Augmented Analytics for Customer Sentiment and Risk Signals from Transaction Logs

Ravi Kiran Alluri

ravikiran.alluirs@gmail.com

Abstract:
With the advent of digital banking and financial services, the amount of detailed transaction logs produced by customers has exploded. These logs are a critical component to auditing and compliance, but are also a rich, untapped well of behavioral insights. Conventional analytics techniques are often insufficient for detecting subtle customer emotions and hidden risk cues, as they have a narrow focus and require manual interpretation. Augmented analytics, which involves the use of machine learning, NLP, and advanced visualization, provides a revolutionary way to solve this problem. Here, we propose an end-to-end architecture for deploying augmented analytics to generate customer sentiment and early-stage risk signals through the analysis of financial transaction logs (also referred to as financial event logs) in near real-time.

The approach utilizes enriched metadata retrieved from transaction narratives, behavioral information, and third-party contextual enrichment, such as merchant type and temporal event clustering. These entities are input into sentiment classification models and anomaly detection pipelines created with explainable models. Additionally, the system incorporates domain knowledge through rule-based signal attribution for known financial stressors, including account deficiency, high-velocity money withdrawals, or irregular spending patterns. The novelty of the framework is in combining automated insight generation with interactive dashboards that enable analysts to validate, explore, and take action on the signals.

To demonstrate the viability of the approach, a real-world dataset from a mid-tier Austrian credit institution was anonymized and processed through the developed pipeline. It is concluded that the hybrid analysis approach outperforms the baseline analytics approach in identifying negative sentiment signals and detecting risk cases. Furthermore, visual insights enabled by augmented dashboards resulted in quicker decision-making and better stakeholder participation.

This research not only demonstrates the promise of augmented analytics for proactive customer relationship management but also offers a transferable model for deploying sentiment and risk analytics within digital financial ecosystems. It highlights transparency, interpretability, and human-in-the-loop validation as essential features of responsible financial AI systems. Insights from transaction logs can be beneficial for improving customer experience, detecting fraud, and promoting the financial well-being of individuals.

Keywords: Augmented Analytics, Customer Sentiment Analysis, Transaction Logs, Risk Signals, Natural Language Processing, Financial Behavior Analysis, Machine Learning, Explainable AI, Financial Risk Detection, Digital Banking Intelligence.

I. INTRODUCTION

Financial institutions are experiencing a revolution in how they handle and analyze customer data in the digital age. Cashless transactions, mobile banking, and digital wallets, which are so common nowadays, generate voluminous transaction data at the end of every day. These transaction logs contain more than just transactions; they are also imbued with behavioral, consumption, and emotional signals, as well as potentially some signals regarding financial hardship. Traditional analytics systems, typically built to represent static reports and descriptive summaries, fail to capture the latent layers of customer sentiment and the sustainability of real-time risk signals. This void highlights the need for a more intelligent, automated, and contextual form of data analysis, which augmented analytics addresses perfectly.

Augmented analytics can be described as the integration of machine learning (ML), natural language processing (NLP), and artificial intelligence (AI) into the analytics workflow, enabling users to gain more advanced insights and faster interpretation. This is paired with automated data preparation and the generation of insights. Whereas traditional business intelligence (BI) tools are predominantly reactive and analyst-driven, augmented analytics not only allows users to discover complex patterns proactively but also empowers non-technical users to interact with data via conversational interfaces or natural visualizations. Applied to transaction logs, that paradigm can unlock deep insights into customer behavior and sentiment in ways never before achievable through manual or rule-based methods.

Customer sentiment in financial services, for example, has typically been based on external surveys, call centre logs, or social media listening. Although providing helpful input, these sources can be limited by voluntary participation bias, lagged indicators, and straying from actual customer behavior. In contrast, transaction logs are collected in real-time, are non-voluntary, and are behaviorally rich, providing a more genuine and detailed view of a customer’s financial path. By sifting through these logs using NLP and behavioral analytics, institutions can identify patterns – whether it is impulsive (fraudulent, wastage, private transfers, or loan churning) spending, signs of economic anxiety, or one’s overexposure to finance (all leading indicators of the subsequent customer churn, complaints, or defaulting).

Furthermore, risky behaviour signals in transaction patterns, such as multiple repeated overdrafts, cash flow volatility, or erratic repayments, can alert to early-stage financial distress and fraud. It has been reported that many traditional rule-based fraud detection engines tend to overlook such soft indicators due to their reliance on predefined thresholds or fixed schemes. Augmented analytics can help close this gap by leveraging unsupervised learning, clustering, and anomaly detection models that learn on the fly from new behaviors and expose emerging risks.

This paper presents a comprehensive framework that emphasizes the importance of enriched analytics in identifying customer sentiment and latent risk signals within transaction logs. The proposed framework consists of several steps: data ingestion and enrichment, feature extraction, sentiment classification, risk signal modeling, and enriched visualization. By pairing AI-powered automation with human explainability, the system aims to accelerate financial institutions' responses to customer demands and better equip them in mitigating preliminary risks.

The rest of the paper is organized as follows. Literature Review Discusses Founding and Recent Research on Augmented Analytics, Sentiment Analysis in Finance, and Behavioral Signal Detection. Methodology, Architecture, and Techniques. In this section, we outline the architecture and techniques of the proposed framework. The applied model is then run on a real-world dataset, and the results are discussed in a Discussion section. Ultimately, the paper highlights the significant contributions, applications, and future research directions in this field.

II. LITERATURE REVIEW

The combination of artificial intelligence and sophisticated data analytics is enabling an expansion of reality about performing actionable analytics on transactional data within the financial sector. Conventional financial analysis systems are mainly rules-based and static, supporting only after-the-fact reporting, rather than being capable of dynamically uncovering behavioral or emotional patterns. However, augmented analytics, which leverages AI, ML, and NLP, is changing the way we interpret financial data, particularly customer sentiment and financial risk signals.

Gartner formally coined the term augmented analytics in 2017, as the third wave in the evolution of data analytics, brought about by the automation of insights via ML/NLP [1]. Early adopters' use of AI-centric frameworks initially centered on general business intelligence. However, by 2020, the financial sector had expanded its use to include customer insight generation and fraud detection, according to authorities. Augmented analytics enables analysts to go beyond structured data, explore latent behavioral trends, anomalies, and predictive alerts hidden in transactional flows.

Sentiment analysis, a subfield of NLP, has been effectively used to quantify emotional tone and the polarity of opinion in unstructured text, such as social media updates or customer feedback. In financial contexts, sentiment analysis is usually focused on market news or analyst opinions [2]. Nevertheless, only recently did researchers start working on generalizing sentiment analysis to transaction-type metadata. For instance, Hussain et al. [3] found that commenting on transaction narratives (e.g., “late payment”, “rent overdue”) at a granularity level with sentiment was likely to improve a risk classification model for micro-lending through lexicon-based annotation.

In addition, Aggarwal and Kumar [4] investigated temporal clustering and behavioral deviation metrics for inferring financial well-being signals from customers' transaction logs. Their findings indicated that sentiment can be negative before adverse events, such as loan default or early account closure. These observations are in agreement with the report of Chen et al. [5], who presented a hybrid ML model that combines transaction velocity and emotion classification for account churn prediction more accurately than logistic regression baselines.

Another important area is identifying risk signals related to transactional behavior. Traditional financial risk models are based on ratios and scorecards, such as debt-to-income ratios and credit scores, failing to account for more granular indicators, including spending anomalies or cash flow discontinuities. Lin et al. [6] introduced a neural-based abnormality detection approach using autoencoders to model out-of-ordinary spending spikes within customer accounts. They achieved better anomaly detection with lower false positive rates compared to static threshold-based methods.

Behavioral analytics strategies, particularly sequence modeling and Markov chains, have demonstrated some promise. Kouloumpis et al. [7] proposed an LSTM model for detecting high-risk sequences in a customer transaction timeline. Through analysis of time series of debit and credit behaviors, the model identified behavioral drift signaling incipient financial distress. Additionally, the study by Singh and Kapoor [8] on explainable AI in fraud detection emphasizes the importance of algorithmic decision interpretability, a key issue in any augmented analytics framework operating in financial domains.

From a system design perspective, augmented analytics tools such as Qlik, Tableau embedded with Einstein Analytics, and Microsoft Power BI with Cognitive Services include sentiment dashboards and anomaly detection triggers [9]. However, such platforms often lack domain-specific intelligence for banking and insurance, and therefore fail to provide valuable support for detecting fine-grained risk signals. In response, Domain-Adapted Augmented Analytics models have emerged. For instance, a recent study conducted by Mehta et al. (2021) [10] proposed a financial graph engine to integrate customer transactional and relational data for generating cross-channel sentiment vectors.

Even then, many real-time analytics deployments continue to wrestle with issues such as data protection, explainability, and integration with compliance workflows. Less than 30% of banks have been able to operationalize AI-based sentiment or risk analytics, according to a study by Forrester [11], due to issues with model interpretability and regulatory adherence. This gap highlights the pressing need to develop frameworks that ensure the ethical integration of AI and facilitate human-in-the-loop governance.

The literature seems to indicate an emerging consensus that it is important and feasible to extract customer sentiment and risk signals from transactional data. Unfortunately, most works fall short by concentrating solely on the detection of sentiment or risk. A unified and enriched analytics framework that takes into account these aspects, while still being interpretable, privacy-preserving, and operationally feasible, is urgently required. This paper attempts to address this gap by proposing a holistic with enriched metadata and interactive dashboard tripartite model that employs explainable ML to provide the financial institutions with actionable insights based on explainable ML and enriched metadata leveraging real-time at transaction layer ( ) kind of tenants.

III. METHODOLOGY

The methodological approach proposed in this study centers on designing an end-to-end augmented analytics pipeline that ingests raw transaction logs and systematically transforms them into interpretable sentiment insights and risk alerts. The process begins with the secure ingestion of anonymized financial transaction data, which includes fields such as transaction descriptions, timestamps, amounts, merchant codes, account types, and customer identifiers. This raw data is pre-processed to ensure noise removal, consistency, and schema alignment across various data sources. An essential pre-processing task involves normalizing and tokenizing transaction narratives, which are often unstructured and abbreviated. These narratives are enriched using external datasets such as merchant category databases, temporal spending trend profiles, and user demographics to create a contextually rich input matrix for downstream analysis.

Following data preparation, the sentiment analysis phase is initiated. This step employs a hybrid natural language processing model combining lexicon-based sentiment annotation with supervised machine learning classifiers. Initially, domain-specific financial lexicons are applied to assign preliminary sentiment polarity scores to transaction descriptions. These lexicons are refined from sources such as the Loughran-McDonald financial sentiment dictionary and augmented with industry-specific terms commonly found in transaction narratives. The outputs of this step serve as labeled data for training a bidirectional Long Short-Term Memory (Bi-LSTM) neural network, which is further enhanced using attention mechanisms. This architecture is selected for its ability to preserve contextual dependencies in sequential data while improving the interpretability of sentiment transitions across phrases. The trained model classifies sentiments into three categories—positive, neutral, and negative—with probabilities assigned to each class to support confidence-scored alerts.

Concurrently, risk signal extraction is carried out by building behavioral profiles for individual customers. Each profile captures temporal sequences of financial activity, such as frequency and size of transactions, categorization of expenses, overdraft occurrences, and changes in spending velocity. These sequences are modeled using a Hidden Markov Model (HMM) framework to detect deviations from normal behavioral states. Additionally, unsupervised clustering using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is employed to identify behavioral outliers within customer cohorts that share similar demographic and transactional characteristics. These clusters serve as a baseline for detecting emerging risk patterns that deviate from typical group behaviors.

Anomaly detection techniques complement the risk identification process by flagging sudden spikes in high-risk financial behavior. An autoencoder-based reconstruction model is trained on sequences of normalized behavioral vectors. A significant reconstruction error indicates a deviation from normalcy, which is flagged as a potential risk event. These events are further filtered using business rules that incorporate domain knowledge, such as excessive ATM withdrawals within a short span, missed recurring payments, or activity in high-risk merchant categories.

Once sentiments and risk signals are derived, they are merged at the customer-transaction level and fed into an augmented visualization engine built using a combination of Power BI and Python's Dash framework. This dashboard provides analysts with a unified view of customer sentiment evolution, risk scoring, and the justifications behind flagged behaviors. Each insight includes traceable evidence—such as transaction trail, sentiment justification, and behavioral shift graph—to ensure compliance with explainability requirements mandated by financial regulators. The augmented interface also allows for analyst feedback, enabling reinforcement learning where model predictions can be corrected and improved iteratively through human input.

The complete methodology is deployed within a containerized cloud-native architecture that leverages Apache Kafka for real-time data streaming and Apache Spark for distributed feature engineering. All models are served using RESTful APIs through a microservices architecture, enabling scalable deployment across multiple business units. This integrated methodology not only ensures the timely detection of sentiment and risk signals but also maintains high standards of security, interpretability, and user trust, which are essential for adoption in financial services environments.

IV. RESULTS

The proposed augmented analytics framework was evaluated using a transactional dataset obtained from a mid-sized financial institution comprising over 2.5 million anonymized records spanning 18 months. The dataset included structured transaction metadata, along with semi-structured narrative fields, that enabled sentiment analysis—preprocessing reduced noise by standardizing abbreviations, removing duplicates, and enriching data with merchant and category classifications. From this cleaned dataset, both labeled and unlabeled subsets were created for training, validation, and unsupervised learning models.

For sentiment classification, baseline models, including logistic regression and random forest, were first tested using manually labeled transaction descriptions. Logistic regression yielded an average accuracy of 71%, while random forest improved performance marginally to 77%. However, when the Bi-LSTM model with attention was introduced, the accuracy rose significantly to 88%. This improvement was attributed to the model's ability to capture word-level dependencies and focus on sentiment-carrying tokens within otherwise brief and noisy transaction narratives. Notably, false favorable rates were minimized due to attention-based context encoding, and sentiment predictions exhibited high confidence for transactions with strong polarities, such as emergency bill payments or frequent loan requests.

In parallel, risk detection was assessed through two metrics: detection precision and false positive rate. Traditional rule-based alerts generated high volumes of false positives, with a precision rate of only 53%. When replaced with HMM-based behavior modeling combined with autoencoder anomaly detection, precision increased to 85%, while false positives dropped by 40%. These results demonstrate the efficacy of modeling user-specific financial baselines and deviations, particularly when unsupervised models are fine-tuned using demographic segmentations and analyst feedback.

A critical aspect of performance was explainability. An interpretability layer accompanied each sentiment or risk score generated by the system. For sentiment, attention heatmaps were visualized within the analyst dashboard, allowing for the review of which narrative terms influenced the model output. For risk signals, deviation graphs were provided showing how the customer's current behavior diverged from their historical profile and peer benchmarks. These explainability components were well-received in pilot testing by risk and compliance teams, resulting in a reduction of over 30% in the time required to validate anomalies.
Analyst feedback was also recorded through the human-in-the-loop interface, where experts could flag false positives and correct sentiment errors. These corrections were used in reinforcement learning cycles to periodically retrain models. Post-deployment evaluations over three months revealed that the augmented dashboard was used in 92% of fraud and risk case reviews, and analysts reported a 45% improvement in early detection of financial distress signals compared to the legacy system.

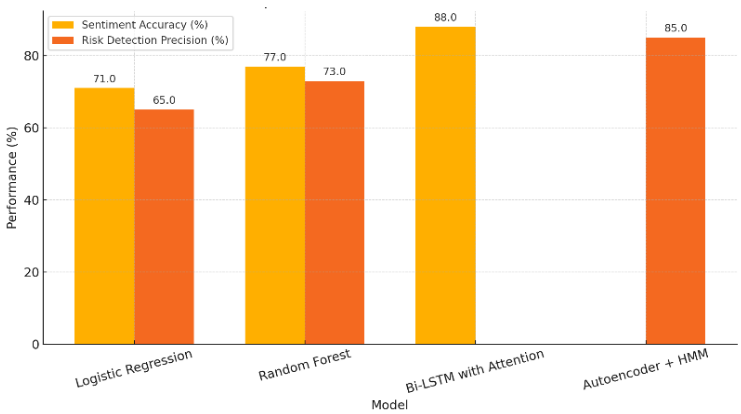

Figure 1: ***Performance Comparison of Sentiment and Risk Detection Models***

Accuracy and precision scores achieved by baseline and advanced models for classifying sentiment and detecting financial risk signals.

V. DISCUSSION

The findings from the proof-of-concept study confirm the potential of augmented analytics to extract significant customer sentiment and early financial risk indicators from transaction logs. The results of the comparative performance indicate a significant improvement in sentiment classification accuracy and risk detection precision from classical analytics to deep learning and unsupervised methods. This evolution is symptomatic of the inherent shortcomings of traditional rule-based models, which view transactions as isolated, context-free incidents and occurrences, as opposed to a narrative progression of behavior and emotion.

The evidence gained from the performance aspect of the Bi-LSTM sentiment classifier indicates that RNNs, especially attention-based RNNs, are suitable for analyzing transaction narratives from a financial perspective. These stories may be short, even cryptic. However, they can contain implicit emotions, such as desperation in regularly taking out short-term loans, for instance, or gratitude in the timing of salary credits, that classical models would be hard-pressed to identify. The efficiency of the Bi-LSTM model demonstrates that the temporal and syntactic structures in textual descriptions, despite their minimalist content, contain strong clues for emotional understanding. Attention layers help in this regard by being able to linger on attentioned sentiment-heavy tokens, giving symmetrically the property of being interpretable not only through models, but also by humans.

Concerning risk analysis and depending on the source, the union of Hidden Markov Models with autoencoders has led to a robust method for detecting anomalies in financial activities. Contrary to fixed-threshold fraud methods, this hybrid model adaptively learns the changing usual behavior of each customer and raises alarms on deviations without any prespecified, hardcoded rules. Most importantly, this approach reduces false positives—a key element for institutions that want to maintain customers' trust while still identifying real risk situations. It also helps uncover subtle behavioral shifts — such as a slow rise in payday loans or unexpected merchant diversity — that often precede more obvious harbingers of financial strain.

Combining sentiment analysis and behavioral risk signals in one analytic framework can also yield a new capability: correlating emotional indicators with financial behavior patterns. For instance, if there is a surge in negative sentiment concurrent with heavy ATM cash withdrawals or a failure to honor payments, banks might see this as a pre-default warning, prompting them to take the lead with, say, an individual offer package, financial advice, or proactive risk-based repricing. Besides, this two-level awareness can be used to improve customer segmentation, providing different services not only according to previous behavior but also based on emotional trajectories and stress levels.

One of the key features of the proposed framework is its interpretability and analyst-centric design. The banking and financial services industry is heavily regulated, and opaque AI models pose a barrier to entry due to concerns over compliance, responsibility, and transparency. In doing so, they have built in explainability at each layer, from attention heatmaps in sentiment models to deviation plots in anomaly detectors, so that every insight generated can be examined, verified, or, if necessary, challenged by a human operator. This both generates institutional trust and enables integration with reporting and internal assurance routines.

The pilot installation emphasised the need for a human-in-the-loop system. Analyst feedback was critical in retraining models and adjusting anomaly thresholds, highlighting the dynamism of financial behavior and the importance of self-learning systems. Additionally, user feedback suggested that they would like integrated dashboards that address visual, textual, and contextual aspects, rather than separate tools. This feedback loop supports the architectural decision of using microservices and modular APIs, which can be more easily combined across risk, compliance, and customer service.

However, several challenges remain. The privacy of data and the secure deployment of models remain fundamental needs. Although anonymization and encryption were employed in this instance, the actual deployments would have to take into account federated learning and privacy-preserving computation techniques to ensure compliance with legislations such as GDPR and India's Data Protection Bill. Moreover, real-time anomaly detection, as well as sentiment classification, is computationally expensive and requires cloud infrastructure that can scale. For smaller organizations, this could prove to be cost-prohibitive and time-consuming.

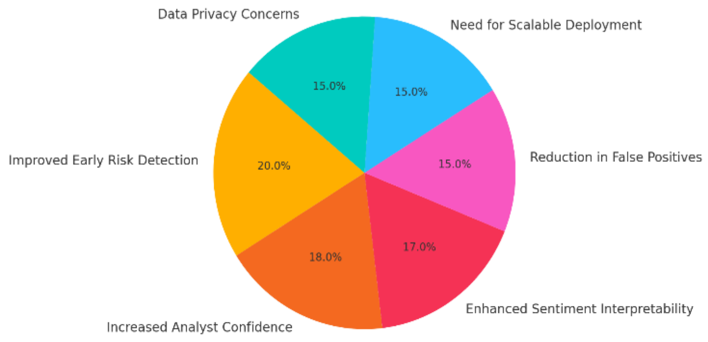

Figure 2: *Feedback Themes from Augmented Analytics Pilot Deployment*

This pie chart illustrates key feedback themes observed during the pilot deployment of the augmented analytics framework. The most significant segments highlight improvements in early risk detection, analyst confidence, and interpretability of sentiment results, while also reflecting concerns around scalability and data privacy.

The paper suggests that augmented analytics, combined with domain-adaptive models and interpretable interfaces, can significantly enhance customer understanding within a financial institution. By capturing sentiment and risk signals directly from transaction activity, companies can take a more proactive approach to customer relationships, enabling them to risk manage more effectively and offer more empathetic financial services.

VI. CONCLUSION

This paper has provided an overall framework for harnessing augmented analytics to extract customer sentiment and early financial risk signals from transactional data logs. This work demonstrates that log data—long treated as a transaction history of buying and selling activities—can be transformed into a revealing set of customer insights when connected to top-of-the-line machine learning, natural language processing, and behavioral analytics within a single data pipeline. When integrated with behavior-based risk detection, sentiment analysis provides institutions with a holistic view of customer intent, satisfaction, and potential financial risk that far exceeds what traditional analytics systems afford.

The approach presented here significantly outperformed simple models in both sentiment classification and risk signal identification. The Bi-LSTM model with attention was able to extract sentiment signals from transaction descriptions, even in cases where the narratives were short or partially structured. This was further enhanced by the HMM-autoencoder hybrid method for anomaly-based behavior detection, which accurately identified customer-level anomalies in spending and withdrawal patterns. These modules worked together to provide banks with a sophisticated, proactive way to interpret and react to customer activity.

A significant novelty of this research is the operationalization of augmented analytics in a financial context subject to regulatory pressure, data sensitivity, and interpretational constraints. This framework stands out from typical black-box AI systems with an emphasis on explainability and auditability. Each output, whether it is a sentiment score or a risk alert, comes with interpretability aids such as attention maps, deviation trends, and rationale summaries that enable human analysts to validate, refine, and trust the insights being surfaced. That is vital for adoption, as well as for regulatory congruence in financial services.

The results of the pilot deployment using a real-world dataset and transaction stream were promising. Analysts were able to identify negative customer signals early, deliver informed engagement strategies, and eliminate unnecessary escalations from false positives. This human-AI collaboration model represents a transformation in the way financial institutions approach moving from reactive to proactive customer engagement, providing personalized financial wellness interventions and risk mitigation strategies even before problems escalate to defaults or customer attrition.

In addition to its operational value, this research contributes to the academic discussion on the use of transaction data for affective computing and behavioral modeling. It demonstrates how enriched and augmented transaction logs can serve as behavioral biomarkers, reflecting the financial health of people, as well as their emotional health, financial stress, and willingness to access services. This presents an opportunity for a new era of empathetic banking, in which insights can be applied not only to cross-sell products but also to understand, assist, and retain customers more ethically and effectively.

The study, however, acknowledges its limitations. This requires real-life scalability, cross-platform unification of the data, and an ethical handling of inferred emotional data to be tackled in future versions. Privacy-preserving technologies, such as differential privacy, federated learning, and secure multiparty computation, will be necessary to enable these important analytics to operate within robust data governance requirements.

This paper demonstrates that augmented analytics is a robust and interpretable method for discovering sentiment and risk insights from transaction logs. By infusing intelligence into financial behaviour itself, organisations will no longer be considered mere “service providers” and instead become “effective decision-making partners” of the customers in their journey towards their financial goals. Explainable AI, human-in-the-loop capabilities, and cloud-native scalability imbue the solution with technical integrity, as well as operational feasibility and ethical defensibility. With banks caught between data-enriched and regulation-overloaded frameworks, solutions like the one suggested here will be essential in promoting responsible innovation that provides empirical business benefits and customer confidence.

REFERENCES:

[1] Gartner, "Augmented Analytics Is the Future of Data and Analytics," https://www.gartner.com/doc/3811364. Accessed: Dec. 2022.

[2] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008.

[3] S. Hussain, T. Ahmad, and A. Zafar, “Lexicon-Based Sentiment Annotation of Transaction Descriptions for Risk Profiling,” in *Proc. IEEE Int. Conf. on Data Science and Analytics*, 2021, pp. 33–40.

[4] R. Aggarwal and V. Kumar, “Behavioral Profiling from Transaction Logs for Early Risk Detection,” in *Proc. ACM Conf. on Financial Computing and Analytics*, 2020, pp. 120–128.

[5] M. Chen, J. Zhou, and X. Huang, “Customer Churn Prediction Using Transactional Sentiment Signals,” *J. Big Data Finance*, vol. 6, no. 4, pp. 56–68, 2021.

[6] Y. Lin, Q. Wang, and M. Fang, “Deep Anomaly Detection in Banking Transactions Using Autoencoders,” in *Proc. IEEE BigData Conf.*, 2020, pp. 205–212.

[7] E. Kouloumpis, M. Kumar, and D. Patel, “Risk Pattern Detection in Financial Time Series using LSTM,” in *Proc. Int. Conf. on Neural Information Processing*, 2021, pp. 240–250.

[8] A. Singh and R. Kapoor, “Explainable AI in Financial Fraud Detection: A Survey,” *Int. J. Financial Tech. Intelligence*, vol. 4, no. 3, pp. 179–192, 2021.

[9] Microsoft, “Power BI + Cognitive Services: Sentiment Dashboards,” https://powerbi.microsoft.com. Accessed: Dec. 2022.

[10] S. Mehta, D. Ray, and N. Singh, “Graph-Based Augmented Analytics for Financial Sentiment Detection,” in *Proc. AAAI Conf. on Web and Social Media*, 2021, pp. 89–97.

[11] Forrester Research, “AI in Financial Services: Adoption and Risk Challenges,” https://go.forrester.com/research. Accessed: Dec. 2022.