CASE STUDY

N.1 | AUGUST 2024

FINANCIAL CONSUMER PROTECTION SUITE WITH WEB SCRAPING AND MACHINE LEARNING-BASED ANALYSIS

SUPTECH WORKING PROTOTYPES DEVELOPED BY THE CAMBRIDGE SUPTECH LAB AND ITS PROJECT PARTNERS, THE BANGKO, SENTRAL NG PILIPINAS, THE SECURITIES AND EXCHANGE COMMISSION OF THE PHILIPPINES, AND WINNOW TECHNOLOGIES

This case study outlines the development of a functional prototype featuring a financial consumer protection suite with social media and web scraping, sentiment analysis, topic modelling, and other machine learning-based analytics. The solution was developed by the Cambridge SupTech Lab in collaboration with the Bangko Sentral ng Pilipinas (BSP), the Securities and Exchange Commission of the Philippines (SEC), and suptech vendor Winnow Technologies.

Cambridge Centre for Alternative Finance
UNIVERSITY OF CAMBRIDGE Judge Business School

CAMBRIDGE SUPTECH LAB

www.cambridgesuptechlab.org

Project overview

Sentiment analysis is a field of study that analyses people's opinions, sentiments, evaluations, attitudes, and emotions from written language, including in social media, using natural language processing (NLP) and text analysis. It classifies the text into positive, negative or neutral categories, and also to more sophisticated categorizations like emotions or intentions. Topic modelling uses machine learning to automatically identify themes or topics present in a collection of texts, making it easier to understand and organize large amounts of information without needing to read every single document.

By collecting, processing, and analysing topics and sentiments, financial supervisors can gauge the market's mood regarding specific companies, financial products, or the economy in general. This can help in early detection of potential market manipulation or fraud. Risks can be proactively surfaced by identifying unusually positive or negative sentiments that do not align with other market data or identifying anomalous spikes in volumes of posts relating to a particular topic.

For example, negative sentiments can sometimes precede market downturns or can indicate potential volatility. Some financial authorities use sentiment analysis to complement traditional economic indicators, providing a more real-time and granular view of economic health.

Using artificial intelligence (AI) to provide a nuanced understanding of consumer sentiment, The prototype delivered interactive dashboards for visualizing data and reporting insights from web-based sources. A centralized data warehouse was established to consolidate various data streams, including consumers' complaints filed with the agencies, enabling comprehensive correlation studies and analyses.

The dynamic business intelligence dashboard offered supervisors a straightforward platform to examine and interpret real-time data through graphs depicting values and percentages, as well as tables for detailed data scrutiny.The prototype included a feature for seamlessly transmitting

data at any point in the cleansing or analysis phase via an API, enhancing existing complaints systems and enriching supervisory intelligence.

These cutting-edge tools enabled the BSP and the SEC to offer more robust assistance and recourse to financial consumers and derive meaningful analytical insights, thus fine-tuning their supervisory methodologies and making more efficient and effective use of limited resources.

The project underscores the benefits of collaborative development between the financial authorities, the Lab and its technology partners, showcasing how joint efforts can lead to the swift creation of advanced suptech solutions while also achieving cost savings.

In reference to the Suptech Generations framework, the solution progressed technology deployment and data science methodologies from 0G to 4G in every layer.

Project partners

- **Bangko Sentral ng Pilipinas (BSP)**: Ensures financial system stability and consumer protection.
- **Securities and Exchange Commission of the Philippines (SEC)**: Supervises the corporate sector and capital markets.
- **Winnow Technologies**: Specializes in web-based data mining, natural language processing, and advanced analytics.

Challenges

- **Manual data processing limitations**: Substitute and augment manual, obsolete supervisory methods with deeper, real-time analysis of public sentiment and risk assessment.
- **Limited data sources**: Diversify data sources for a deeper understanding of financial institutions' market conduct and consumers' behaviours, preferences and sentiment.
- **Reactive supervision**: Shift from reactive to more proactive supervision via an early warning system that identifies unusual patterns of transactions.

Key features

- **Centralized data warehouse**: Consolidates various data streams for comprehensive analysis.
- **API integration**: Allows seamless data transmission and integration with existing systems.
- **Social media and web scraping**: Automates the data collection from web platforms and stores it in a data warehouse for correlation with internal datasets.
- **Sentiment analysis**: Interprets opinions, sentiments, and emotions from text data, classifying them into positive, negative, or neutral categories.
- **Topic modeling**: Identifies themes in text data, aiding in the organization and understanding of large information sets.
- **Interactive dashboards**: Visualizes data and insights in user-friendly formats.

Benefits

- **Real-time analysis**: Provides supervisors with a nuanced understanding of market sentiment and trends, aiding in proactive risk management and detection of potential market manipulation and fraud.
- **Automated data aggregation and processing**: Reduces the manual workload by automatically collecting and analysing data from public domainssources like news articles, blogs, financial forums and social media.
- **Enhanced supervision**: Advanced analytics provides a more robust capability to detect emerging patterns of market misconduct and the related risksIntegrates with existing complaint management systems to offer a comprehensive view of consumer sentiment and market conduct for a more preventive and proactive approach.

1. BACKGROUND AND SUPERVISORY CHALLENGES

Over the past five years, BSP has significantly enhanced its consumer protection and supervision capabilities, accessing timely, accurate, and actionable insights that have driven more effective market conduct oversight and have also informed policy and regulatory development. In 2020, BSP distinguished itself as the first central bank to deploy a chatbot-assisted integrated complaints management system, innovatively designed to streamline the collection and administration of financial consumer grievances – BSP received 23,142 complaints in 2022 and 43,115 in 2023.

This system was developed in collaboration with the RegTech for Regulators Accelerator (R²A), which has since evolved into the Cambridge SupTech Lab. Additionally, BSP systematically collects data from regulated financial institutions detailing consumer complaints and their resolutions. BSP has been looking to augment these processes with data from web-based public domains, such as social media, news outlets, and app-store reviews, with the intention to leverage this rich array of alternative data to acquire a deeper understanding of financial institutions' market conduct and consumers' behaviours, preferences and sentiments.

Similarly, the SEC supervisors and investigators have been exploring the use of diverse data solutions to enhance their analytical capabilities, aiming to better understand financial consumer behaviour and to bolster market conduct oversight. For instance, they have used "open sources" like search engines and social media to manually cross-reference "closed sources" such as the SEC's company registry and internal databases that contain periodic reports from registered companies. However, these open and closed sources often generate a significant amount of data that is not always efficiently processable by manual means, leading to an excess of unfiltered, irrelevant information – referred to as "noise". In some instances, the SEC's oversight tools and methodologies have allowed misconduct to slip through their system undetected, with investigations triggered by complaints received or misconducts reported only once substantial harm has been incurred by users.

The BSP and SEC acknowledged the necessity to substitute and augment manual, obsolete supervisory methods with deeper, real-time analysis of public sentiment and risk assessment of the financial entities under their watch. By harnessing the advanced analytics prototyped through this project, the BSP and SEC anticipate a more robust capability to detect emerging patterns of market misconduct and the related risks. Such insights would enable an early warning system that identifies unusual patterns of transactions that may indicate fraudulent activities or misconduct, as well as signs of a financial firm's distress, and even signals of broader financial instability (e.g., rapid credit growth, assets bubbles, widespread over indebtedness).

This could help prevent the escalation of these issues into full-blown financial crises, reducing systemic risks and supporting macroprudential policy by providing data-driven insights into the overall health of the financial system and identifying potential threats to financial stability. The shift towards suptech enables these supervisory agencies to evolve from a reactive posture to one that is decisively proactive and preventive, elevating the practices of financial supervision to meet the exigencies of the digital era.

2. PROJECT CONCEPTUALISATION AND INCEPTION

In 2022, teams from BSP and SEC participated in the Cambridge SupTech Lab's Capacity Building and Education (CB&E) online programmes. There, they forged separate but complementary proofs of concept (POCs) of suptech applications that would enhance their supervisory capabilities in consumer protection.

Both teams sought an automated solution that would streamline the web scraping of public web content, employing AI to dynamically adjust search terms through sentiment analysis, and pattern correlations with internal complaints data. This would enable early detection of business practices, scams, and fraudulent activities that would potentially harm financial consumers.

The concept was to equip the supervisors in their agencies with immediate access to comprehensive synthesised information needed for prompt and actionable insights, enhancing the understanding of financial consumer behavior and fortifying market conduct oversight. Such a system would produce sharper insights while reducing the workload and making resource allocation more efficient.

Addressing shared challenges, BSP and SEC set out to harness suptech to:

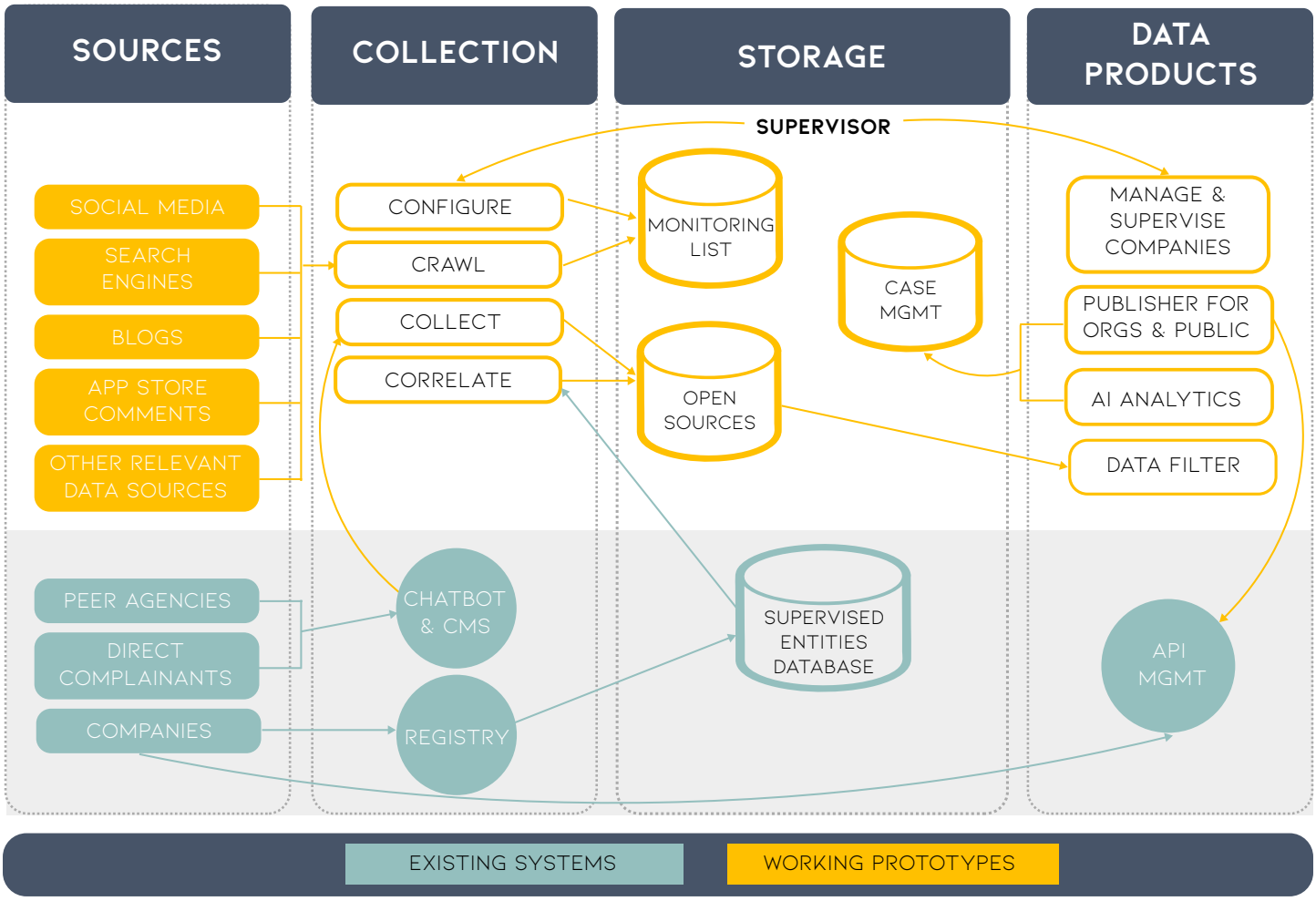
- Integrate data from public scraping, chatbot interactions, and regulatory reports from financial firms and other entities under their supervision.
- Leverage this data to identify new market patterns, trends, and emerging risks.
- Investigate the range of financial products and services in the market and the channels through which they are transacted.
- Identify the spectrum of entities and business models associated with new trends, pinpointing their roles and the degree of their involvement.
- Evaluate the impacts of specific regulations on the financial industry, correlating them with prevailing trends and issues.
- Analyse the interconnections between market trends and diverse data sets – social media, news sites, Appstore reviews, and consumer complaints – for a holistic perspective.
- Apply the derived insights from integrated data analytics to synchronise regulatory actions better and judiciously distribute resources for effective interventions.

> *"We envision integrating the publicly scraped data with our chatbot data and the data from regulatory reports from the entities supervised by BSP. We would want to have a view of emerging trends or financial issues, types of financial products and services, transaction channels, specific regulations, implicated the involved entity types and the correlation on each of the emerging trend within these different data sources. This can help us coordinate our efforts and actually allocate the resources effectively to facilitate prompt and tailored intervention."*
>
> *Alvin Tolosa*
> *Consumer Protection and Market Conduct office, BSP*

Figure 1 provides a diagram of the envisioned solution. The pre-existing complaints management platform components are shown in teal, while the novel web-based data collection, data warehouse, and advanced analytics components of the working prototype are shown in yellow.

With its mission to accelerate the roll-out of scalable, cutting-edge suptech solutions – streamlining everything from de-risked vendor selection and procurement to agile prototyping and testing – the Lab, in January 2023, selected these two POCs for prototype development from the dozens of projects spawned by CB&E programme attendees. These POCs stood out as they made a stronger case for web scraping among many agencies interested in this type of tool. In addition, the teams both had strong buy-in from the executive management for such an initiative. The Lab partnered with the BSP and SEC to align their POCs and refine the technical specifications.

FIGURE 1. SCHEMATIC DIAGRAM OF THE FINANCIAL CONSUMER PROTECTION SUITE WITH WEB SCRAPER AND ML-BASED ANALYSIS



3. LEAN VENDOR PROCUREMENT AND SELECTION

In March 2023, the Lab procured the working prototypes on behalf of the agencies, executing global competitive bidding via a Request for Proposal (RFP) and spearheading an expedited yet rigorous vendor selection process. An independent expert panel reviewed the anonymised bids from a global cohort of applicants, selecting the proposal by Winnow.

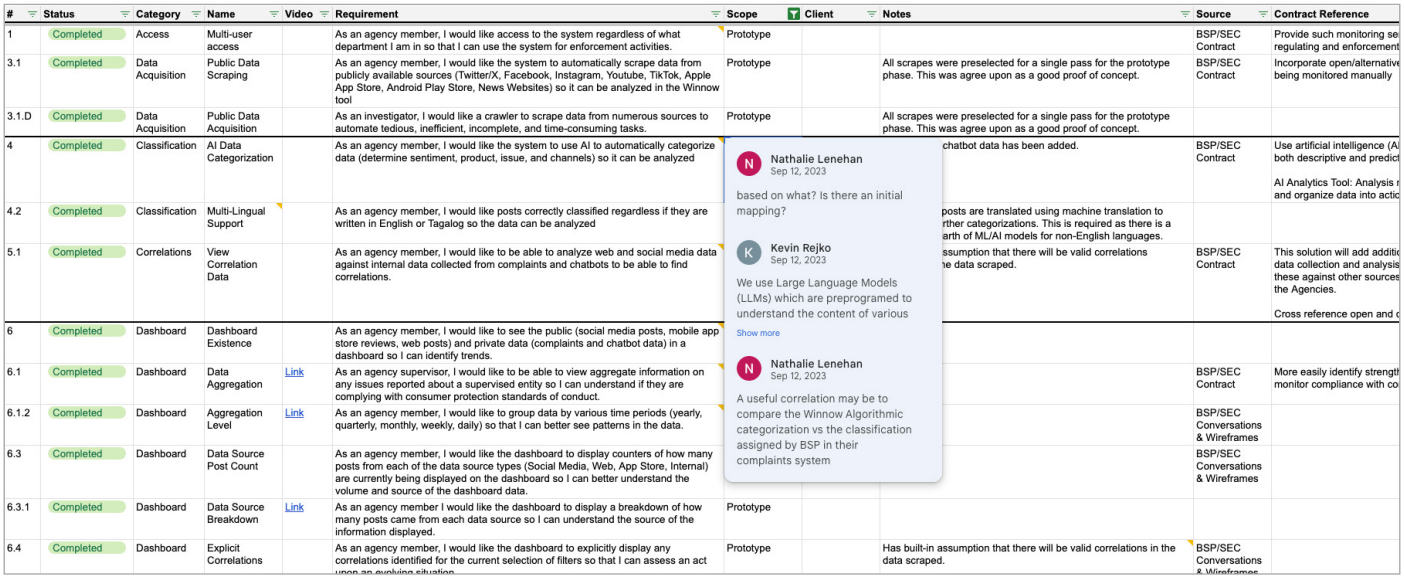
First, experts from the Lab and the University shortlisted responses to a request for expressions of interest (REOI) based on three criteria: (1) relevant experience (60%), technical and managerial expertise (30%), and adequate resourcing (10%). Next, each firm on the shortlist was issued an RFP, and ensuing proposals were reviewed by an independent panel of judges comprised of leading subject-matter experts and innovators from around the world, including, at minimum an expert in financial regulation and supervision, a technologist with experience deploying the tools being assessed, and a specialist with expertise in the particular product at hand. The criteria panelists considered during this second stage were topic responsiveness (65%), execution plan (25%), and innovative approach (10%).

Once the vendor was selected by the panel and run through a no-objection process with the partner financial authorities, the Lab and the University of Cambridge conducted necessary due diligence and formalised the settlement of legal issues with regard to data sharing and storage, licensing (of intellectual property), and public procurement. All items were addressed by a project agreement between the University and the vendor, inclusive of non-disclosure agreement (NDA) terms.

After completing due diligence, on behalf of the Lab, the University of Cambridge contracted the selected vendor (Winnow Technologies Inc.) at the end of April. Once the vendor was contracted, the Lab provided project management through the development and testing phases of the working prototype. The project kicked off in May 2023.

Starting in June, the Lab led design sprints with cross-functional project teams from each agency and the vendor to gather user stories and further detail the specific functionalities envisioned (Figure 2). In this phase, the project teams worked asynchronously, contributing comments, questions and suggestions through a truly collaborative, interactive and iterative process facilitated by the Lab.

FIGURE 2. USER STORIES & ASYNCHRONOUS COLLABORATION



#	Status	Category	Name	Video	Requirement	Scope	Client	Notes	Source	Contract Reference
1	Completed	Access	Multi-user access		As an agency member, I would like access to the system regardless of what department I am in so that I can use the system for enforcement activities.	Prototype			BSP/SEC Contract	Provide such monitoring se regulating and enforcement
3.1	Completed	Data Acquisition	Public Data Scraping		As an agency member, I would like the system to automatically scrape data from publicly available sources (Twitter/X, Facebook, Instagram, Youtube, TikTok, Apple App Store, Android Play Store, News Websites) so it can be analyzed in the Winnow tool	Prototype		All scrapes were preselected for a single pass for the prototype phase. This was agree upon as a good proof of concept.	BSP/SEC Contract	Incorporate open/alternative being monitored manually
3.1.D	Completed	Data Acquisition	Public Data Acquisition		As an investigator, I would like a crawler to scrape data from numerous sources to automate tedious, inefficient, incomplete, and time-consuming tasks.	Prototype		All scrapes were preselected for a single pass for the prototype phase. This was agree upon as a good proof of concept.		
4	Completed	Classification	AI Data Categorization		As an agency member, I would like the system to use AI to automatically categorize data (determine sentiment, product, issue, and channels) so it can be analyzed			chatbot data has been added.	BSP/SEC Contract	Use artificial intelligence (A both descriptive and predict AI Analytics Tool: Analysis and organize data into actic
4.2	Completed	Classification	Multi-Lingual Support		As an agency member, I would like posts correctly classified regardless if they are written in English or Tagalog so the data can be analyzed			posts are translated using machine translation to rther categorizations. This is required as there is a arth of ML/AI models for non-English languages.		
5.1	Completed	Correlations	View Correlation Data		As an agency member, I would like to be able to analyze web and social media data against internal data collected from complaints and chatbots to be able to find correlations.			ssumption that there will be valid correlations e data scraped.	BSP/SEC Contract	This solution will add additi data collection and analysis these against other sources the Agencies. Cross reference open and
6	Completed	Dashboard	Dashboard Existence		As an agency member, I would like to see the public (social media posts, mobile app store reviews, web posts) and private data (complaints and chatbot data) in a dashboard so I can identify trends.					
6.1	Completed	Dashboard	Data Aggregation	Link	As an agency supervisor, I would like to be able to view aggregate information on any issues reported about a supervised entity so I can understand if they are complying with consumer protection standards of conduct.				BSP/SEC Contract	More easily identify strength monitor compliance with co
6.1.2	Completed	Dashboard	Aggregation Level	Link	As an agency member, I would like to group data by various time periods (yearly, quarterly, monthly, weekly, daily) so that I can better see patterns in the data.				BSP/SEC Conversations & Wireframes	
6.3	Completed	Dashboard	Data Source Post Count		As an agency member, I would like the dashboard to display counters of how many posts from each of the data source types (Social Media, Web, App Store, Internal) are currently being displayed on the dashboard so I can better understand the volume and source of the dashboard data.				BSP/SEC Conversations & Wireframes	
6.3.1	Completed	Dashboard	Data Source Breakdown	Link	As an agency member I would like the dashboard to display a breakdown of how many posts came from each data source so I can understand the source of the information displayed.	Prototype				
6.4	Completed	Dashboard	Explicit Correlations		As an agency member, I would like the dashboard to explicitly display any correlations identified for the current selection of filters so that I can assess an act upon an evolving situation	Prototype		Has built-in assumption that there will be valid correlations in the data scraped.	BSP/SEC Conversations & Wireframes	

Nathalie Lenehan
Sep 12, 2023
based on what? Is there an initial mapping?

Kevin Rejko
Sep 12, 2023
We use Large Language Models (LLMs) which are preprogramed to understand the content of various
Show more

Nathalie Lenehan
Sep 12, 2023
A useful correlation may be to compare the Winnow Algorithmic categorization vs the classification assigned by BSP in their complaints system

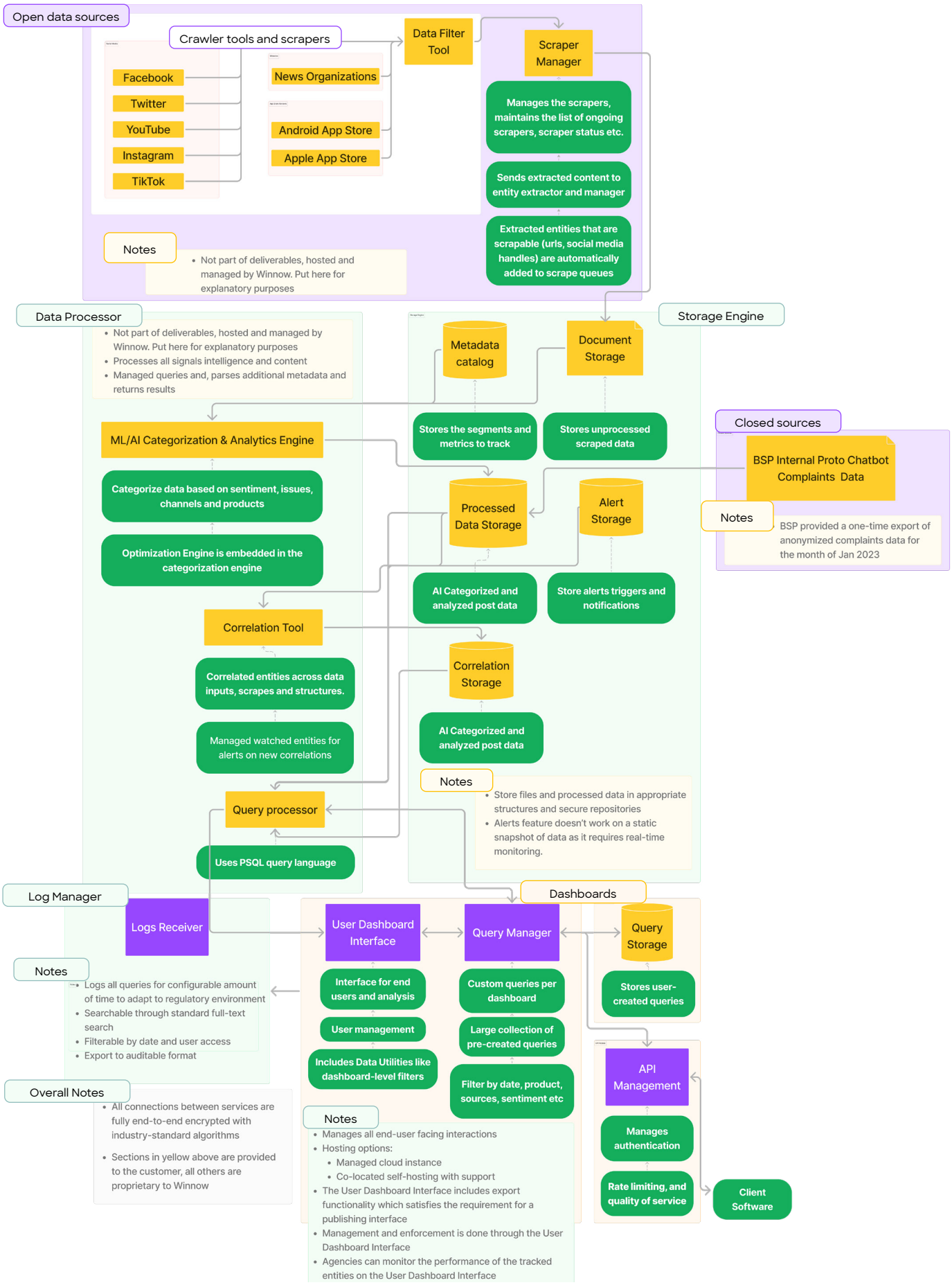
4. WORKING PROTOTYPE AGILE DEVELOPMENT

Winnow developed and configured the platform in accordance with the system architecture shown in Figure 3, and quickly moved from mock-ups to development of a working prototype with agile development and hands-on testing that delivered incremental enhancements of the solution each week when the project team (Winnow, BSP, SEC and the Lab) met to review progress, discuss new features, interact with the technology hands-on, and exploring ideas to further the impact of the project.

The Lab's approach fosters a transparent environment where progress and details of the development are visible to all stakeholders. This ensures that the system or solution being developed does not become a "black box" – a system or application whose inner workings are hidden or poorly understood by its users. By integrating user feedback throughout the development phases, agile methods help evolve the solution to align closely with user needs and expectations, making the entire process open and understandable.

Six months after the kick-off meeting, in November 2023, user acceptance testing was concluded, and the working prototypes were delivered. This included production roadmaps detailing the options and requirements for a lean solution deployment informed by the prototyping process.

FIGURE 3. WINNOW'S SYSTEMS ARCHITECTURE



5. THE APPLICATION

Winnow delivered a platform that automated the ingestion, processing, storage, analysis, and visualisation of data streams from social media channels, app-store reviews, and news sites, along with APIs for integration with existing complaints management systems. Winnow's advanced technology harmonised public datasets with internal agency data, applying sophisticated sentiment analysis, topic modelling, and other machine learning techniques.

This data was systematically organised in a centralised database system, ensuring secure and efficient storage. Access to the data was facilitated through:

1. An API for seamless integration with existing on-premises systems and processes for the SEC

For the SEC Philippines, the API approach has proved to be particularly beneficial as it allows the integration of the processed public data into their existing systems. This enables the agency to continue leveraging its own proprietary algorithms for conducting correlations and other AI-driven analytics. This integration significantly enhanced the SEC analytical capabilities by incorporating external insights provided by the vendor.

In the prototype's testing phase, SEC Philippines opted to monitor two entities previously subjected to enforcement actions. Many of these entities' accounts had been overtaken by spam or were deleted from social media platforms, along with their post histories, thus the resulting scraped dataset volume was low.

An API was developed and tested to transmit specified data elements including date-time stamps, sentiment analysis, details of the posting entities, and the content of the posts. The system's automation allowed for this information to be processed and integrated into cases for consideration by different regulatory departments of SEC Philippines to augment and focus on multi-monitoring to facilitate filtering out the noise and gather actionable data for regulation and oversight enforcement.

2. A dashboard with intuitive visualisations for the BSP

The BSP opted to develop an interactive dashboard that could be integrated with its existing complaint management platform. The interface features user-friendly access, navigation, and analysis, with intuitive and insightful visualisations. Custom filtering options allow users to sift through the data seamlessly. For instance, users can opt to view only entries with low sentiment scores, sort data by date, or focus on specific products, among other options.

The dashboard offers the functionality of side-by-side comparison between different filtered datasets. This feature is extremely valuable for supervisors, allowing them to benchmark one supervised entity against others within the same market segment or to compare one market segment with another. Such comparisons are particularly useful for detecting anomalies. For instance, supervisors can now identify if a bank is an outlier compared to its competitors, or if a larger segment of the market is disproportionately affected by a significant event, signalling the need for additional investigation and oversight. Furthermore, the design of the system allows for the customisation of queries.

Users can save their preferred filters and settings for future reference, streamlining the process of data retrieval and analysis. The system also includes an alert feature, which can be configured to notify users of specific patterns or thresholds being met. This functionality enhances the proactive

monitoring capabilities of financial supervisors, enabling them to respond swiftly and effectively to emerging trends or issues within the financial landscape.

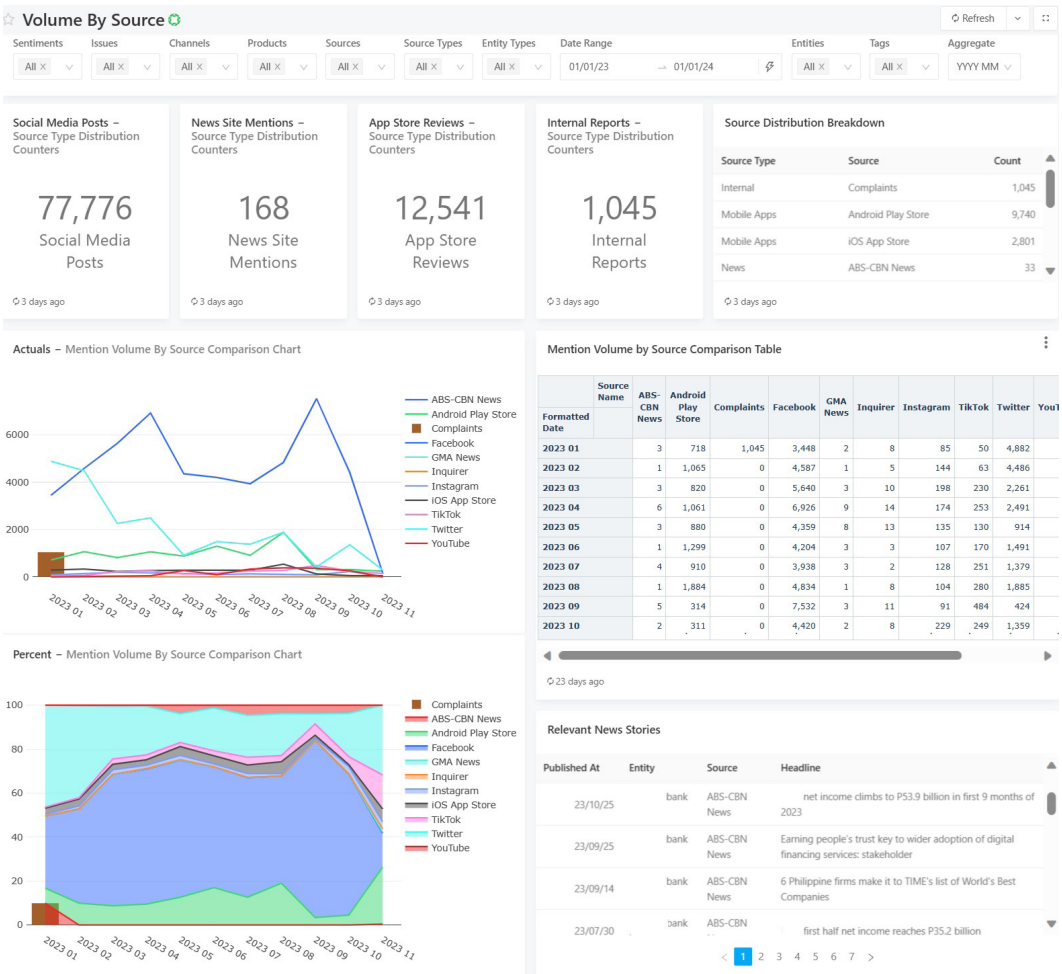
Regarding the real-time aspect of this system, the envisioned production configuration enables supervisors to collect and analyse social media data, app store reviews, websites, and locally sourced data such as complaints and interactions from the supervisor's own forms and chatbots, all in real-time giving users a constantly updated visualisation of their supervisory landscape.

The solution is designed to meet diverse operational and regulatory demands through its flexible deployment options. It can be set up on-premises in a co-located server, within a secure cloud environment, or in a hybrid configuration. This adaptability positions the system to align within legal constraints, data security standards, and the specific needs of the supervisory authority. Such automation ensures that the system remains current with the latest changes and trends in web and social media platforms, enhancing the efficiency and accuracy of data collection.

By automatically adjusting to new formats, terminologies, and data sources, the platform maintains a continuous and uninterrupted flow of relevant and up-to date information. This feature is particularly valuable in the fast-paced digital world where financial products and services, as well as consumer behaviours and sentiments, are constantly evolving.

The business intelligence dashboard in Figure 4 displays time-series charts of aggregate issues, topics and financial channels with real time data represented in actual figures and percentages.

FIGURE 4. WINNOW'S WORKING PROTOTYPE VOLUME BY SOURCE DASHBOARD



The AI-driven sentiment analysis categorises posts as positive, neutral and negative sentiment, helping supervisors stay ahead of emerging situations by observing spikes in negative sentiment.

The granularity of the dashboards enables supervisors to focus on particular areas of interest or concern. The filters can be applied to both the visualised data and the underlying raw data, which is meticulously stripped of any personally identifiable information and other sensitive data to ensure compliance with privacy regulation.

The dashboard displays volume filtered by source, dates, sentiment, issues, channels, products, source type, entity, and entity type as classified by the AI-model developed for each agency. For example, users can filter to search for specific issues related to account management, fraud, loan restructuring, poor customer service, unauthorised disclosures to third party, or unauthorised online transactions. Or by channels such as ATMs, online, over the counter/branch issues. Issues relating to products such as credit card, deposit, e-money wallets, loans. The correlation table combines the public data with internal datasets, providing deep context for emerging issues.

Figure 5 shows additional graphs developed for the working prototype. Supervisors can also configure dashboards to visualise data from different perspectives, observe which social media platforms are most often used by customers, and find trends with specific entity banks or products. Additionally, they can download raw data for a more granular analysis powered by continuous monitoring which updates instantly with incoming data to provide a near real-time view.

> *"We've chosen to use the API because of its ability to collect that data and have the freedom to correlate it with our internal sources. The result of that is that we can keep data sovereignty or localisation by having our datasets protected by our data privacy laws. I think that speaks to the bigger insight about hybrid computing and not just cloud computing."*
>
> *Oliver Chato*
> *Director of Information and Communications Technology Department, SEC Philippines.*

FIGURE 5. WINNOW'S WORKING PROTOTYPE DASHBOARD

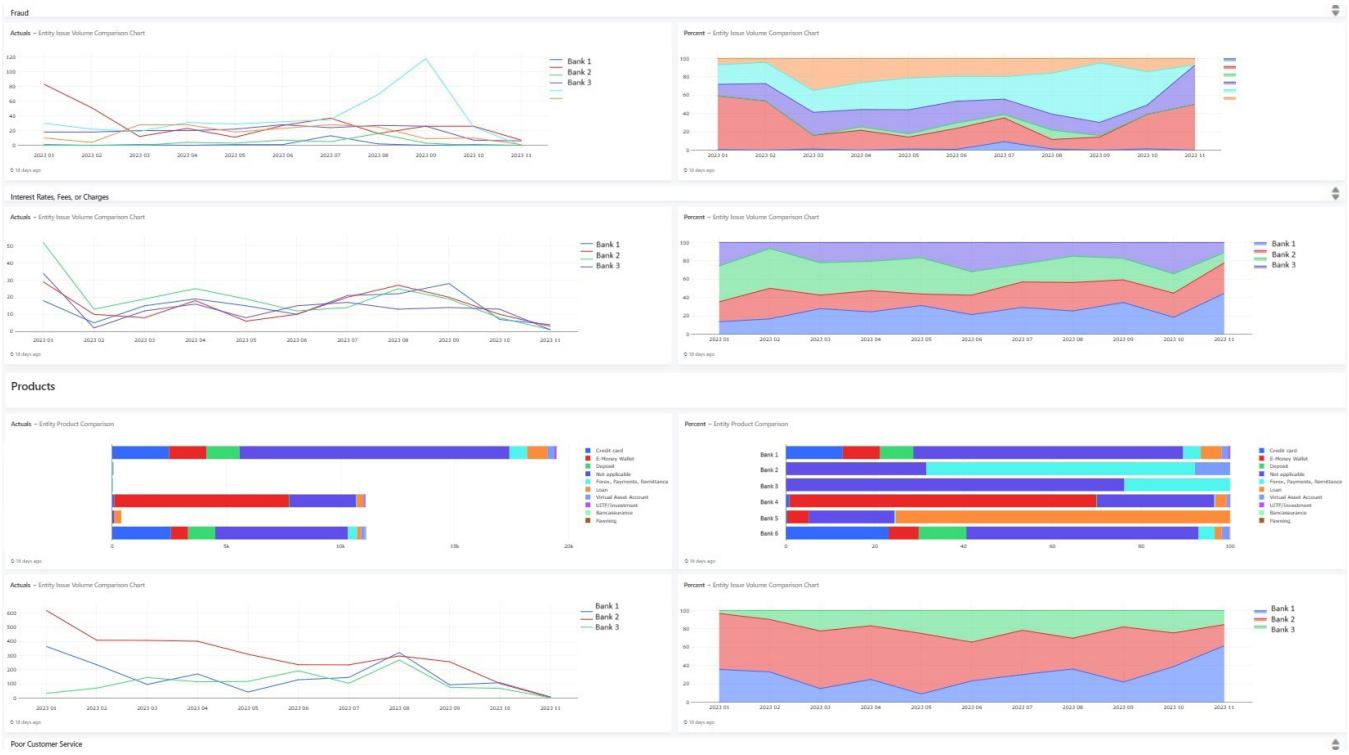


TABLE 1. WINNOW WORKING PROTOTYPE SCRAPING CAPABILITIES, POINTING THE SCRAPER AT THREE ENTITIES SUPERVISED BY BSP ENTITIES AND TWO ENTITIES SUPERVISED BY SEC

WEB SCRAPING	BSP	SEC
CHANNELS SCRAPED	Facebook Instagram News Sites TikTok Twitter Youtube IOS App and Android Play Store Reviews	Facebook Instagram Twitter
COMMENTS SCRAPED	90, 485	1,780

6. IMPACT

The integration of social media and other public datasets into the supervisory process significantly enhances the monitoring capabilities of financial authorities by enabling them to process and analyse large volumes of data effectively. The implementation of this technology is pivotal for supervisors, offering them insights into public discussions and perceptions about the services, goods, and support provided by regulated entities. These discussions often take place on social media platforms and in online reviews, where customers frequently voice their experiences and opinions. This tendency arises from a general reluctance to lodge formal complaints, leading customers to express their concerns on more familiar and accessible communication channels.

By tapping into these sources of informal feedback, supervisors gain a more immediate and candid view of customer sentiments. This approach allows for a proactive stance in identifying and addressing potential issues with financial products and services. Instead of relying solely on formal complaint channels, which may only capture feedback after an issue has escalated, the use of Winnow's technology enables financial authorities to detect early warning signs and trends. As a result, they can act swiftly to address emerging problems, enhancing consumer protection, monitoring market conduct, and preserving the integrity and stability of the financial system.

Winnow's platform offers an advanced level of automation that adeptly adjusts to the dynamic nature of web platforms. This adaptability is crucial, as it alleviates the burden on supervisors of constantly maintaining and updating data collection mechanisms. With this automated adaptability, supervisors are better equipped to focus on interpreting the data and applying their expert judgement in financial supervision, rather than being bogged down by the technicalities of data collection and system maintenance. Instead, supervisors can concentrate on their primary responsibility: effective financial oversight.

Supervisors benefit from the ability to:

- Transform unstructured consumer experiences shared on social media into structured insights with minimal effort, improving the understanding of consumer grievances and trends.
- Automate the collection of social media data, unlocking invaluable insights into consumer preferences and sentiments, facilitating informed decision-making.
- Merge data from web scraping with other data sources to effectively identify risks, guiding the allocation of supervisory efforts based on the severity and surveillance of identified risks.
- Use fully integrated data to lay a strong foundation for the development of future policies, ensuring they are well-informed and targeted towards actual market needs.
- Continuously assess and adapt regulatory policies to align with evolving market dynamics and emerging consumer protection needs, ensuring regulations remain relevant and effective.
- Leverage an automated alert system to enhance surveillance capabilities, enabling swift identification and immediate response to new issues, irregularities, or potential risks to financial consumers.

7. WHAT'S NEXT

After the prototype was tested, both agencies confirmed that the solution accelerates their digital transformation by demonstrating the value of injecting artificial intelligence into the supervisory process, and that the prototype exercise helped them to better understand how to move into production.

The Lab's approach is designed to avoid vendor lock-in. BSP and SEC will continue to have access to the working prototype space and perpetual license to the object and source code of the solutions, including Winnow's data and analytics, as they consider the path to production. After the delivery of the prototype, the agencies have the options to i) continue working with the vendors that developed it, ii) contract different vendors, or iii) deploy in-house resources to further the development of the solutions and maintain it. The agencies are currently pondering which pathway to production is more suited for each of them.

> *"The analytical capability to provide structured insights into consumer experiences in financial services that would otherwise require significant effort to interpret. In the Philippines, social media mostly serves as a channel for consumers to voice their complaints. Implementing the prototype's features would allow automated collection to gather and efficiently utilise social media data, which offers invaluable insights into consumer preferences and sentiments.*
>
> *Integrating data from the web scraping application with the various other data sources would significantly contribute to risk assessment initiatives and provide guidance in prioritising supervisory efforts based on the severity and surveillance of identified risk. Also, fully integrated data from the web scraping application and various added data sources form a valuable foundation for shaping future consumer protection and market conduct policies.*
>
> *This comprehensive dataset also allows for continuous evaluation of regulatory policies adapting to evolving market dynamics and addressing emerging consumer protection needs.*
>
> *The automated alert system feature would enhance real-time surveillance capabilities and would facilitate prompt identification and immediate response to emerging issues, irregularities or potential risks affecting financial consumers."*
>
> *Alvin Tolosa*
> Consumer Protection and Market Conduct
> Off ce of the BSP

Bangko Sentral ng Pilipinas

BSP promotes and maintains price stability, a strong financial system, and a safe and efficient payments and settlements system conducive to sustainable and inclusive economic growth. In attaining this mission, one of the BSP's mandates is to ensure that appropriate mechanisms are in place to protect the interests of consumers of financial products and services under conditions of transparency, fairness, and sound market best practices.

BSP has pioneered suptech innovation, having developed two significant projects in partnership with the RegTech for Regulators Accelerator (R2A, now the Cambridge SupTech Lab). The first project has enhanced the efficiency and accuracy of BSP's reporting system by using APIs to streamline the collection and analysis of financial data from all regulated banks. The second project has modernised consumer protection by developing the first chatbot-assisted integrated complaints management system for financial authorities.

The Securities and Exchange Commission of the Philippines

The SEC is the national governmental agency responsible for supervising the corporate sector, capital market participants, securities and investment instruments markets, and for protecting the investing public. As the registrar and supervisory authority of the Philippine corporate sector, the SEC also develops and regulates the capital market, an integral element of the Philippine financial system and its economy. The SEC's mission is to foster a competitive and secure environment that facilitates swift and straightforward company registration, efficient capital formation, broad investor participation, and innovative business practices.

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 scalable integration of innovative technologies, data science and agile methodologies by supervisory authorities to address the enduring and emerging challenges of the rapidly changing financial landscape. Through the Lab, financial authorities have championed the adoption of advanced suptech solutions that tackle critical issues such as financial crime, fraud, exclusion, climate change enablers, consumer protection, and artificial intelligence biases.

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 (2024), Case Study: Financial Consumer Protection Suite with Web Scraping and Machine Learning-Based Analysis, Cambridge: Cambridge Centre for Alternative Finance (CCAF), University of Cambridge. Available at www.cambridgesuptechlab.org

AUTHOR
Nathalie Lenehan

DESIGN
Dayna Donovan

ADDITIONAL CONTRIBUTORS
Simone di Castri, Matt Grasser, Samir Kiuhan
Adriana Senior