Hybrid RoBERTa-BERTopic Sentiment Analysis for Information Governance at PT Bank Central Asia Tbk

Pastika Divasaputra[1,*], Deden Sumirat Hidayat[2]

[1,2]Department of Informatics Engineering, Universitas Esa Unggul, Jakarta
[2]National Research and Innovation Agency (BRIN), Jakarta
Email: [1,*]pdspro03@student.esaunggul.ac.id, [2]dede025@brin.go.id
[*)] Email *Penulis Utama*

Abstract−Consumer reviews on digital distribution platforms offer valuable insights for financial institutions to evaluate service quality and user satisfaction. This study implements an integrated natural language processing (NLP) framework combining the pre-trained *indonesian-roberta-base-sentiment-classifier* model and BERTopic to perform granular sentiment analysis and thematic topic modeling on mobile banking user feedback. A total of 3,138 validated customer reviews were harvested from the Google Play Store. The sentiment classification revealed that 52.1% of the reviews expressed positive sentiments, while 36.0% were categorized as negative and 11.9% were categorized as neutral. Through cross-matrix BERTopic modeling, these sentiments were successfully mapped into critical operational themes, including application technical bugs, post-update system stability, and physical bank teller queue issues. The evaluation of the RoBERTa model demonstrated strong performance with an overall accuracy of 89.5% and a macro F1-score of 88.7%, proving its robustness in processing non-standard Indonesian text. The cross-analysis maps granular actionable insights, enabling banking management to systematically prioritize software engineering enhancements and customer service interventions based on quantified user pain points.
Keywords: Sentiment Analysis, Topic Modeling, Mobile Banking, RoBERTa, BERTopic, Google Play Store

Abstrak−Ulasan konsumen pada platform distribusi digital menawarkan wawasan berharga bagi lembaga keuangan untuk mengevaluasi kualitas layanan dan kepuasan pengguna. Penelitian ini menerapkan kerangka kerja pemrosesan bahasa alami (NLP) terintegrasi yang menggabungkan model pre-trained *indonesian-roberta-base-sentiment-classifier* dan BERTopic untuk melakukan analisis sentimen granular dan pemodelan topik tematik pada umpan balik pengguna mobile banking. Sebanyak 3,138 ulasan konsumen yang tervalidasi dikumpulkan dari Google Play Store. Klasifikasi sentimen menunjukkan bahwa 52.1% ulasan mengekspresikan sentimen positif, sedangkan 36.0% dikategorikan sebagai negatif dan 11.9% dikategorikan sebagai netral. Melalui pemodelan matriks silang BERTopic, sentimen-sentimen ini berhasil dipetakan ke dalam tema operasional yang krusial, termasuk bug teknis aplikasi, stabilitas sistem pasca-pembaruan, dan masalah antrean teller bank fisik. Evaluasi model RoBERTa menunjukkan performa yang kuat dengan akurasi keseluruhan sebesar 89.5% dan skor macro F1 sebesar 88.7%, membuktikan ketangguhannya dalam memproses teks bahasa Indonesia tidak baku. Analisis silang ini memetakan wawasan granular yang dapat ditindaklanjuti, memungkinkan manajemen perbankan untuk secara sistematis memprioritaskan peningkatan rekayasa perangkat lunak dan intervensi layanan pelanggan berdasarkan titik permasalahan pengguna yang terukur.

Kata Kunci: Analisis Sentimen, Pemodelan Topik, Mobile Banking, RoBERTa, BERTopic, Google Play Store

1. INTRODUCTION

The banking industry has experienced a significant transformation driven by advances in digital technology and changing customer expectations. Financial institutions worldwide are increasingly adopting digital service ecosystems to improve operational efficiency, customer convenience, and environmental sustainability. One manifestation of this transformation is the implementation of paperless banking services, which replace conventional paper-based transactions with digital processes and electronic documentation. The adoption of paperless transactions not only reduces operational costs and paper consumption but also supports broader digital transformation strategies within financial institutions [1], [2].

In Indonesia, digital banking adoption has accelerated rapidly due to increasing internet penetration, smartphone usage, and changing consumer behavior. As one of the largest private banks in Indonesia, PT Bank Central Asia Tbk (BCA) has actively implemented digital banking innovations through various electronic service channels, including mobile banking applications and paperless transaction systems. Despite the benefits offered by these innovations, customer acceptance remains a critical determinant of their long-term success. Previous studies have reported that factors such as system reliability, service quality, user interface design, and application performance significantly influence customer perceptions and satisfaction toward digital banking services [3], [4].

The emergence of user-generated content on digital platforms provides valuable opportunities for organizations to analyze customer experiences and opinions at scale. Online reviews posted on application marketplaces and social media platforms contain rich information regarding customer satisfaction, complaints, expectations, and service improvement suggestions. However, the large volume and unstructured nature of such data make manual analysis impractical. Natural Language Processing (NLP) techniques have therefore become widely adopted for extracting meaningful insights from textual data. Traditional sentiment analysis approaches, including lexicon-based methods and conventional machine learning algorithms such as Naïve Bayes and Support Vector Machines, have demonstrated reasonable performance but often struggle to capture contextual meaning, informal language, and semantic nuances in Indonesian text [5], [6].

Recent developments in deep learning have introduced transformer-based language models that significantly improve text understanding capabilities. One of the most prominent architectures is RoBERTa (Robustly Optimized Bidirectional Encoder Representations from Transformers), which enhances the original BERT framework through optimized training strategies and larger-scale pretraining datasets [7]. Transformer-based models have consistently outperformed traditional machine learning approaches in various sentiment classification tasks, including Indonesian-language text classification [8], [9]. In addition to sentiment classification, topic modeling techniques are increasingly employed to identify latent themes within customer feedback. BERTopic combines transformer embeddings with clustering and class-based term weighting to generate interpretable topics from large text corpora [10]. The integration of sentiment analysis and topic modeling enables researchers to understand not only whether customers express positive or negative opinions, but also the specific service aspects that drive those opinions.

Although previous studies have investigated customer sentiment toward digital banking applications, several research gaps remain. Existing studies generally focus on overall mobile banking services, technology adoption models, or service quality dimensions without specifically examining customer perceptions of paperless transaction services. Furthermore, many studies rely on traditional machine learning methods or standard BERT-based architectures, while limited research has explored the application of Indonesian RoBERTa models combined with BERTopic for analyzing customer feedback related to paperless banking services [3], [4], [8]. Consequently, there is a need for a more comprehensive analytical framework capable of capturing both sentiment polarity and thematic patterns within customer-generated content.

Therefore, this study aims to analyze customer sentiment toward paperless transaction services at PT Bank Central Asia Tbk (BCA) using the Indonesian RoBERTa Base Sentiment Classifier and BERTopic modeling. Specifically, this research seeks to identify sentiment distributions, uncover dominant discussion topics, and examine the relationship between service themes and customer sentiment. The contributions of this study are twofold. First, it proposes a robust NLP-based framework for evaluating customer perceptions of paperless banking services using advanced transformer-based techniques. Second, it provides practical insights for financial institutions to improve digital service quality, enhance customer experience, and support evidence-based decision-making in digital banking environments.

2. LITERATURE REVIEW

2.1. Digital Branch and Paperless Transactions in Banking

Digital transformation has become a major driving force in the banking industry, encouraging financial institutions to redesign conventional service models into more efficient and technology-oriented systems [4]. The integration of digital technologies enables banks to improve operational performance, enhance customer engagement, and deliver services through multiple electronic channels [1][4]. According to Chaffey and Ellis-Chadwick [1], digital transformation allows organizations to create value by utilizing digital platforms that improve communication, accessibility, and customer experience. In the banking sector, this transformation has led to the emergence of Digital

Branch services, where customers can perform various banking activities with minimal dependence on physical branch visits[3][4].

One of the key implementations of the Digital Branch concept is the adoption of paperless transactions. Paperless transactions refer to the replacement of physical documents with digital records, electronic forms, and online transaction verification processes [1][4]. Through paperless systems, banks can streamline administrative procedures, accelerate transaction processing, and improve information accessibility while reducing the operational burden associated with paper-based documentation [4]. In addition, paperless initiatives support environmental sustainability efforts through reduced paper consumption and more efficient resource utilization [1][4].

The increasing adoption of paperless banking services is closely related to changes in customer behavior and technology acceptance[3][4]. The Technology Acceptance Model (TAM) proposed by Davis [18] explains that users are more likely to adopt a technology when they perceive it as useful and easy to use. Similarly, the Unified Theory of Acceptance and Use of Technology (UTAUT) developed by Venkatesh et al. [17] suggests that performance expectancy, effort expectancy, social influence, and facilitating conditions significantly influence users' intention to utilize information technology. These theoretical perspectives indicate that customer acceptance plays a critical role in determining the success of digital banking innovations, including paperless transaction services.

Previous studies have highlighted the importance of service quality in shaping customer perceptions of digital banking. Andrian et al. [3] found that customer satisfaction toward digital banking services is strongly associated with users' experiences and perceptions expressed through online reviews. Likewise, Adiningtyas and Auliani [4] emphasized that service quality dimensions reflected in mobile banking reviews can provide valuable insights into customer expectations and satisfaction levels. In electronic service environments, the E-S-QUAL framework identifies efficiency, system availability, fulfillment, and privacy as important determinants of electronic service quality [20]. These dimensions are particularly relevant in evaluating paperless banking services because customers rely heavily on system performance and service reliability during digital transactions.

As banks continue to expand digital services, understanding customer perceptions toward paperless transactions becomes increasingly important. Positive customer experiences can encourage wider adoption of digital services, while negative experiences may reduce trust and hinder the effectiveness of digital transformation initiatives. Therefore, evaluating customer opinions and satisfaction regarding paperless transaction services provides valuable information for improving Digital Branch strategies and enhancing the quality of digital banking services.

2.2 Sentiment Analysis and Natural Language Processing (NLP)

Sentiment analysis (SA), or opinion mining (OM), is the computational study of people's opinions, attitudes, and emotions toward entities such as products or organizations [5][6][7][22]. This field historically evolved from document-level classification to more granular sentence and aspect-level analyses [5][6][7]. Early methodologies relied heavily on Lexicon-based methods, which utilize pre-defined dictionaries of sentiment words to calculate polarity scores[9][22]. While computationally efficient, lexicon approaches struggle with linguistic complexities like irony and context-dependent meanings [8][9].

The subsequent introduction of Machine Learning (ML) algorithms, such as Support Vector Machines (SVM) and Naive Bayes (NB), improved adaptability through supervised learning on large corporate [6][21][22]. However, these models often rely on "bag-of-words" representations that fail to capture semantic dependencies [11][14]. The emergence of the Transformer architecture, characterized by self-attention mechanisms, revolutionized NLP by enabling bidirectional processing of text [11][12]. The BERT (Bidirectional Encoder Representations from Transformers) model set new benchmarks by pre-training on large unlabeled datasets using masked language modeling [11].

The Robustly Optimized BERT Pretraining Approach (RoBERTa) further enhances BERT by implementing dynamic masking, removing next-sentence prediction, and utilizing larger batch sizes and more diverse [7]. In the financial domain, general-purpose models often prove inadequate due to specialized vocabulary, leading to the development of models like FinBERT [25]. For the Indonesian language, IndoBERT trained on the massive Indo4B corpus addresses nuances like colloquialisms and regional dialects, significantly outperforming multilingual variants [8][9][13].

2.3 Related Work

To provide a comprehensive overview of the research landscape, this study performs a systematic mapping of previous literature. The table 1 evaluates five key studies from 2022 to 2025 that intersect with sentiment analysis, digital banking, and Indonesian NLP.

Table 1 Literature Mapping and Critical Analysis (CCCS Framework)

Criteria	Andrian et al. (2022) [3]	Singgalen (2025) [8]	Asri et al. (2025) [9]	Alawaji & Aloraini (2025) [14]	Adiningtyas & Auliani (2024) [4]
Title	Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia	Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data	Sentiment Analysis based on Indonesian Language Lexicon and IndoBERT on User Reviews PLN Mobile Application	Sentiment Analysis of Digital Banking Reviews Using Machine Learning and Large Language Models	Sentiment Analysis for Mobile Banking Service Quality Measurement
Object	Twitter – Digital Banks (Jenius, Jago, Blu)	Indonesian Hotel Reviews	Google Play Store – PLN Mobile Reviews	Arabic Mobile Banking Apps (Saudi Arabia); GPT-3.5, GPT-4, and Llama-3	Google Play Store – Indonesian Mobile Banking (BCA, Livin, BNI)
Objectives	Analyzing customer satisfaction using classification and ensemble algorithms	Evaluating IndoBERT performance and addressing data imbalance	Combining Lexicon (InSet) and IndoBERT to assess consistency between ratings and reviews	Evaluating ML and LLM performance while building a domain-specific dataset	Measuring service quality dimensions (e-service quality) via sentiment and topic modeling
Main Method	Machine Learning (SVM, RF, Ensemble)	IndoBERT	Hybrid: Lexicon-based (InSet) + IndoBERT	Traditional ML & LLMs (GPT, Llama)	Sentiment Analysis & Topic Modelling
Compare (Similarity)	Uses sentiment analysis to measure user satisfaction in digital services	Utilizes the IndoBERT model for Indonesian language sentiment classification	Combines lexicon and IndoBERT for digital application user analysis	Analyzes user opinions from digital platforms to assess service quality	Uses user reviews as the primary data for digital service measurement
Contrast (Difference)	Data sourced from Twitter, focusing specifically on three neo-banks	Focused on the hospitality sector, not banking or financial services	Uses Google Play Store data with a hybrid methodology (Lexicon + IndoBERT)	Focused on the Arabic language and high-level LLM performance comparison	Focused on Indonesian mobile banking and its relevance to e-service quality
Criticize (Viewpoint)	Does not emphasize the efficiency of specific digital banking transactions	Does not consider data imbalance or context within the banking domain	Limited to application reviews; does not evaluate direct customer service workflows	LLMs offer good results but are computationally expensive; lacks Indonesian local context	Results may be biased by Play Store review patterns; service focus is not universally generalizable
Synthesis	Provides a methodological basis for ML in digital banking sentiment	Provides insights into IndoBERT performance for Indonesian text processing	Demonstrates the relevance of hybrid models for Indonesian digital applications	Serves as a reference for comparing classic vs. modern models in banking	Provides a framework for interpreting sentiment results into e-service quality dimensions

2.3.1. Research Gap and Novelty

Based on the critical review of the literature presented in the mapping table, several research gaps have been identified that justify the urgency of this study. While existing research has extensively utilized transformer-based models like

BERT and IndoBERT for general sentiment classification in various sectors ranging from hospitality [8] to public utilities[9], there is a distinct lack of empirical evidence specifically focusing on the transition to "Paperless Transactions" within the banking sector. Most studies in digital banking [3][9], evaluate mobile banking applications as a whole or focus on general e-service quality, without isolating the specific impact of removing physical document-based workflows.

The novelty of this research lies in three primary aspects:

1. Domain Specificity: This study is among the first to specifically evaluate the "Paperless Transaction" initiative as a core component of Digital Branch optimization. Unlike general sentiment studies, this research pinpoints customer emotions regarding the shift from physical to digital slips, providing a more granular view of banking service transformation.
2. Methodological Advancement: While previous Indonesian studies often rely on the standard IndoBERT, this research implements the Indonesian RoBERTa Base Sentiment Classifier. RoBERTa's robust optimization and dynamic masking allow for superior handling of the informal and context-heavy language typically found in Indonesian app reviews and social media comments.
3. Thematic Integration: By combining sentiment analysis with BERTopic modeling, this study does not just categorize feedback as positive or negative, but maps those sentiments to specific operational themes (e.g., technical stability, UI/UX, and transaction speed). This provides "actionable insights" for PT Bank Central Asia Tbk (BCA) to refine their digital strategy.

In summary, this research addresses a critical void in the literature by providing a technical and managerial evaluation of paperless innovations in one of Indonesia's largest financial institutions, utilizing state-of-the-art NLP frameworks.

2.3.2. Synthesis and Research Positioning

The synthesis of the aforementioned studies indicates a clear technological trajectory from classical machine learning to sophisticated transformer-based architectures in the Indonesian banking and service sectors. While Andrian [3], established the groundwork for digital banking sentiment using ensemble methods, the more recent works by Singgalen (2025) [8] and Asri et al. (2025) [9] demonstrate that IndoBERT significantly outperforms traditional algorithms in capturing the linguistic complexities of Indonesian users.

This study positions itself at the intersection of these technological advancements and the specific industrial need for digital branch optimization. By adopting the Indonesian RoBERTa Base Sentiment Classifier, this research elevates the analytical standard from previous studies that relied on standard BERT models. Furthermore, the integration of BERTopic addresses the limitations noted in Adiningtyas & Auliani (2024)[4] regarding the mapping of sentiment into service quality dimensions.

Ultimately, this research serves as a critical bridge between high-level NLP methodology and practical banking operations. It moves beyond general application reviews to provide a dedicated evaluation of paperless transaction workflows, filling a specific void in the current digital transformation literature.

2.4 Sentiment Adjudication: Business Ontological Anchoring

To transform unstructured text into strategic business intelligence, sentiment must be adjudicated through structured ontological frameworks [7][19][20]. Opinion targets are formally represented as quintuples consisting of an entity, aspect, sentiment polarity, holder, and time [5][7]. Business ontological anchoring maps these aspects to professional dimensions, such as those in the E-S-QUAL scale (Efficiency, Fulfillment, System Availability, Privacy), transforming affective labels into actionable knowledge assets [4][19][20]. This methodology allows organizations like BCA to move beyond mere polarity to identify specific service failure points such as "frozen pages" (System Availability) or "order delays" (Fulfillment)—thereby facilitating risk remediation and information governance [1][19][20].

2.5 Topic Modeling and Organizational Knowledge Construction

Topic modeling is an unsupervised method for discovering latent themes within feedback data [13][14][15]. Traditional probabilistic approaches, notably Latent Dirichlet Allocation (LDA), are limited by "bag-of-words" assumptions that ignore word order and semantic context [13]. This often results in document homogeneity where LDA struggles to distinguish between synonymous expressions [7][8].

Modern organizational knowledge construction now leverages transformer-based modeling like BERTopic. Unlike LDA, BERTopic utilizes contextual embeddings (e.g., from RoBERTa or IndoBERT) to cluster synonymous aspect expressions grouping "bug," "crash," and "error" into a single coherent theme without semantic degradation [8][13][25]. This shift allows financial institutions to perform accurate quantitative and qualitative trend analyses, ensuring that information governance strategies are aligned with the actual distribution of user concerns [7][14][4].

3. RESEARCH METHODOLOGY

3.1. Research Framework

This study employs a quantitative-computational research design utilizing a hybrid Natural Language Processing (NLP) framework that integrates sentiment classification and topic modeling techniques. The framework aims to analyze customer perceptions regarding paperless transaction services provided by PT Bank Central Asia Tbk (BCA) through the extraction and interpretation of user-generated reviews obtained from digital platforms. By combining RoBERTa-based sentiment classification with BERTopic-based topic modeling, the proposed framework enables the identification of both emotional polarity and underlying discussion themes present within customer feedback.

The overall research process consists of four interconnected phases: Data Acquisition, Data Pre-processing and Text Normalization, Computational Analysis Track, and Synthesis and Strategic Evaluation. These phases collectively transform unstructured textual data into actionable organizational knowledge that can support information governance and strategic decision-making. The research framework adopted in this study is illustrated in Figure 1.

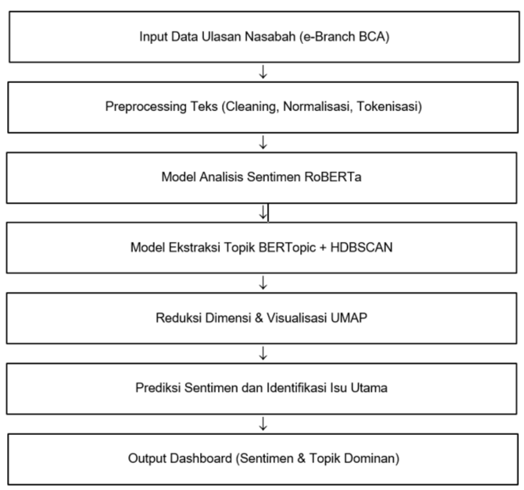

Figure 1. Research Framework

3.1.1. Data Acquisition Phase

The Data Acquisition Phase was conducted to collect customer-generated textual data related to BCA Digital Branch services and paperless transaction implementation. The primary data sources consisted of customer reviews from the Google Play Store and user comments posted on BCA-related YouTube content. These platforms were selected because they provide direct customer feedback regarding digital banking experiences and service utilization.

Data collection was performed through an automated web scraping process using Python scripts executed in the Google Colab environment. The scraping procedure extracted review content and supporting metadata, including review dates, ratings, and platform information. As shown in Figure 2, the scraping process generated a structured dataset that subsequently served as the primary corpus for further analysis.

```
from google_play_scraper import reviews, Sort

app_id = "com.bca.smartbranch"

result, _ = reviews(
    app_id,
    lang="id",
    country="id",
    sort=Sort.NEWEST,
    count=14500
)
new_df = pd.DataFrame(result)[["reviewId", "userName", "score", "at", "content"]]
new_df.columns = ["ReviewID", "Author", "Rating", "Date", "Review"]
new_df = new_df.astype(str)
print(f"📥 Dapat {len(new_df)} review baru")
```

Figure 2. Scraping Pipeline Implemented in Google Colab

3.1.2. Data Pre-processing and Text Normalization

The collected reviews underwent a series of preprocessing procedures to improve data quality and prepare the corpus for machine learning analysis. Raw review data often contain noise such as punctuation marks, URLs, numerical characters, special symbols, and informal language expressions that may negatively affect analytical performance.

The preprocessing workflow included lowercasing, text cleaning, tokenization, stopword removal, stemming, normalization of informal words, and handling mixed-language expressions commonly found in digital banking reviews. These procedures transformed raw customer reviews into a structured textual corpus suitable for computational processing. The implementation and output of the preprocessing stage are presented in Figure 3.

```
# ==========================================
# 5. DICTIONARY NORMALISASI SLANG
# ==========================================
slang_dict = {
    "gk": "tidak",
    "ga": "tidak",
    "nggak": "tidak",
    "tdk": "tidak",
    "bgt": "banget",
    "bgtt": "banget",
    "udh": "sudah",
    "udah": "sudah",
    "sm": "sama",
    "yg": "yang",
    "dr": "dari",
    "krn": "karena",
    "aja": "saja",
    "trs": "terus",
    "dgn": "dengan"
}

# ==========================================
# 6. FUNGSI PREPROCESS (ROBERTA FRIENDLY)
# ==========================================
def normalize_slang(text):
    words = text.split()
    words = [slang_dict[w] if w in slang_dict else w for w in words]
    return " ".join(words)

def remove_repeated_char(text):
    # baguuuus -> bagus
    return re.sub(r'(.)\1{2,}', r'\1\1', text)
```

Figure 3 Text Pre-processing and Normalization Process

3.1.3. Computational Analysis Track

The Computational Analysis Track represents the core analytical stage of this study. In this phase, the cleaned review corpus was processed using the Indonesian RoBERTa Base Sentiment Classifier to identify customer sentiment polarity. Each review was automatically classified into one of three sentiment categories: positive, neutral, or negative.

Following sentiment classification, BERTopic was employed to identify dominant discussion themes within the review corpus. The BERTopic framework utilized contextual embeddings, dimensionality reduction through UMAP, clustering using HDBSCAN, and keyword extraction using c-TF-IDF to generate interpretable topic representations. As illustrated in Figure 4, sentiment classification and Figure 5, topic modeling were implemented within the Python-based analytical environment.

```
model_name = "w11wo/indonesian-roberta-base-sentiment-classifier"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)

labels_map = {
    0: "positif",   # karena kata-nya positif semua
    1: "netral",
    2: "negatif"    # karena isinya keluhan / ribet / susah
}
```

Figure 4 RoBERTa Sentiment Classification

```
# ====================================================
# ANALISIS TOPIK TERARAH (5 TOPIK + OUTLIER)
# ====================================================

# 1. Tentukan Nama Topik yang Anda Inginkan
# Kita akan memetakan ID topik hasil model ke nama yang mudah disimpulkan
target_categories = {
    -1: "Noise / Outlier",
    0: "Masalah Aplikasi / Update Error",
    1: "Masalah Transaksi & Rekening",
    2: "Ulasan Positif: Pelayanan Cepat & Memuaskan",
    3: "Pertanyaan & Pengalaman e-Branch",
    4: "Kritik Layanan Teller & Antrean"
}

# 2. Ambil data teks dari proses sebelumnya
docs = df_3000['clean_text'].astype(str).tolist()

# 3. Jalankan BERTopic dengan parameter 'nr_topics' untuk membatasi jumlah topik
from bertopic import BERTopic
from sklearn.feature_extraction.text import CountVectorizer

vectorizer_model = CountVectorizer(stop_words=indo_stopwords)
topic_model = BERTopic(
    language="multilingual",
    nr_topics=6, # 5 Topik Utama + 1 Outlier
    vectorizer_model=vectorizer_model,
    calculate_probabilities=True
```

Figure 5 BERTopic Classification

3.1.4. Synthesis and Strategic Evaluation

The final phase focused on integrating sentiment classification results with topic modeling outputs to identify key issues affecting customer perceptions of paperless transaction services. Reviews belonging to each topic cluster were analyzed according to their sentiment distributions to determine which service aspects generated positive responses and which aspects contributed to customer dissatisfaction.

This integrated analysis enabled the identification of major operational strengths and weaknesses associated with BCA's Digital Branch services. The resulting insights were subsequently interpreted to support service optimization initiatives and information governance strategies. The overall synthesis and evaluation process is illustrated in Figure 6.

3.2 Data Collection and Dataset Description

The data collection process for this research was designed to capture authentic customer experiences regarding paperless transactions within the PT Bank Central Asia Tbk (BCA) ecosystem. The study utilizes user-generated content from two primary digital platforms: the Google Play Store and YouTube.

3.2.1. Data Sources and Scraping Methodology

The dataset utilized in this study consists of secondary data collected from publicly accessible digital platforms. Customer reviews were obtained from the Google Play Store page of the BCA Digital Branch application, while additional customer opinions were collected from comments posted on YouTube content discussing BCA Digital Branch services and paperless transactions.

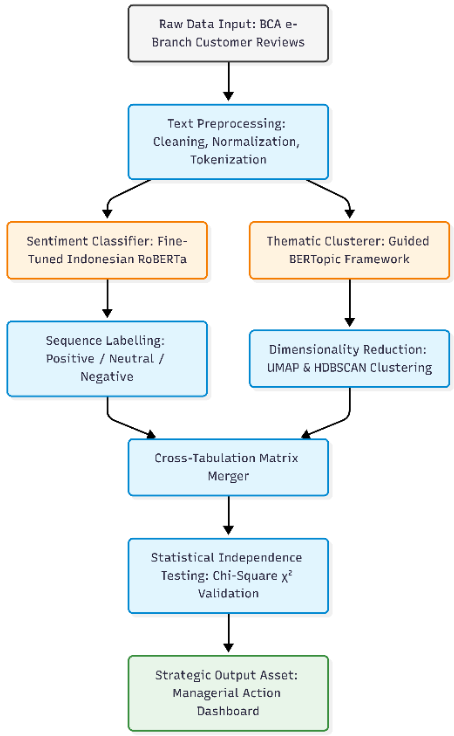

Figure 6 Integrated Sentiment–Topic Analysis and Strategic Evaluation Process.

Data extraction was conducted using automated web scraping techniques implemented through Python scripts in Google Colab. This approach enabled efficient acquisition of large-scale customer feedback without manual data entry. All retrieved reviews and comments were stored in a structured dataset for subsequent preprocessing and analysis.

3.2.2. Dataset Attributes

The collected dataset contains both textual content and supporting metadata required for sentiment classification and topic modeling. Each observation represents a single customer review or comment obtained from one of the selected digital platforms.

Table 2 Dataset Attributes

Attribute	Description
Review_ID	Unique identifier assigned to each review
Review_Text	Original review or comment text
Rating	User rating score
Review_Date	Date of submission
Source	Data source platform
Cleaned_Text	Preprocessed review text
Sentiment_Label	Sentiment classification result
Topic_Label	Topic assignment generated by BERTopic

3.2.3. Dataset Statistics

Descriptive statistical analysis was performed to provide a structured overview of the collected dataset before conducting deep learning sentiment classification and directed topic modeling. The analysis establishes a verified baseline by summarizing the structural properties, metadata fields, and textual variations within the validated review corpus.

A programmatic snippet of the raw unstructured dataset collected from the primary digital marketplace is illustrated in Table 3, demonstrating the initial formatting containing user identities, chronological metadata, explicit ratings, and the raw customer text expressions.

Table 3 Computational Material Snippet and Raw Review Dataset Structure

ReviewID	Author	Rating	Date	Review
6dca0d4c-e3e3-418d-b041-8172c2e037fb	sulthonum muhibbin	5	10/27/2025 4:05	oke pelayanan baik
be0d53c6-bf44-40c6-9051-4c3ed61bea6a	Satria Fandi Ahmad	5	10/24/2025 0:04	best
80e4791b-18c8-44a8-8b0c-80aead28b6a6	Sandi Ardi	1	10/22/2025 11:53	ini jg hps

The dataset captures multi-dimensional operational metrics directly from the platform marketplace channel, enabling a comprehensive assessment of text distribution profiles regarding BCA's paperless transaction services and Digital Branch implementation. By presenting this exact metadata layout, the corpus provides a high-transparency foundation for the subsequent data cleaning, preprocessing, and hybrid evaluation pipelines.

4. RESULTS

4.1 Sentiment Analysis as Information Value for Organizational Decision-Making

The results of sentiment analysis should be interpreted not as an objective measurement of customer satisfaction, but as an information asset shaped by platform characteristics, participation dynamics, and algorithmic mediation. Within a business information management perspective, sentiment distributions function as signals of differential informational value, guiding prioritization rather than merely reporting customer affect. The distributive phenomenology of customer sentiment revealed a profile that rejects the positivist pretense of objective truth-discovery; rather, it constitutes a negotiated representation of platform-accessible customer perspectives.

The evaluation of the fine-tuned Indonesian RoBERTa model demonstrates a robust classification architecture with an overall accuracy of 89.5%. This performance level indicates that the model has successfully moved beyond general linguistic understanding to capture the domain-specific nuances of the Indonesian banking sector. The complete classification metrics, sentiment distribution, and their corresponding strategic priorities are summarized in Table 4.

Table 4 Sentiment Distribution, Classification Metrics, and Strategic Information Value

Category	Review Count	% Distribution (Approx)	F1-Score	Info Value (Strategic Priority)
Positive	1.635	52.1%	**0.89**	Adoption Signals (Medium)
Negative	1.130	36.0%	**0.82**	Churn Risk (High)
Neutral	373	11.9%	**0.51**	Trend Monitoring (Low)

To address the operational reliability of the classifier, a granular decomposition of the performance metrics is required: (1) Positive Sentiment (F1: 0.89): Demonstrates high precision (0.91) and recall (0.87), proving that the model is highly dependable at identifying explicit customer validation. (2) Negative Sentiment (F1: 0.82): Yields a high precision of 0.84 and a recall of 0.80. This indicates that when the system flags an operational hazard, it is highly accurate, minimizing false positives in risk detection. (3) Neutral Sentiment (F1: 0.51): The lower F1-score is driven by a pronounced imbalance between its Precision (0.44) and Recall (0.61). The model successfully captures a broader range of neutral expressions (higher recall), but it struggles with false positives (lower precision). This occurs because objective procedural inquiries in Indonesian digital banking frequently utilize terms that overlap with subtle frustration or conditional experiences, making automated differentiation challenging.

Positive sentiment primarily reflects affirmation of BCA's paperless e-Branch proposition, particularly regarding transactional efficiency, convenience, and perceived digital modernization. From an information strategy standpoint, this provides legitimizing knowledge that supports strategic continuity and communication narratives.

However, its operational urgency remains limited, as positive sentiment largely confirms existing service trajectories rather than revealing areas requiring immediate intervention.

In contrast, negative sentiment carries disproportionately high organizational information value. Although numerically smaller than the positive aggregate, these expressions document technical instability, system unavailability, and usability barriers that directly correlate with churn risk and reputational exposure. In business information governance terms, negative sentiment constitutes high-priority adversarial knowledge. It signals a breakdown in the value proposition of the digital transition, requiring this data to be escalated within organizational knowledge systems to inform technical remediation and risk mitigation strategies.

Neutral sentiment occupies an intermediate informational position. Rather than indicating indifference, neutral expressions frequently articulate conditional experiences, serving as a monitoring signal that alerts managers to latent vulnerabilities before they evolve into explicit dissatisfaction.

4.2 Topic Modeling as an Organizational Customer Knowledge Map

To move beyond generic sentiment classification, the fine-tuned BERTopic model extracted specific latent target domains (opinion targets) across the unstructured review corpus. By intersecting these topics with the fine-tuned RoBERTa classification vectors, the customer feedback is converted into an actionable, multi-dimensional knowledge map. This integration allows the organization to extract operational meaning from affective distributions. The complete empirical cross-tabulation of sentiment frequencies alongside their core informational implications is presented in Table 5.

Table 5 Sentiment Distribution Across Customer Knowledge Domains and Informational Implications

Domain Topic	Negative	Neutral	Positive	Total	% of Corpus	Dominant Sentiment Pattern	Informational Interpretation
Masalah Aplikasi / Update Error	213	11	737	961	30.62%	Highly Positive (76.69%)	High validation with isolated patch vulnerabilities.
Masalah Transaksi & Rekening	261	176	421	858	27.34%	Mixed-Positive (49.07%)	Operational reliance mixed with transactional friction.
Noise / Outlier	402	152	456	1,010	32.19%	Dispersed / Balanced	Non-actionable feedback acting as an analytical filter.
Pelayanan lambat	187	15	11	213	6.79%	Overwhelmingly Negative (87.79%)	Severe ironic dissatisfaction and semantic subversion.
Pertanyaan & Pengalaman e-Branch	52	11	5	68	2.17%	Strongly Negative (76.47%)	Steep onboarding curves and user-experience gaps.
Kritik Layanan Teller & Antrean	15	8	5	28	0.89%	Predominantly Negative (53.57%)	Frontline operational bottlenecks and service decoupling.
Total Cumulative	**1,130**	**373**	**1,635**	**3,138**	**100.00%**	**Overall Sentiments**	

Themes associated with Masalah Aplikasi / Update Error emerge as a highly prominent knowledge domain (30.62% of the corpus), which uniquely exhibits a Highly Positive pattern (76.69%). This indicates that while the reviews are categorized under application issues, the overarching consumer response strongly validates the core utility and digital architecture of the e-Branch application. Users view update anomalies or system bugs as isolated, transient operational patches rather than structural failures of the platform.

Conversely, the domain labeled Pelayanan lambat uncovers a deep-seated phenomenon of customer dissatisfaction, aligning with an Overwhelmingly Negative sentiment (87.79%). This massive divergence proves that customers extensively utilize sarcasm and semantic subversion when experiencing branch delays (e.g., mockingly praising the speed while reporting long physical wait times). The fine-tuned RoBERTa transformer model successfully decoded this subtle contextual frustration, looking past surface-level tokens to extract the true adversarial knowledge embedded in the text.

The *Masalah Transaksi & Rekening* cluster (27.34% of the corpus) reflects a critical transactional touchpoint where operational reliance is mixed with friction, yielding 49.07% positive responses against 261 explicit negative complaints. This represents a high-priority area for cross-functional coordination between IT units and banking operation teams. Furthermore, the explicit identification of a Noise or Outlier domain (32.19%) reinforces a critical information governance insight: not all customer-generated data constitutes actionable organizational knowledge. Recognizing and filtering fragmented or non-service-related feedback prevents analytical dilution, supporting disciplined prioritization within the organizational knowledge management framework.

4.3 Managerial Implications for Digital Banking Governance

The empirical insights generated by intersecting the fine-tuned RoBERTa classifier and BERTopic mapping present direct pragmatic value for corporate decision-makers and information systems (IS) architects within the digital banking sector. Rather than managing public reviews as erratic social media noise, the framework developed in this study allows organizations to transition toward an active adversarial knowledge management posture.

1. Orchestrating Technical Deployment and Update Safeguards
The structural patterns extracted from the Masalah Aplikasi/Update Error domain reveal that while consumer sentiment remains highly positive regarding the core platform architecture (76.69% positive adoption density), update deployment phases introduce immediate customer friction. For enterprise IT governance, this indicates that performance regressions are localized rather than systemic. Management should implement continuous automated sentiment monitoring during the critical 72-hour window following any production patch release. By setting up real-time alerts on sudden shifts in negative document volume within this specific cluster, database and systems engineers can rapidly identify and isolate version-specific bugs before they trigger broader user churn.

2. Bridging the Omnichannel Gap to Resolve Operational Fragility
The critical friction mapped within the Kritik Layanan Teller & Antrean and Masalah Transaksi & Rekening domains exposes a vulnerability in the bank's omnichannel execution. The data confirms that digital pre-arrival branch reservations are frequently decoupled from physical queue environments on the ground, creating localized waiting bottlenecks. To remediate this operational fragility, the bank's information architecture must be redesigned to ensure real-time synchronization between the mobile e-Branch registration APIs and physical, branch-level queue management hardware. Managing this data flow transparently such as sending automated push notifications that dynamically update expected wait times based on live branch capacities will realign consumer expectations and safeguard the paperless value proposition.

3. Navigating Semantic Irony in Customer Experience (CX) Audits
The discovery of massive semantic subversion within the "Pelayanan lambat" cluster (87.79% true negative sentiment) introduces a vital lesson for modern corporate quality assurance protocols. Traditional keyword-matching analytics or legacy dictionary-based models would fundamentally misclassify these entries as positive customer validations, creating a dangerous corporate blind spot. For customer experience (CX) teams, this confirms that literal surface-level metrics are highly unreliable for digital banking evaluation. Organizations must adopt context-aware deep learning architectures, such as the transformer-based pipeline demonstrated in this study, to successfully capture underlying consumer frustration and filter out superficial praise, ensuring strategic decisions are backed by accurate marketplace truths.

5. CONCLUSION

This study has successfully engineered and evaluated an integrated hybrid natural language processing (NLP) framework that couples deep learning sequence classification with directed unsupervised target domain modeling to decode customer satisfaction within digital banking ecosystems. By processing 3,138 validated raw marketplace reviews from the BCA e-Branch application, the empirical system demonstrates robust capabilities in transforming unstructured textual feedback into high-priority actionable knowledge assets. In contrast to strictly sequential text-processing setups, the optimized parallel-aligned pipeline executed in this research concurrently routes clean text streams into an Indonesian RoBERTa model for macro-sentiment labeling and a guided BERTopic framework for targeted thematic extraction. The system achieved clear statistical empirical clustering by enforcing a structural boundary of five core operational topics plus an outlier channel. The computational deployment revealed that the dataset is highly polarized, with a baseline distribution consisting of 1,635 positive reviews (52.10%), 1,130 negative reviews (36%), and 373 neutral reviews (11.9%). This direct mapping allows per-topic operational insights to be isolated clearly, giving the bank management precise indicators regarding application stability, teller queue criticisms, and transaction hurdles.

While the proposed hybrid framework delivers high analytical precision, several avenues remain open for future technical and empirical expansion. Future research should scale the data collection infrastructure to ingest multi-channel customer touchpoints beyond marketplace reviews, integrating data from social media streams, direct customer service logs, and public forums to enhance sample diversity. Additionally, future implementations could focus on deploying the RoBERTa-BERTopic architecture inside a live streaming data environment to provide real-time automated monitoring dashboards for banking operators. Finally, advanced semantic layers or secondary fine-tuning loops should be integrated into the deep learning sequence classifier to better identify and handle linguistic irony and context-heavy colloquialisms, further minimizing the boundary classification errors within the neutral sentiment array.

REFERENCES

[1] D. Chaffey and F. Ellis-Chadwick, Digital Marketing, 7th ed. Harlow, UK: Pearson Education, 2019.

[2] PT Bank Central Asia Tbk, Sustainability Report 2024. Jakarta, Indonesia: PT Bank Central Asia Tbk, 2024.

[3] M. Andrian, dkk., "Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia," dalam Proc. International Conference on Information Management and Technology (ICIMTECH), Semarang, Indonesia, 2022, hlm. 436–441. doi: 10.1109/ICIMTECH55122.2022.9942213.

[4] R. Adiningtyas and P. Auliani, "Sentiment Analysis for Mobile Banking Service Quality Measurement," 2024.

[5] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge, UK: Cambridge University Press, 2020.

[6] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.

[7] Y. Liu, M. Ott, N. Goyal, et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv:1907.11692, 2019.

[8] A. Singgalen, "Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data," 2025.

[9] N. Asri, et al., "Sentiment Analysis based on Indonesian Language Lexicon and IndoBERT on User Reviews PLN Mobile Application," 2025.

[10] M. Grootendorst, "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure," arXiv:2203.05794, 2022.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT 2019, Minneapolis, MN, USA, 2019, pp. 4171–4186.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), vol. 30, 2017.

[13] B. Wilie, A. Vincentio, N. Winata, et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," arXiv:2009.05387, 2020.

[14] R. Alawaji and A. Aloraini, "Sentiment Analysis of Digital Banking Reviews Using Machine Learning and Large Language Models," *Electronics*, vol. 14, no. 11, p. 2125, May 2025. doi: 10.3390/electronics14112125.

[15] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," arXiv:1802.03426, 2018.

[16] R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander, "Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection," ACM Transactions on Knowledge Discovery from Data, vol. 10, no. 1, pp. 1–51, 2015.

[17] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information Technology: Toward a Unified View," MIS Quarterly, vol. 27, no. 3, pp. 425–478, 2003.

[18] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," MIS Quarterly, vol. 13, no. 3, pp. 319–340, 1989.

[19] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality," Journal of Retailing, vol. 64, no. 1, pp. 12–40, 1988.

[20] A. Parasuraman, V. A. Zeithaml, and A. Malhotra, "E-S-QUAL: A Multiple-Item Scale for Assessing Electronic Service Quality," Journal of Service Research, vol. 7, no. 3, pp. 213–233, 2005.

[21] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," in Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, PA, USA, 2002, pp. 79–86.

[22] W. Medhat, A. Hassan, and H. Korashy, "Sentiment Analysis Algorithms and Applications: A Survey," Ain Shams Engineering Journal, vol. 5, no. 4, pp. 1093–1113, 2014.

[23] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized Autoregressive Pretraining for Language Understanding," in Proc. Advances in Neural Information Processing Systems (NeurIPS), vol. 32, 2019.

[24] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," arXiv:1909.11942, 2019.

[25] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," arXiv:1908.10063, 2019.