ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 5 May 2024

Advanced Data Technologies in Banking: A Comparative Review and Methodological Analysis

[1] P. Abhiram, [2] Narsimha Reddy Alipeddi, [3] Kailash Guguloth, [4] Ramakrishna Kolikipogu, [5] K.H. Vijaya Kumari*

[1] [2] Department of Information Technology, Chaitanya Bharathi Institute of Technology (CBIT), Hyderabad, India
Corresponding Author Email: [1] pedamalluabhiram@gmail.com, [2] narsimha9788@gmail.com, [3] gogulothkailash@gmail.com, [4] ramakrishna it@cbit.ac.in, [5] vijayakumari it@cbit.ac.in

Abstract*— In the evolving landscape of banking, the integration of advanced data technologies and frameworks stands as a testament to the sector's commitment to innovation and efficiency. It offers a comprehensive exploration of the advancements in data management within banking, underpinned by a thorough analysis of literature and comparative studies. It underscores the significance of data-driven decision-making in modern bank- ing, the paper delves into a curated selection of studies and methodologies that have shaped the contemporary banking data ecosystem. Methodologically, the paper employs a systematic approach to dissect and compare various data management technologies, providing insights into their efficacy, limitations, and potential implications for the banking industry. Through this lens, the review elucidates the pivotal role of advanced data technologies in enhancing customer experiences, optimiz- ing operational processes, and mitigating risks. Furthermore, it underscores the importance of continuous evaluation and adaptation, positioning banks to navigate the challenges and opportunities of an increasingly digital financial landscape.

Index Terms*— Data Management, Advanced Technologies, Banking, Comparative Studies, Financial Applications, Decision-making, Innovation in Banking.

I. INTRODUCTION

In today's interconnected digital landscape, sentiment anal- ysis has become pivotal in deciphering emotions and opinions from vast online data. This technology categorizes textual content into positive, negative, or neutral sentiments, offering invaluable insights into public perception. The banking sector, undergoing rapid digital transformation, is leveraging this tool to enhance customer engagement, optimize operations, and drive strategic decision-making. As banks embrace sentiment analysis, it becomes a strategic imperative, ushering in a new era of data-driven banking.

The rise of digitalization has facilitated the proliferation of online banking platforms, mobile applications, and real-time transactional capabilities, reshaping customer expectations and redefining traditional banking paradigms. The banking industry, traditionally characterized by its rigorous regulatory frame- work and data-driven approach, now witnesses a paradigm shift towards customer-centricity, fueled by the digital rev- olution. In this context, sentiment analysis emerges as a game-changer, enabling banks to proactively gauge customer satisfaction, identify emerging trends, and mitigate potential

risks. By analyzing customer feedback and sentiments ex- pressed across digital platforms, banks can tailor their services, optimize customer experiences, and foster a more personalized relationship with their clientele.

By harnessing the power of sentiment analysis, banks can gain valuable insights into customer sentiments, prefer- ences, and behaviors, thereby enabling them to craft targeted strategies, optimize service offerings, and cultivate enduring customer relationships in an increasingly competitive and dynamic marketplace.

Furthermore, sentiment analysis proves indispensable in enhancing fraud detection mechanisms, identifying unusual patterns, and mitigating risks associated with online transac- tions. By leveraging real-time sentiment insights, banks can swiftly respond to customer grievances, address concerns, and cultivate a culture of transparency and trust in the dig- ital banking ecosystem.Moreover, in an era characterized by information overload and rapid technological advancements, sentiment analysis equips banks with the tools to navigate the complex digital landscape, distill actionable insights from disparate data sources, and stay attuned to evolving customer expectations and preferences.

Sentiment analysis stands as a cornerstone of the modern digital age, empowering the banking sector with actionable insights, and driving innovation. As banks continue to embrace digital transformation, the integration of sentiment analysis emerges not merely as an option but a strategic imperative, heralding a new era of data-driven decision-making and cus- tomer engagement.

The burgeoning digital transformation in the banking sector underscores the imperative to harness innovative technologies that enhance customer engagement, optimize operations, and drive strategic decision-making. Amidst the proliferation of online platforms and the exponential growth of digital inter- actions, understanding and interpreting customer sentiments emerge as paramount factors that influence organizational suc- cess and sustainability.The motivation to delve into sentiment analysis within the banking realm stems from its transmuting potential to

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 5 May 2024

revolutionize customer interactions, mitigate risks, and foster a culture of continuous improvement and inno- vation. By unraveling the intricacies of customer sentiments

expressed across diverse digital channels, banks can glean invaluable insights that inform product development, refine service offerings, and cultivate lasting relationships with their clientele.

Furthermore, the escalating prevalence of online fraud and cybersecurity threats accentuates the need for robust sentiment analysis frameworks that bolster fraud detection mechanisms, enhance security protocols, and safeguard the integrity of the banking ecosystem.In essence, the motivation to explore senti- ment analysis within the banking sector encapsulates a broader vision to leverage data-driven insights, enhance customer ex- periences, and forge a resilient and adaptive banking landscape that resonates with the evolving needs and expectations of contemporary consumers.

The primary objective of this paper is to critically evaluate and compare three distinct methodologies employed in senti- ment analysis within the banking sector. By systematically an- alyzing the methodologies' principles, advantages, limitations, and applications, the paper aims to elucidate their efficacy, relevance, and potential implications for enhancing customer engagement, risk mitigation, and strategic decision-making in the digital banking landscape. Through a comprehensive exploration of these methodologies, the paper endeavors to contribute valuable insights, foster scholarly discourse, and inform future research endeavors in the burgeoning field of sentiment analysis within the banking sector.

Tools like the Twitter API facilitate data collection from platforms like Twitter [1]. Platforms such as RapidMiner offer advanced data analysis functionalities [2], while Apache Kafka and Apache Spark provide capabilities for real-time data processing and large-scale data analysis, respectively [3].

The paper's overview follows this structure: Section 2 delves into the examination of relevant literature, Section 3 elucidates the methodologies employed in various research papers, and Section 4 provides the concluding remarks.

II. RELATED WORK

Sentiment analysis, also known as opinion mining, is an important business intelligence tool that helps companies improve their products and services. Benefits of Sentiment Analysis include providing objective insights, building better products and services, analyzing at scale and real-time results [4]. A. Mahmud et al. in [5] they have conducted sentiment analysis on public perceptions of the Padma Bridge, utilizing data gathered from Facebook, YouTube, and online news portals and have applied various machine learning algorithms and have introduced an innovative voting mechanism to ag- gregate sentiments produced by the models. Dailing Zhang et al. in [6] uses Sentiment Analysis for Credit Risk, and introduced an innovative approach by incorporating a Gener- alized Multiple Kernel Learning model to automate decision making but the only problem lies that the study's limitation lies in its reliance on a single dataset obtained from a Chinese commercial bank, and could potentially restrict the adaptability of the proposed method to various settings or data sources. Analysis of customer complaints and feedback from Indian

banks was conducted by Gutha Jaya Krishna et al. in [7] from the data extracted from "complaintsboard.com". Pre- processing of raw textual data,which encompassed techniques such as creating a document term matrix (DTM) through Term Frequency-Inverse Document Frequency (TF-IDF), employing an embedding model like Word2Vec, and utilizing a psycho- linguistic method known as Linguistic Inquiry and Word Count (LIWC).The results of the sentiment analysis can be used to improve the business functionality of Indian banks. Nadhila Idzni Prabaningtyas et al. in [8], Mahardhian Anjar Ligiarta et al. in [9] and Lekha R. Nair et al. in [10] has performed Sentiment Analysis using Twitter data.Prabaningtyas along with her co-authors created a n-gram model of of Go-Pay whereas Mahardhian Anjar Ligiarta along with co-authors Rapid Miner to assist with classification procedures and Lekha

R. Nair et al. used Apache Spark and Spark Streaming. The study of Prabaningtyas focuses on an aspect-based sentiment analysis of mobile payment reviews, specifically using word pairs or bigrams related to services and applications.

Amira Marouani et al. in [11] emphasizes the significant role of employing data science algorithms and methodolo- gies to analyze customer behavior in the banking sector. The research hones in on a case study conducted in Tunisia, employing statistical predictive models to probe into the interplay between loan status and various factors including loan type, loan purpose, customer location, and outstanding loan amounts.

Advanced Data Technologies have revolutionized the fi- nancial landscape, introducing transformative solutions and capabilities that redefine traditional banking and investment paradigms. Leveraging technologies such as Big Data analyt- ics, machine learning, and blockchain, financial institutions can now process vast volumes of data with unprecedented speed and accuracy. This enables real-time risk assessments, fraud detection, and personalized customer experiences. More- over, machine learning algorithms predict market trends, optimizing investment strategies and portfolio management. Blockchain, with its decentralized and immutable nature, ensures transparent and secure transactions, reducing inter- mediaries and associated costs. Collectively, these advanced data technologies empower the financial sector to innovate continuously, foster trust among stakeholders, and navigate the complexities of the modern digital economy.

Tabitha Kemboi in [12] and Narayana Darapaneni et al. in [13], have simply performed prediction of loan status using data analysis methods to predict loan status in a sample

IFERP connecting engineers...developing research

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 5 May 2024

dataset spanning from 2007 to 2011 from Lending Club loan.Narayana Darapaneni et al. in [14] also mentions its pricing model, revenue projections, and costs associated with development and operation. Its objective is to provide financial institutions with a Software as a Service (SaaS) solution, assisting them in making well-informed lending decisions. Arujothi et al. in [15] endeavoured to construct a credit scoring model to precisely evaluate credit-related information and effectively discern between individuals who default and legitimate cus- tomers. With the objective in mind, they utilized a fusion of

Min-Max normalization and the K-Nearest Neighbors (KNN) classifier as part of their approach.

P. Mukherjee et al. in [16] has tackled a crucial issue concerning P2P lending platforms by focusing on the identifi- cation of potential defaulters. Hamayel in [17] have performed identification of potential defaulters on P2P lending platforms using machine learning by using the data from Quds Bank in Palestine and emphasizing credit limitations and regulatory directives, the research seeks to improve the precision of loan approval forecasts.These discoveries offer significant perspec- tives for financial institutions aiming to enhance their loan approval procedures while adeptly managing potential risks. The goal of Sun, X. in [18] is to forecast loan repayment behaviour for borrowers using peer-to-peer (P2P) lending networks. The article discusses the difficulties encountered by the P2P lending sector when evaluating borrowers' cred- itworthiness, primarily because of the unsecured nature of the loans. Big Data in banking refers to the vast volume of structured and unstructured data generated daily, offering insights into customer behaviors, transaction patterns, and risk assessments. By harnessing advanced analytics and machine learning algorithms, banks can derive actionable insights to enhance customer experiences, optimize operational efficien- cies, detect fraudulent activities in real-time, and make data-driven strategic decisions. This transformational shift enables banks to remain competitive, drive innovation, and deliver personalized services tailored to individual customer needs in today's digital era.

Anusmita Poddar et al. in [19] underscores the signifi- cance of Big Data analytics within the financial and bank- ing sector, highlighting its instrumental role in decision- making processes. It elucidates how Big Data can contribute to informed choices in banking, encompassing areas such as customer segmentation, understanding spending patterns, promoting product cross-selling, enhancing risk management, performing sentiment and feedback analysis while strength- ening fraud management. Furthermore, the paper explores a wide range of applications of Big Data within the banking sector, elucidating the methodologies employed, and outlines the sector's current utilization while shedding light on the future prospects and opportunities associated with Big Data. The research methodology comprises the collection of both qualitative and quantitative data from various sources, includ- ing esteemed journals housing numerous use cases. Despite these valuable insights, it is essential to acknowledge certain limitations in the study. Firstly, concerns pertaining to data quality and storage are prevalent in numerous instances, which can hinder the effectiveness of Big Data analytics. Secondly, the study assumes a descriptive nature, relying on available use cases, this might not provide a complete overview of every aspect of Big Data analytics in the banking industry. Shweta Yadav et al. in [20], and Uma Maheswari et al. in [21] have utilized Big Data for one of the task mentioned by Anusmita Poddar et al. in [19]. Shweta Yadav et al. have proposed a methodology have proposed a methodology for assessing cred- itworthiness and measuring the performance of borrowers by

loan performance analysis and credit risk analysis. The specific focus was on variables like interest rates, loan purposes, loan status, and the borrowers' financial grades. Uma Maheswari et al. in [21] have applies Fuzzy logic and Big Data techniques for sentiment analysis of mobile product reviews collected from sources like Twitter, e-commerce websites, and Flipkart. Aspect-based classification and sentiment score calculation are performed, with adjustments for negation words and modifiers. Fuzzy Logic enhances sentiment classification, leading to categorization into various sentiment levels. Summaries are generated for aspect categories, and sentiment is predicted for each category.

The banking sector is undergoing a profound transformation driven by the infusion of advanced technologies and frame- works in data management. Furthermore, the emergence of cloud computing has revolutionized data storage and acces- sibility, allowing banks to scale their operations seamlessly while ensuring data security and compliance. Additionally, the incorporation of Artificial Intelligence (AI) and Machine Learning (ML) algorithms empowers banks to derive pre- dictive analytics, enhancing customer service through per- sonalized offerings and proactive risk management.Moreover, frameworks such as Apache Kafka streamline data pipelines, ensuring efficient data ingestion and processing across diverse platforms. The adoption of robust data governance frameworks ensures the integrity, quality, and security of data assets, align- ing with regulatory requirements and bolstering stakeholder trust.In essence, the integration of advanced technologies and frameworks in data management propels the banking sec- tor towards a future characterized by enhanced operational efficiencies, superior customer experiences, and innovative financial solutions.

H. Huang in [22] emphasized on understanding and estab- lishing an effective loan pricing mechanism for small enter- prises provided by city commercial banks. As an alternative approach, the article introduces the concept of relationship lending, which emphasizes the importance of enduring partner- ships between banks and small businesses. The article closes by highlighting the crucial nature of factoring in both risk and relational elements when setting loan pricing for small businesses.

Mallidi et al. in [23] explores the deployment of a real-time streaming platform in the financial services and banking sector through the use of Kafka. It explains the challenges faced by Banking, financial services and insurance (BFSI) enterprises and how Kafka can help overcome those challenges. The file also provides an overview of Kafka's architecture, its features, and how it can be used to build a real-time streaming platform. The input configuration parameters Han Wu et al. in [24] for the queueing based packet flow model proposed in this paper are log retention time, partition number, and batch size. The model helps estimate how specific configuration parameters affect the performance metrics of Kafka. It achieves this by employing a queuing model (M/P H/1) to assess the average

time it takes for a piece of data to traverse a Kafka cluster. Comparative studies and evaluations play a pivotal role

in the banking sector, offering insights into best practices, performance benchmarks, and areas of improvement. By juxtaposing different banking models, strategies, or technologies, these studies enable institutions to gauge their standing in the market, identify competitive advantages, and strategize for future growth. Through rigorous analysis of metrics such as customer satisfaction, operational efficiency, risk management, and financial performance, banks can benchmark themselves against industry peers, uncovering opportunities for innovation and optimization. Furthermore, comparative evaluations often encompass regulatory compliance, ethical considerations, and sustainability practices, ensuring banks align with evolving global standards. In essence, these studies foster a culture of continuous improvement, driving the banking sector towards enhanced transparency, resilience, and stakeholder value.

Hlaing et al. in [25] introduce a streamlined and adaptable data ingestion framework designed to handle substantial streams of data. They substitute Apache Nifi with StreamSets Data Collector (SDC) for data acquisition and employ Kafka for data distribution. In performance evaluations, SDC surpasses Apache Nifi in terms of data loading effectiveness and operator performance. Notably, the framework achieves an average throughput of 1.7 seconds, making it a robust choice for applications like real-time monitoring of breaking news articles, offering improved data flow and real-time capabilities. Sapthami et al. in [26], focuses on sentiment analysis, a computational method for classifying around 4000 reviews. The system offers multiple benefits, such as providing highly accurate reviews' assessments, graphically comparing algo- rithm accuracies, and outperforming existing systems regard- ing precision.In study conducted by Margocahyo in [27], the research process is for sentiment prediction is carried out using the Naive Bayes algorithm. Notably, the ”Techno-Insecurity” aspect achieves the highest accuracy at 86%, highlighting its significance in understanding customer sentiment.The study's findings lead to conclusions and recommendations that en- hance customer feedback interpretation. Boulegane in [28], the study addresses challenges in interpreting customer sentiment from online reviews. It suggests alternative metrics, such as sentiment indicators derived from text mining and word

counts, to provide more accurate insights.

III. METHODOLOGY

The suggested approach outlined in reference delineates the process of data acquisition and its subsequent examination. Mahardhian Anjar Ligiarta et al. [9] harnessed the Twitter API in combination with Rapid Miner for the retrieval of Twitter data. They conducted searches to identify tweets containing specific keywords, with data collection spanning from mid-February to the conclusion of February 2022.

To guarantee the quality and dependability of classification models, it is imperative to perform effective data preprocessing

III. In the study referenced as [9], the researchers applied a sequence of data refinement methods.These measures were employed to enhance the dataset and eliminate potential dis-

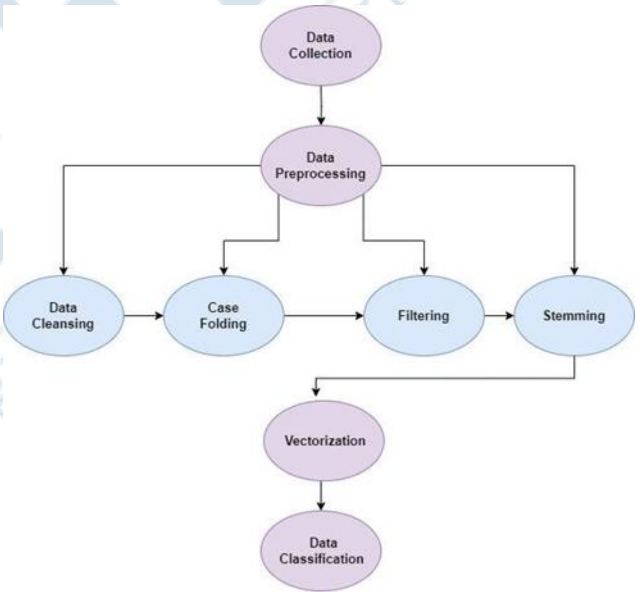

Fig. 1. Workflow

ruptions that could impact the performance of the classification models.

Data Cleansing: The initial step involved data cleansing, where systematic removal of elements such as unrecognized characters and URLs from the dataset was performed. This meticulous process was carried out to create a clean and consistent dataset, free from any irregularities or noise.

Case Folding: Following data cleansing, case folding was employed by converting all text to lowercase. This uniformity in letter casing is crucial to ensure that the model treats words with different cases as identical, avoiding unnecessary complexity and redundancy.

Stopword Filtering: Stopword filtering was applied using the Python library called Sastrawi. This step involved

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Volume 11 Issue 5 May 2024

removing common stopwords in the Indonesian language. By eliminating words with minimal semantic significance, the dataset was re- fined to contain only words with substantive meaning, thereby improving the focus of the research on relevant information.

Stemming: The final stage of data preprocessing involved stemming using the Sastrawi library. This procedure reduced words to their fundamental bases, promoting semantic co- herence while concurrently diminishing the dataset's dimen- sionality.Stemming is particularly valuable in languages like Indonesian, which exhibit rich inflectional morphology.

Vectorization After performing preprocessing, vectorization served as the pivotal process for transforming textual docu- ments into matrices, enabling their mathematical analysis. In this research article, The study utilized the widely recognized vectorization technique known as Term Frequency-Inverse Document Frequency (TF-IDF). In this research, the TF-IDF vectorizer provided by the Rapid Miner operator was selected as the preferred tool for vectorization within the proposed model.

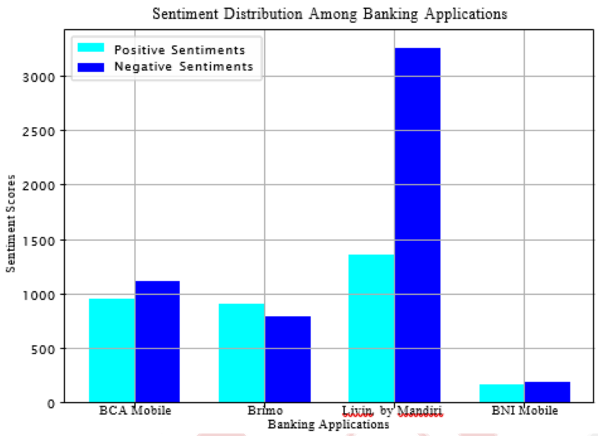

Fig. 2. Sentiment analysis

The TF-IDF (Term Frequency-Inverse Document Fre-rate of 92.5%, particularly in the context of identifying positive sentiment.

The research results pointed out that, in comparison to other platforms, Brimo showed the most prominent positive sentiment. In contrast, both Livin' by Mandiri and Brimo demonstrated noteworthy concerns regarding their dependabil- ity. As for user satisfaction, MB and BRI should prioritize reliability improvements due to the substantial prevalence of negative sentiment in this aspect. BCA, meanwhile, should concentrate on enhancing the utility of its mobile banking platform. Ultimately, BNI ought to place more emphasis on enhancing the responsiveness of its mobile banking system.

The scholars in [9] arrived at the determination that these observations may offer significant value to financial insti- tutions aiming to assess and improve their mobile banking systems. By doing so, they could foster positive sentiment, el- evate customer satisfaction, and potentially drive profit growth. Moreover, the general public can find this research useful quency) equation serves as a method for evaluating the sig-nificance of a term (word) within an individual document compared to a larger collection of documents (a corpus). It is a frequently applied technique in the realm of information retrieval and text analysis for gauging the importance of terms within documents. This formula comprises two essential elements:

Term Frequency (TF): This element quantifies the fre- quency of a term within a specific document. It is computed as follows:

$$TF\ (tm, doc) = \frac{\text{Number of occurrences of the tm in the doc}}{\text{Total number of tm in the doc}}$$

Inverse Document Frequency (IDF): This aspect assesses the significance of a term across a collection of documents. It is determined by the formula:

$$IDF\ (tm) = \frac{\text{Total number of doc in the corpus}}{\text{Number of doc containing the tm}}$$

when making choices regarding mobile banking platforms. Various social networking platforms may serve as data sources, and a variety of diverse Big Data software and non-relational databases can be employed for constructing the architecture. The design of the system is oriented towards the gathering and handling of substantial data originating from social media networks through a real-time, multi-platform social media input source, which incorporates a central component for extracting and analyzing sentiment [29].

The suggested structure comprises four primary elements: the Data Input Layer, the Analytics Layer, the Resource Supervision Layer, and the Storage Layer. The architecture is shown in figure III.

Here, "tm" signifies the specific word, and "Total number of documents in the corpus" denotes the overall count of documents in the collection. The TF-IDF score for a given "tm" in a particular "doc" is established by the multiplication of the TF and IDF components: TF-IDF (tm, doc) = TF (tm, doc) * IDF (tm)

In this research, SVM (Support Vector Machine) was chosen as the classification method for tweets due to its proficiency in managing high-dimensional data, its suitability for binary classification, its resilience to outliers, its capacity to handle non-linear separations using kernel functions, and its focus on optimizing the margin. SVM has a proven history of success in text classification tasks, particularly in sentiment analysis, rendering it a dependable option for the analysis and categorization of tweets.

The figure III above illustrates the distribution of senti- ments among different banking applications. The investigation carried out by [9] utilized three key performance assessment metrics: accuracy, precision, and recall. The model achieved a 92.5% accuracy rate, a precision score of 93.1%, and a recall

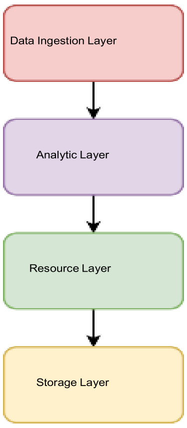

Fig. 3. Workflow

- **Data Input Layer:**
  - Apache Kafka is employed to gather and stream real- time data from incoming tweets, managing data with high velocity originating from various origins.
  - Apache Kafka efficiently distributes data across mul- tiple nodes in a cluster.
- **Analytics Layer:**
  - Apache Spark is utilized to establish sequences of Spark tasks for the processing of data received from Kafka.
  - Spark performs sentiment analysis on incoming data and extracts valuable insights.
  - Data cleaning techniques are used to remove irrele- vant data like stop words, punctuation, and special characters.
  - A lexicon-based classifier is used to categorize data into positive, negative, or neutral sentiments, utiliz- ing a pre-built sentiment lexicon.
- **Resource Supervision Layer:**
  - YARN clusters are utilized for task execution man- agement, efficiently overseeing tasks across cluster nodes in the Spark environment.
- **Storage Layer:**
  - MongoDB, a NoSQL database, is employed for storing large volumes of unstructured data.

The result of the envisioned framework is the evaluation of sentiment within the extensive data produced by social media networks. The framework is devised to extract various data labels and categorize them as favorable, unfavorable, or neutral emotional expressions. The sentiment analysis results can be used for various applications such as brand monitoring, customer feedback analysis, and public opinion analysis.

The proposed framework boasts scalability through Apache Kafka, ensuring rapid and downtime-free system expansion. It excels in high throughput, handling a large volume of messages with low latency, a crucial factor for social media platforms. The framework's fault tolerance leverages Hadoop HDFS for resilience, while Spark Streaming ensures swift recovery, load balancing, and resource optimization. Efficient resource management and availability are maintained through the YARN Cluster Manager and MongoDB.

Evaluating the effectiveness of a given framework [29] remains a significant challenge, primarily due to concerns related to the reliability and dependability of Big Data sourced from social media. Furthermore, the paper does not address the assessment of Big Data quality, a crucial element in any Big Data framework. The authors propose that future endeavors could integrate methodologies for assessing data quality to substantiate its quality. Furthermore, the proposed framework is evaluated using a limited dataset, and the authors do not provide a detailed analysis of the performance of the frame- work on a larger dataset. Finally, the proposed framework is not compared with other existing frameworks for sentiment analysis, which limits the ability to assess the effectiveness of the proposed framework compared to other frameworks.

Kafka serves as a valuable technical solution for creating a real-time streaming platform that facilitates seamless data transfer across various applications and systems. These stream- ing applications, known for their speed, scalability, and reliabil- ity, offer distinct advantages over traditional Message Queues. These benefits include expandability, maintaining message history, catering to various users, duplication, safeguarding message sequence, and utilizing the TCP protocol [23].

In the banking and financial application backend systems, there is a noticeable shift away from traditional batch pro-

cessing in favor of streaming platforms. This transition aims to achieve nearly real-time data processing, leading to substantial reductions in processing and settlement times within these critical sectors. Three case studies explore the strengths and weaknesses of streaming platforms, shedding light on their potential to enhance efficiency and real-time data processing. Kafka's architectural framework plays a pivotal role in the Banking, Financial Services, and Insurance (BFSI) domain, offering technical and business advantages. In the banking and financial domain, Kafka significantly reduces trade set- tlement times, moving from T+2 to T+1 or even T+0 days. Additionally, it enables the delivery of real-time alerts and notifications to customers, especially in response to suspicious

account activities.

The Kafka-based architecture delivers the key benefit of decoupling components within a system, establishing a clear division between event production and consumption. This ar- chitectural approach enhances the resilience and fault tolerance of event systems while allowing independent platforming, scalability, and efficient packaging. Kafka's KStream feature proves powerful for handling streams of key-value pairs.

Adopting Kafka-based architecture results in improved user interfaces, heightened security,and enhanced system perfor- mance. Kafka stands as a robust technical solution for building real-time streaming platforms that seamlessly transfer data between applications and systems.

Case Study 1: Prominent Singaporean banks integrated Kafka solutions to enhance fraud detection during mobile application logins to their banking portal. These applications were constructed using Kafka, Kafka Streams, and APIs, with Kafka brokers primarily handling event streaming rather than computational logic. Events were processed in a real-time, event-at-a-time model with minimal latency in milliseconds. The processing applications employed Kafka's consumer and producer APIs in conjunction with microservices to enhance user interface, security, and system efficiency. Case Study 2: This case study highlights Kafka's role in cloud-based trade and settlement systems in conjunction with existing enterprise service bus (ESB) architectures. Kafka's architecture facili- tated the decoupling of system components and established a logical separation between event production and consumption. This design bolstered the resilience and fault tolerance of event-driven systems, offering scalability and independent packaging. The benefits included a reduction in settlement time from Ti+2 to Ti+1/Ti+0 days, real-time customer alerts, and enhanced user interface, security, and system efficiency. Case Study 3: A Proof of Concept (POC) was carried out to integrate a legacy core banking system, which ran on a DB2 database, with a Kubernetes-based container engine. In the financial sector, we have a complex project at hand. It's all about creating a complete system architecture that covers everything from the core banking mainframe's DB2 database to modern containerized cloud solutions. This process happens in several steps. Notably, one of these steps involves moving data from Kafka topics to MQ, which is quite interesting. The advantages of this solution encompass improved user interface,

heightened security, and optimized system efficiency, along with the ability to seamlessly integrate legacy systems with modern containerized solutions. In conclusion, this review pa- per underscores the transmute potential of Kafka in enhancing the efficiency, security, and user experience within banking and financial systems. The showcased case studies vividly illustrate the myriad advantages of Kafka adoption in the BFSI sector, spanning fraud detection, cloud-based solutions, and trade and settlement systems. Throughout, the paper accentuates the compelling call for modernization in the BFSI domain, placing a spotlight on the indispensable role of enterprise application integration in realizing both technical and business benefits.

IV. CONCLUSION

The papers seem to focus on analyzing customer satisfac- tion and sentiment towards mobile banking applications and financial systems using social media data. The papers propose various frameworks and methodologies for collecting and an- alyzing data in real-time to improve customer satisfaction and loyalty. The proposed frameworks include components such as data ingestion, layered analytics, resource management, and storage. The papers suggest that analyzing customer sentiment and feedback can help banks and financial institutions identify areas of dissatisfaction and improve their services to drive profit growth. Overall, the papers highlight the importance of leveraging social media data to improve customer satisfaction and loyalty in the banking and financial industry.

V. ACKNOWLEDGEMENT

I extend my heartfelt gratitude to the Department Review Committee for their valuable insights and feedback, which greatly contributed to the quality and refinement of this research paper.I am grateful for the assistance of AI in rephrasing certain sentences for this research paper.

REFERENCES

[1] S. M. Alzahrani, "Big data analytics tools: Twitter api and spark," in 2021 International Conference of Women in Data Science at Taif University (WiDSTaif), pp. 1–6, 2021.

[2] T. A. Mat, A. Lajis, and H. Nasir, "Text data preparation in rapidminer for short free text answer in assisted assessment," in 2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), pp. 1–4, 2018.

[3] B. R. Hiraman, C. Viresh M., and K. Abhijeet C., "A study of apache kafka in big data stream processing," in 2018 International Conference on Information, Communication, Engineering and Technology (ICI- CET), pp. 1–3, 2018.

[4] A. W. Services, "Sentiment analysis - amazon web services," Accessed in 2023. Information on sentiment analysis, its applications, methods, challenges, and tools provided by Amazon Web Services.

[5] T. A. Mahmud, S. Sultana, T. I. Chowdhury, and F. R. Anando, "A new approach to analysis of public sentiment on padma bridge in bangla text," in 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), pp. 1–6, 2022.

[6] D. Zhang, W. Xu, Y. Zhu, and X. Zhang, "Can sentiment analysis help mimic decision-making process of loan granting? a novel credit risk evaluation approach using gmkl model," in 2015 48th Hawaii International Conference on System Sciences, pp. 949–958, 2015.

[7] G. J. Krishna, V. Ravi, B. V. Reddy, M. Zaheeruddin, H. Jaiswal, P. S. R. Teja, and R. Gavval, "Sentiment classification of indian banks' customer complaints," in TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), pp. 429–434, 2019.

[8] N. I. Prabaningtyas, I. Surjandari, and E. Laoh, "Mining customers opinion on services and applications of mobile payment companies in indonesia using sentiment analysis approach," in 2019 16th International Conference on Service Systems and Service Management (ICSSSM), pp. 1–5, 2019.

[9] M. A. Ligiarta and Y. Ruldeviyani, "Customer satisfaction analysis of mobile banking application based on twitter data," in 2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), pp. 322–327, 2022.

[10] L. Nair and S. Shetty, "Streaming big data analysis for real-time sentiment based targeted advertising," International Journal of Electrical and Computer Engineering, vol. 7, pp.

402–407, 02 2017.

[11] A. Marouani and A. Tick, "Predictive modeling to investigate and forecast customer behaviour in the banking sector," in 2023 IEEE 21st World Symposium on Applied Machine Intelligence and Informatics (SAMI), pp. 000255–000260, 2023.

[12] T. Kemboi and M. R. Islam, "Project: Lending club data analysis," 2019. Available online.

[13] D. Ujjwal, V. Uniyal, S. Pandey, S. V. Akram, V. Pachouri, and P. Negi, "Big data influence on the corporate social responsibility: Benefits and challenges," in 2023 International Conference on Computational Intelli- gence, Communication Technology and Networking (CICTN), pp. 121– 125, IEEE, 2023.

[14] N. Darapaneni, A. Kumar, A. Dixet, M. Suriyanarayanan, S. Srivastava, and A. R. Paduri, "Loan prediction software for financial institutions," in 2022 Interdisciplinary Research in Technology and Management (IRTM), IEEE, 2022.

[15] G. Arutjothi and C. Senthamarai, "Prediction of loan status in com- mercial bank using machine learning classifier," in 2017 International Conference on Intelligent Sustainable Systems (ICISS), IEEE, 2017.

[16] P. Mukherjee and Y. Badr, "Detection of defaulters in p2p lending platforms using unsupervised learning," in 2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS), IEEE, 2022.

[17] M. J. Hamayel, M. A. Abu Mohsen, and M. Moreb, "Improvement of personal loans granting methods in banks using machine learning methods and approaches in palestine," in 2021 International Conference on Information Technology (ICIT), IEEE, 2021.

[18] X. Sun, "Prediction of the borrowers' payback to the loan with lending club data," in 2020 International Conference on Modern Education and Information Management (ICMEIM), IEEE, 2020.

[19] A. Poddar, P. Kulkarni, and N. Natraj, "Application of big data for better decision management in banking," in 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), pp. 1404–1407, 2023.

[20] S. Yadav and S. Thakur, "Bank loan analysis using customer usage data: A big data approach using hadoop," in 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), IEEE, 2017.

[21] S. U. Maheswari and S. S. Dhenakaran, "Aspect based fuzzy logic sentiment analysis on social media big data," in 2020 International Con- ference on Communication and Signal Processing (ICCSP), pp. 0971– 0975, 2020.

[22] H. Huang, "Loan pricing of city commercial banks for small enterprises based on relationship-lending," in 2009 International Conference on Business Intelligence and Financial Engineering, IEEE, 2009.

[23] R. K. Mallidi, M. Sharma, and S. R. Vangala, "Streaming platform implementation in banking and financial systems," in 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), pp. 1–6, 2022.

[24] H. Wu, Z. Shang, and K. Wolter, "Performance prediction for the apache kafka messaging system," in 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Con- ference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 154– 161, 2019.

[25] N. N. Hlaing and T. T. Soe Nyunt, "Developing scalable and lightweight data stream ingestion framework for stream processing," in 2023 IEEE Conference on Computer Applications (ICCA), pp. 405–410, 2023.

[26] I. Sapthami, B. M. Krishna, T. Bhaskar, and C. Ravela, "Sentiment anal- ysis using machine learning algorithms for customer product reviews," in 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), pp. 447–451, 2023.

[27] T. R. Margocahyo, M. Saputra, and R. Y. Fa'rifah, "Unlocking user satisfaction: Evaluating stress levels of financial technology users in indonesia through sentiment analysis," in 2023 10th International Con- ference on ICT for Smart Society (ICISS), pp. 1–6, 2023.

[28] D. Boulegane, N. Radulovic, A. Bifet, G. Fievet, J. Sohn, Y. Nam, S. Yu, and D.-W. Choi, "Real-time machine learning competition on data streams at the ieee big data 2019," in 2019 IEEE International Conference on Big Data (Big Data), pp. 3493–3497, 2019.

[29] K. Fahd, S. Parvin, and A. de Souza-Daw, "A framework for real-time sentiment analysis of big data generated by social media platforms," in 2021 31st International Telecommunication Networks and Applications Conference (ITNAC), pp. 30–33, IEEE, 2021.