, a legitimate app may still have bugs or other features that make users leary, resulting in negative sentiment. This empowers RBI to not only request removal of bad actors from the app store, but to assess the pain points that customers have with legitimate actors.

It is expected that the full implementation and deployment of the tool within RBI/RBIH will, over time, deliver quantification of smartphone app listings, including the total number of negative reviews versus positive, the total number of fraudulent apps versus non-fraudulent, and other important metrics that are now feasible with the use of the new platform.

Ultimately, metrics and enforcement action results generated by this suptech solution can be a means for deeper risk-based supervision. Although end consumers of the lending apps are not the users of the suptech solution itself (and are therefore unaware of the quantitative, formal basis by which to trust an app), the risk-based supervision that the solution provides means that customers have more access to trustworthy tools, which will increase overall trust in the market, in turn providing a stronger foundation to further RBI's financial inclusion goals.

7. WHAT'S NEXT

The tool will allow RBI to increase trustworthiness and market value by empowering supervisors to act upon bad actors more quickly. Crucially, this has a trickle-down effect in the market as the end customers of fintech apps are expected to experience increased trust in the market due to reduced incidence and scale of fraud.

In terms of the underlying technologies, code and fine-tuned model have been shared with RBI for their use. RBI and RBIH intend to work on the prototype and develop it into a full-fledged model which monitors apps on a real-time basis and is exploring further engagement with both the Lab and Winnow Technologies. Currently, RBI and RBIH have a team of developers and analysts who can work on the model.

The first step will be to prepare more training data to fine-tune the existing model and further its capabilities to detect fraudulent apps. The model will continue to provide a more precise probability score, which later will be monitored by RBI supervisors who act as humans in the loop to (a) confirm the model's predictions as to whether the app is fraudulent, and (b) flag where the model has indicated an erroneous probability so the model can be iteratively trained to account for avoiding such anomalies. In this way, the feedback provided by the supervisors at RBI will be fed back to the model to enhance the accuracy of the model and learn from it, further improving its capability to detect fraudulent apps.

Additional features, including a live dashboard, management information system module, customer queries, role-based access and periodic reports, are facets that the RBI and RBIH teams expressed their collective desire to work on.

Reserve Bank of India (RBI)

The apex bank in India, licences, controls, and regulates the financial and banking system in India, including banks, non-banking financial companies (NBFCs), and cooperative banks.

Reserve Bank of India Hub (RBIH)

The Innovation Hub is a wholly owned subsidiary of the RBI that was set up to promote and facilitate an environment that accelerates innovation across the financial sector. The RBIH aims to foster and evangelise innovation across the financial sector to enable a billion Indians to access suitable, sustainable financial products securely and frictionlessly. In addition, the RBIH intends to create internal capabilities by building applied research and expertise in the latest technology while collaborating with financial sector institutions, policy bodies, the technology industry, and academic institutions and coordinating efforts to exchange ideas and develop prototypes related to financial innovations.

Winnow Technologies Inc.

Winnow Technologies (Winnow) is a vendor that specialises in web-based data mining tooling, natural language processing and advanced analytics to assist public agencies in fulfilling their mandates to citizens and support the development of inclusive, sustainable and resilient markets, economies, and societies. The tools developed and deployed by Winnow allow the oversight of regulated firms and unregulated activities by scanning the web, social media, company reports and other communications to flag potential violation of policy and regulations, conduct sentiment analysis, and correlate collected information for supervisors on an ongoing basis.

About the Cambridge SupTech Lab

The Cambridge SupTech Lab accelerates the digital transformation of financial supervision to nurture resilient, transparent, accountable, sustainable, and inclusive financial sectors.

The Lab catalyses the integration of innovative technologies and data science into supervisory processes to address enduring and emerging challenges in the rapidly evolving financial landscape. Through the Lab, financial authorities have championed the adoption of advanced suptech solutions to address pressing issues such as financial crime, fraud, exclusion, climate change enablers, consumer protection, artificial intelligence biases, and the supervision of fintech and digital assets.

Our global, multidisciplinary team partners with financial authorities' executives, supervisors, and data scientists to craft solutions across the entire innovation lifecycle – from data governance to AI-powered strategies, from the initial design to the full-scale deployment and scaling of cutting-edge suptech applications.

The Cambridge SupTech Lab is an initiative of the Cambridge Centre for Alternative Finance (CCAF) at the Cambridge Judge Business School, leveraging foundational intellectual property and know-how from the RegTech for Regulators Accelerator (R²A).

The mention of specific companies, manufacturers, or software does not imply that they are endorsed or recommended by the Cambridge SupTech Lab in preference to others of a similar nature that are not mentioned.

SUGGESTED CITATION

Cambridge SupTech Lab (2025), Case Study: Intelligence Platform for Flagging Fraudulent Fintech Apps, Cambridge: Cambridge Centre for Alternative Finance (CCAF), University of Cambridge. Available at www.cambridgesuptechlab.org

AUTHOR

Nathalie Lenehan, Manish Muralidharan

DESIGN

Dayna Donovan

ADDITIONAL CONTRIBUTORS

David Allen, Samir Kiuhan-Vasquez, Kalliopi Letsiou, Susu Smaili