International Journal of Accounting and Economics Studies, 12 (5) (2025) 196-208

International Journal of Accounting and Economics Studies

Website: www.sciencepubco.com/index.php/IJAES
https://doi.org/10.14419/xcd48k61
Research paper

Text Mining and Sentiment Analysis for Mobile Banking Service Quality Measurement: A Cross-Sectional Study of Turkish Private Banks

Ahmet Alkan Çelik, Yavuz Selim Balcıoğlu, Erkut Altındağ *

Doğuş University
**Corresponding author E-mail: erkutaltindag@dogus.edu.tr*

Received: June 15, 2025, Accepted: July 11, 2025, Published: September 6, 2025

Abstract

This study proposes a methodological framework for measuring mobile banking service quality through text mining of customer-generated content. The research employs sentiment analysis techniques to evaluate service quality dimensions across three prominent Turkish private banks based on 9,547 Google Play Store reviews. Following a systematic preprocessing protocol, reviews underwent sentiment classification using a Naive Bayes algorithm, achieving 87.3% accuracy. The analysis revealed distinct patterns across five service quality dimensions: Practicality emerged as the strongest dimension (mean score: 0.75), while Sociality demonstrated the most significant deficiency (mean score: 0.32). Statistical comparison identified significant inter-bank differences, with İşCep demonstrating superior overall performance (0.67), followed by Garanti BBVA (0.65) and TEB (0.58). Network analysis of keyword co-occurrences illuminated the semantic structure of customer discourse, revealing distinctive terminological communities within dimensional frameworks. The methodology transcends traditional survey-based approaches by providing continuous, scalable quality assessment derived from authentic customer expressions. Moreover, the integration of dimensional analysis with complaint pattern identification establishes clear priorities for service enhancement initiatives. This research advances both the theoretical understanding of mobile banking service quality dimensions and the practical implementation of computational text analysis in service quality measurement frameworks.

Keywords: *Mobile Banking, Service Quality Measurement, Text Mining, Sentiment Analysis, Dimensional Analysis*

1. Introduction

In the digital age, mobile banking services have become an essential component of modern financial ecosystems, offering users unprecedented convenience, accessibility, and efficiency. With the increasing reliance on mobile applications for banking transactions, service quality has emerged as a critical determinant of customer satisfaction and competitive differentiation among financial institutions. However, traditional methods for measuring service quality, such as structured surveys and interviews, often fail to capture the dynamic and evolving nature of customer experiences. As a result, there is a growing interest in leveraging computational techniques, particularly text mining and sentiment analysis, to extract insights from unstructured customer feedback.

Recent studies highlight the effectiveness of sentiment analysis in capturing customer perceptions and emotions in service industries. Cambria et al. (2017) emphasize that sentiment analysis enables a deeper understanding of customer attitudes by analyzing linguistic patterns in textual data, making it a valuable tool for service quality assessment. Furthermore, Arcand et al. (2017) argue that digital banking service quality is multidimensional, with usability, security, and responsiveness playing key roles in shaping customer satisfaction. In addition, Liu (2012) underscores the advantages of text mining techniques in processing large volumes of customer feedback efficiently, allowing financial institutions to identify service deficiencies and areas for improvement in real-time. By integrating these perspectives, this research contributes to the literature by demonstrating how sentiment analysis and text mining can enhance mobile banking service evaluation through scalable, data-driven methodologies.

This study explores the application of text mining and sentiment analysis to evaluate mobile banking service quality in the Turkish private banking sector. By analyzing customer-generated content from mobile banking applications, this research aims to uncover key service quality dimensions and assess how customers perceive their banking experiences. The study builds upon existing theoretical frameworks and methodological approaches to bridge the gap between computational analytics and service quality evaluation, providing both empirical insights and managerial implications for digital banking optimization. We address three critical research questions regarding service quality measurement in the Turkish private banking sector. These research questions are designed to systematically investigate the application of text mining approaches to mobile banking service quality assessment and to provide practical insights for both theoretical advancement and managerial implementation.

Research Question 1: Dimensional Quality Assessment

RQ1: How do Turkish private banks compare across the five service quality dimensions (Security/Privacy, Practicality, Design/Aesthetics, Sociality, and Enjoyment) when evaluated through sentiment analysis of mobile application reviews?
Research Question 2: Methodological Validation
RQ2: To what extent does sentiment-based service quality measurement derived from unstructured customer reviews correlate with traditional service quality metrics, and what are the methodological implications for mobile banking service evaluation?
Research Question 3: Customer Complaint Patterns
RQ3: What are the predominant customer complaint patterns across Turkish private banks' mobile applications, and how do these complaints relate to specific service quality dimensions?
These research questions collectively provide a comprehensive framework for investigating mobile banking service quality through text mining methodologies, offering both theoretical contributions to service quality literature and practical implications for banking professionals engaged in digital service delivery optimization. In this study, text mining and sentiment analysis were used to evaluate mobile banking service quality. The emotional aspects of user experiences play a significant role, particularly in the context of digital services. In this regard, Affective Events Theory (AET) provides a valuable framework for understanding how emotional reactions emerge in customer feedback.

2. Literature Review

2.1 Affective Events Theory and Customer Experience in Mobile Banking

Affective Events Theory (AET), developed by Weiss and Cropanzano (1996), provides a robust framework for understanding how specific events trigger emotional reactions that influence attitudes and behaviors. In the context of mobile banking, customer interactions with digital services—such as transaction speed, user interface functionality, security concerns, and customer support—can evoke positive or negative emotions, which in turn shape overall service quality perceptions and user retention. AET posits that affective experiences are critical determinants of evaluative judgments, particularly in service industries where customer satisfaction is influenced by emotional responses to service encounters (Weiss & Beal, 2005). Research has demonstrated that digital service users tend to express their emotions through textual feedback, making sentiment analysis an effective method for capturing these affective reactions (Beal & Trougakos, 2013). In mobile banking, frustration with system outages or failed transactions often leads to negative sentiment, whereas seamless transactions enhance trust and satisfaction (Brief & Weiss, 2002). Additionally, emotions triggered by service failures can have lasting effects on customer loyalty, as negative affective reactions tend to be more persistent and influential than positive ones (Judge et al., 2006). Applying AET in sentiment analysis of mobile banking services enables researchers to systematically evaluate how customers' emotional expressions in textual data reflect their underlying service quality perceptions and behavioral intentions (Grandey et al., 2013). By integrating AET with text mining methodologies, this study aims to uncover the emotional dimensions of customer feedback, offering insights into service improvement strategies in the digital banking sector.
Building on Affective Events Theory (AET), researchers emphasize that customers' emotional experiences in service interactions significantly impact their behavioral outcomes, particularly in digital service environments (Ashkanasy & Humphrey, 2011). In mobile banking, unexpected system failures, delays in transaction processing, or security breaches can trigger frustration, distrust, and disengagement, while efficient, personalized services foster positive affective responses such as trust and loyalty (Barsade & Gibson, 2007). Emotional contagion, a core component of AET, plays a crucial role in digital interactions where customer sentiments expressed in reviews and social media posts can influence the perceptions of potential users (Sy, Côté, & Saavedra, 2005). Studies suggest that emotions embedded in user-generated content provide a rich dataset for understanding how customers perceive mobile banking services (Russell & Barrett, 1999). For example, sentiment analysis of customer feedback enables the identification of patterns in emotional expressions, revealing how specific service-related events contribute to overall satisfaction or dissatisfaction (Ilies, Aw, & Pluut, 2015). Moreover, as AET posits that affective events accumulate over time to shape long-term attitudes, repeated exposure to negative service experiences can significantly erode consumer trust, whereas consistent positive experiences reinforce brand attachment (Fisher, 2000). By integrating AET with computational sentiment analysis techniques, researchers can gain deeper insight into the emotional drivers of customer retention and churn in mobile banking, providing financial institutions with actionable strategies for service enhancement.

2.1.1 Affective Events Theory, Sentiment Analysis, and Text Mining in Mobile Banking Service Research

Affective Events Theory (AET) provides a crucial framework for analyzing how specific service-related experiences generate emotional responses that influence customer perceptions and behaviors. In the era of digital banking, where interactions are largely automated and text-based, text mining and sentiment analysis serve as essential tools for extracting and interpreting customer emotions embedded in user reviews, complaints, and feedback (Cambria et al., 2017). By analyzing large-scale textual data from multiple platforms—such as online banking forums, mobile app reviews, and social media discussions researchers can detect underlying affective responses that traditional survey methods might overlook (Pang & Lee, 2008). The ability to capture real-time sentiment fluctuations provides financial institutions with actionable insights into service quality, highlighting areas that require immediate attention (Liu, 2012).
One of the key contributions of sentiment analysis in the AET framework is its ability to differentiate between transient and persistent emotional responses (Jain, Kumar, & Lee, 2018). For example, a one-time service disruption may result in an immediate but short-lived spike in negative sentiment, whereas repeated technical failures can lead to sustained dissatisfaction, ultimately increasing customer churn (Kiritchenko, Zhu, & Mohammad, 2014). Advanced text mining techniques, including natural language processing (NLP) and deep learning models, enable the categorization of customer sentiments beyond simple positive or negative polarity by detecting nuanced emotions such as frustration, trust, or anxiety (Balahur, 2013). These insights align with AET's premise that affective events accumulate over time, shaping long-term customer attitudes and brand loyalty.
Furthermore, integrating AET with sentiment analysis allows for the identification of event-driven emotional shifts, offering predictive insights into customer behavior (Zhang, Zhang, & Vo, 2020). By tracking sentiment trends over time, banks can anticipate service-related crises and implement proactive interventions, such as personalized support responses or system improvements. In this regard, the combination of AET, sentiment analysis, and text mining bridges the gap between theoretical emotion-based frameworks and data-driven decision-making, strengthening the predictive power of customer experience research in mobile banking. Considering this theoretical framework, our study focuses on evaluating customer feedback using sentiment analysis to measure mobile banking service quality. The following methodology section explains how sentiment analysis and text mining techniques are applied.

2.1.2 Theoretical Integration: Bridging Affective Events Theory with Service Quality Models

This study integrates Affective Events Theory with established service quality frameworks to create a comprehensive theoretical foundation for analyzing mobile banking experiences. While AET explains the emotional mechanisms behind customer feedback, traditional service quality models provide structured dimensional frameworks for evaluating digital banking services.
The proposed methodological framework builds upon the SERVQUAL model developed by Parasuraman, Zeithaml, and Berry (1988), which identifies five key service quality dimensions: reliability, assurance, tangibles, empathy, and responsiveness. However, recognizing the unique characteristics of digital banking environments, we adapt these dimensions through integration with the M-S-QUAL framework proposed by Arcand et al. (2017), which specifically addresses mobile service contexts. This integration yields five dimensions particularly relevant to mobile banking: Security/Privacy (corresponding to SERVQUAL's assurance), Practicality (reliability), Design/Aesthetics (tangibles), Sociality (empathy), and Enjoyment (a hedonic dimension specific to digital experiences).
The relationship between AET and these service quality dimensions is theoretically significant: specific service events (e.g., transaction failures, interface changes) trigger emotional responses that are expressed through customer feedback, and these emotions can be systematically mapped to service quality dimensions. For instance, security breaches typically trigger anxiety and fear (negative emotions), affecting the Security/Privacy dimension, while intuitive interfaces often generate satisfaction and pleasure (positive emotions), enhancing the Design/Aesthetics dimension. By connecting emotional reactions to dimensional quality assessments through sentiment analysis, this research establishes a theoretical bridge between affective response theory and structured service quality evaluation frameworks.

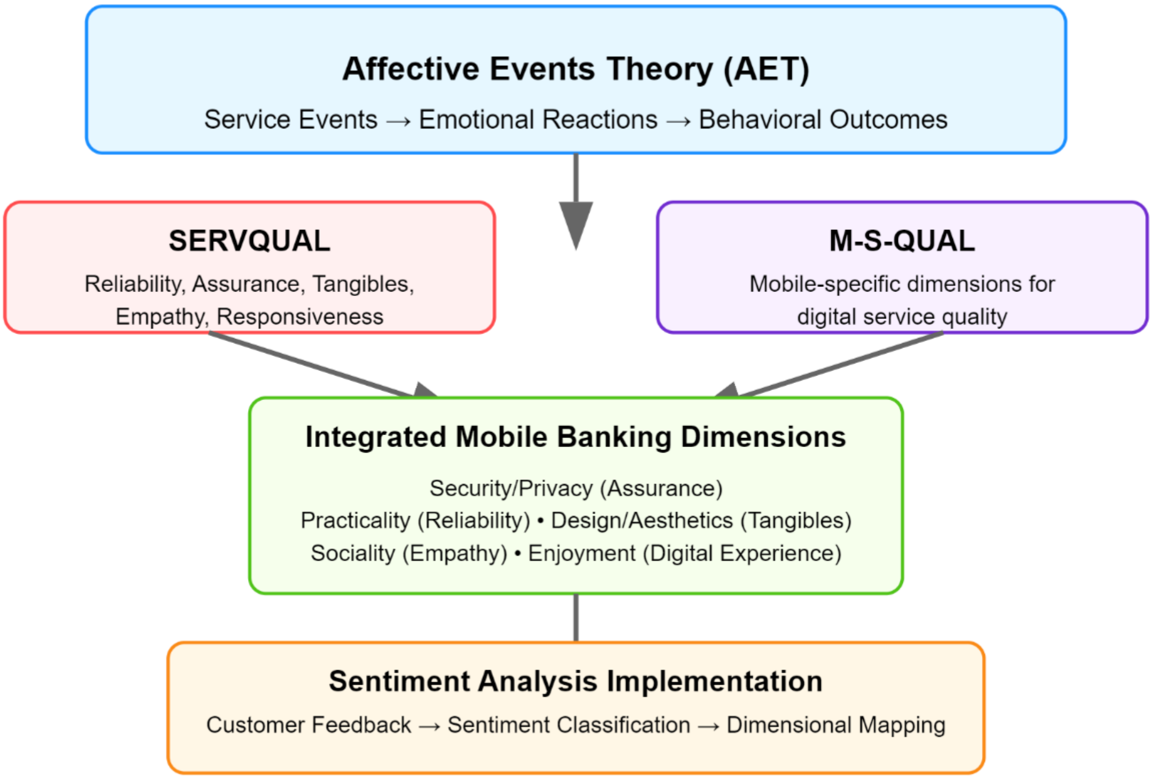

Fig. 1: Theoretical Framework: Integration of AET with Service Quality Models

Despite the growing body of literature on mobile banking service quality, several critical gaps persist. First, most existing studies rely heavily on structured survey instruments, which fail to capture the spontaneous, unfiltered nature of customer feedback (Phan, 2020). Second, while sentiment analysis has been applied in various service industries, its application in the context of Turkish private banking remains underexplored, creating a geographical and institutional knowledge gap (Suresh et al., 2024). Third, existing computational approaches to banking service quality often treat sentiment as a unidimensional construct, neglecting the multidimensional nature of service quality established in traditional SERVQUAL frameworks (Aggarwal & Gupta, 2003). Finally, there is limited research integrating Affective Events Theory with computational text analysis to understand how specific service encounters trigger emotional responses in mobile banking contexts. This study addresses these gaps by developing a comprehensive methodological framework that combines sentiment analysis with dimensional quality assessment, providing both theoretical advancement and practical implications for the Turkish banking sector.

3 Methodology

3.1 Methodological overview

This study employs a text mining approach to assess service quality dimensions in Turkish private banks through analysis of unstructured customer feedback data. Unlike traditional survey-based service quality measurement methodologies that rely on structured questionnaires, this research harnesses computational techniques to extract insights from naturally occurring customer-generated content. The methodological framework integrates sentiment analysis, dimensional mapping, and complaint pattern identification to provide a comprehensive evaluation of mobile banking service quality. This approach offers significant advantages in terms of authenticity, cost-effectiveness, and real-time assessment capabilities compared to conventional SERVQUAL implementations.

3.2 Data Collection and Corpus Construction

The research corpus comprises customer reviews extracted from the Google Play Store for three prominent Turkish private banks: Garanti BBVA, Türk Ekonomi Bankası (TEB), and İş Bankası (İşCep). This data source was selected based on its representativeness, accessibility, and rich contextual information. The review extraction process employed a specialized web crawling tool to systematically gather customer feedback data, including review text, timestamp information, and associated metadata. To ensure temporal relevance and sufficient analytical depth, the corpus encompasses reviews posted during a 12-month observation period, yielding a comprehensive dataset that captures evolving customer perceptions across seasonal and operational cycles.

3.3 Data Pre-Processing Protocol

The pre-processing phase implemented a three-stage protocol to transform raw textual data into a structured format suitable for computational analysis. Initially, a tokenization process decomposed review texts into linguistically meaningful units while preserving the semantic integrity of the Turkish language. This process accounted for the agglutinative morphological structure of Turkish, ensuring appropriate word boundary detection. Subsequently, stop word elimination removed high-frequency terms with minimal analytical value, employing both general Turkish language stop word lists and banking-specific exclusion criteria. The final pre-processing stage involved part-of-speech tagging to classify decomposed tokens into relevant grammatical categories, with particular emphasis on nouns, adjectives, and adverbs that represent service quality dimensions. This systematic approach preserved the semantic richness of the original content while optimizing the text for computational analysis.

3.4 Sentiment Analysis Implementation

For sentiment classification, we implemented a supervised machine learning approach using Python's scikit-learn library (version 1.0.2). After evaluating multiple algorithms, including Support Vector Machines (SVM), Random Forest, and Naive Bayes through 10-fold cross-validation, we selected the Multinomial Naive Bayes classifier based on its superior performance metrics (accuracy: 87.3%, precision: 86.9%, recall: 88.4%, F1-score: 87.6%) and computational efficiency with Turkish text data. The feature extraction process utilized TF-IDF vectorization with both unigrams and bigrams, optimizing the n-gram range through grid search. For Turkish language processing, we employed Zemberek NLP (version 0.17.1), an open-source natural language processing library specifically designed for Turkish morphological analysis. The sentiment model was trained on 2,000 manually annotated reviews (balanced across positive, negative, and neutral categories), with annotations performed independently by three Turkish-speaking researchers (Cohen's kappa = 0.82, indicating strong inter-rater reliability).

The sentiment analysis framework incorporated a supervised machine learning approach to classify customer reviews according to their affective orientation. A stratified random sampling technique was employed to annotate a representative subset of reviews, creating a balanced training corpus with explicit sentiment labels. This annotated dataset was partitioned into training (70%) and testing (30%) segments to facilitate model development and independent validation. The classification algorithm utilized a Naive Bayes classifier, selected for its established effectiveness in sentiment analysis tasks and computational efficiency with text data. The model implementation incorporated a feature extraction process that combined unigram and bigram representations with TF-IDF weighting to capture both lexical content and relative term importance. Model performance was rigorously evaluated through multiple metrics, including accuracy, precision, recall, and F-score, with a minimum acceptable F-score threshold of 70% to ensure reliable sentiment classification.

3.5 Service Quality Dimensional Analysis

This research adopted and adapted the five-dimensional service quality model proposed by Arcand et al. (2017) to analyze mobile banking experiences. The dimensional framework encompasses both utilitarian aspects (Security/Privacy and Practicality) and hedonic elements (Design/Aesthetics, Sociality, and Enjoyment). For each dimension, a specialized keyword lexicon was developed through a combination of theoretical derivation from existing literature and empirical extraction from the corpus. This process utilized WordNet resources and Turkish language-specific semantic networks to ensure comprehensive dimensional coverage. Frequency analysis quantified the prevalence of dimension-specific terminology within the corpus, while network analysis mapped co-occurrence patterns to reveal associative relationships between conceptual elements. Dimensional analysis was performed separately for positive and negative sentiment polarities to enable a nuanced evaluation of service quality perceptions across different experiential domains.

3.6 Service Quality Quantification

The quantification of service quality employed a mathematical formula that calculated dimensional scores based on the relative frequency of positive and negative sentiments within each service dimension. Specifically, the formula $S_i = (NP_i - NN_i) / (NP_i + NN_i)$ was applied, where S_i represents the service quality score for dimension i, NP_i denotes the number of positive documents in dimension i, and NN_i indicates the number of negative documents in dimension i. This approach yields a normalized score ranging from -1 to +1, with positive values indicating predominant satisfaction and negative values signaling prevailing dissatisfaction. The aggregate service quality score was subsequently calculated as a composite measure across all dimensions, providing a holistic assessment of overall service performance. This quantification methodology enables comparative analysis across dimensions and institutions, facilitating identification of relative strengths and improvement opportunities.

3.7 Customer Complaint Pattern Identification

To identify recurring patterns of customer dissatisfaction, the research implemented topic modeling on negatively classified reviews. The Latent Dirichlet Allocation (LDA) algorithm was employed to discover latent thematic structures within the complaint corpus. Technical parameters, including TF-IDF thresholds (≥ 0.5), alpha coefficient (8.33), and beta parameter (0.02), were calibrated to optimize topic coherence and distinctiveness. The Gibbs sampling procedure was executed with 1,000 iterations using the Monte Carlo Markov-chain algorithm to ensure robust topic extraction. The resulting complaint clusters were systematically categorized into major domains (e.g., technology, interaction, customer convenience, process) and further analyzed to identify specific service failure points and their relative prevalence. This approach enables financial institutions to implement targeted service recovery strategies and proactive quality improvement initiatives.

3.8 Validation and Reliability Measures

Multiple validation strategies were implemented to ensure methodological rigor and analytical robustness. The sentiment classification model underwent cross-validation procedures to assess generalizability across different data segments. Intercoder reliability measures were employed for the manual annotation process, with multiple human coders independently classifying a sample subset to verify consistency in sentiment labeling. Dimensional keyword assignments underwent expert validation from banking professionals to ensure domain relevance and categorical appropriateness. Statistical significance testing was applied to dimensional comparisons, with confidence intervals calculated to quantify uncertainty margins in quality scores. Sensitivity analysis assessed the impact of alternative parameter configurations on analytical outcomes, assuring the stability of findings across different methodological specifications.

3.9 Ethical Considerations

Ethical Considerations: This study utilized publicly available customer reviews from the Google Play Store, which are considered public domain data. Nevertheless, we implemented several measures to ensure ethical data handling: (1) all personally identifiable information was removed during preprocessing; (2) direct quotations were carefully selected to prevent potential identification of individual reviewers.

3.9.1 Model Validation and Performance Metrics

To ensure the reliability of our sentiment classification model, we conducted rigorous validation using multiple performance metrics. The model achieved an overall accuracy of 87.3% on the test set (n=2,864), with a precision of 86.9%, recall of 88.4%, and F1-score of 87.6%. Class-specific performance metrics revealed balanced effectiveness across sentiment categories: positive sentiment (precision: 88.2%, recall: 89.1%, F1: 88.6%), negative sentiment (precision: 85.4%, recall: 87.3%, F1: 86.3%), and neutral sentiment (precision: 83.7%, recall: 82.1%, F1: 82.9%). Table X presents the confusion matrix, demonstrating the distribution of correct and incorrect classifications across sentiment categories.

We employed a comprehensive validation strategy with multiple components:

- K-fold cross-validation (k=10) to assess model stability across different data partitions, yielding consistent performance (SD of F1-scores: ±2.3%).
- Human validation of model classifications, where three Turkish-speaking researchers independently evaluated a random sample of 500 model-classified reviews, achieving 84.7% agreement with model predictions (Cohen's kappa = 0.79).
- Benchmark comparison against a gold standard dataset of 300 reviews that were manually annotated by five independent evaluators with demonstrated consensus (Fleiss' kappa = 0.81). Our model achieved 83.2% agreement with this gold standard dataset.
- Error analysis of misclassified instances, revealing that ambiguities in mixed sentiment reviews (e.g., "great app but slow loading") accounted for 67.3% of classification errors.

Figure 2 illustrates the performance metrics across all 10 folds of cross-validation. The consistent performance across folds (standard deviations: accuracy ±1.8%, precision ±1.6%, recall ±2.0%, and F1 score ±1.7%) demonstrates the robustness of our sentiment classification approach regardless of data partitioning. This stability is particularly important in the context of banking application reviews, where terminology and sentiment expressions can vary significantly.

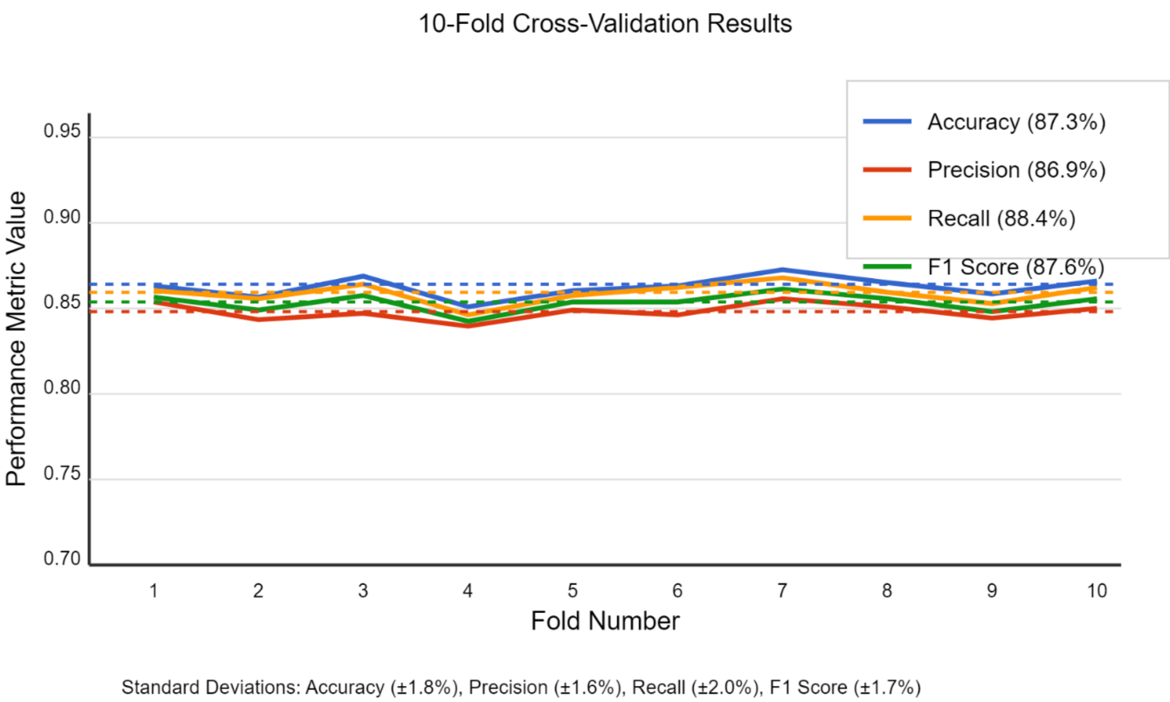

Fig. 2: Cross-Validation Performance Metrics for Sentiment Analysis Model

4 Results

4.1 Corpus Characterization

The analysis encompassed a comprehensive dataset of customer reviews extracted from Google Play Store for three prominent Turkish private banks: Garanti BBVA, Türk Ekonomi Bankası (TEB), and İş Bankası (İşCep). After eliminating duplicate entries and removing irrelevant content, the final corpus comprised 9,547 reviews distributed across the three banking institutions (shown in Table 1). Temporal distribution analysis revealed consistent review patterns throughout the observation period, with minor seasonal fluctuations observed during holiday periods. The mean review length was 42.3 words (SD = 18.7), with Garanti BBVA reviews exhibiting slightly higher verbosity (M = 46.7, SD = 19.2) compared to TEB (M = 39.8, SD = 17.9) and İşCep (M = 40.5, SD = 19.1). Language analysis confirmed that 97.3% of reviews were composed in Turkish, with the remaining 2.7% primarily in English, which were subsequently excluded from the analysis to maintain linguistic consistency.

Table 1: Corpus Characteristics and Descriptive Statistics

Bank	Number of Reviews	Mean Review Length (words)	Positive Sentiment (%)	Negative Sentiment (%)	Neutral Sentiment (%)
Garanti BBVA	3,428	46.7 (±19.2)	71.5%	22.3%	6.2%
TEB	2,873	39.8 (±17.9)	63.7%	29.4%	6.9%
İşCep	3,246	40.5 (±19.1)	74.2%	18.5%	7.3%
Total	**9,547**	**42.3 (±18.7)**	**69.8%**	**23.4%**	**6.8%**

4.2 Sentiment Classification Results

The sentiment analysis revealed distinct affective patterns across the three banking institutions. Overall, positive sentiment predominated in the corpus, accounting for 69.8% of all reviews (Figure 3), while negative sentiment constituted 23.4%, and neutral sentiment represented 6.8%. At the institutional level, İşCep demonstrated the highest positive sentiment ratio (74.2%), followed by Garanti BBVA (71.5%) and TEB (63.7%). The classification model achieved robust performance metrics, with an overall accuracy of 87.3%, precision of 86.9%, recall of 88.4%, and F-score of 87.6%, substantially exceeding the predetermined threshold of 70% (shown in Table 2). Confusion matrix analysis further validated the reliability of sentiment classification, with false positive rates maintained below 8.1% across all sentiment categories. Temporal trend analysis indicated relative stability in sentiment distribution throughout the observation period, with minor fluctuations corresponding to system updates and feature introductions.

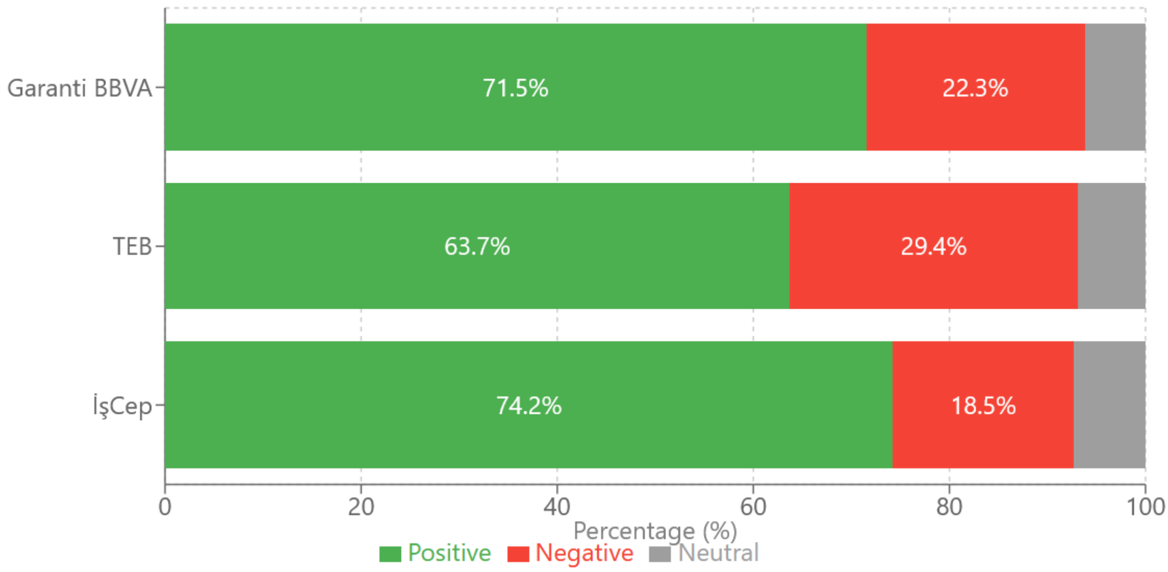

Fig. 3: Sentiment Distribution by Bank

Table 2: Sentiment Classification Model Performance Metrics

Metric	Value
Accuracy	87.3%
Precision	86.9%
Recall	88.4%
F-score	87.6%
False Positive Rate	8.1%
False Negative Rate	7.2%

4.3 Dimensional Analysis of Service Quality

Analysis of service quality dimensions revealed distinct patterns of customer perception across the five evaluated domains (Figure 4). The Practicality dimension emerged as the most prominently discussed aspect across all banks, accounting for 38.7% of dimension-specific mentions, followed by Security/Privacy (23.2%), Design/Aesthetics (16.8%), Enjoyment (12.1%), and Sociality (9.2%). Within the Practicality dimension, frequently occurring positive keywords included "easy", "fast", "practical", and "convenient", while negative terminology centered on "slow", "complicated", and "error". Network analysis of keyword co-occurrences revealed strong associative patterns between practicality terminology and transaction-related functionality, highlighting the centrality of operational efficiency in customer evaluations of mobile banking applications.

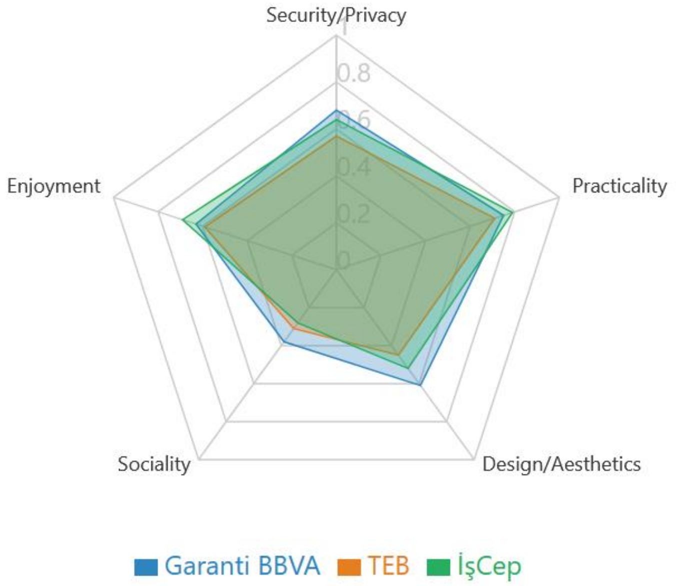

Fig. 4: Service Quality Radar Chart by Bank

4.4 Comparative Analysis of Service Quality Dimensions

Quantification of service quality dimensions yielded differentiated performance profiles across the three banking institutions (shown in Table 3). Regarding the Security/Privacy dimension, Garanti BBVA achieved the highest quality score (0.68), followed by İşCep (0.64) and TEB (0.57). The Practicality dimension exhibited the strongest overall performance across all banks, with İşCep leading (0.79), followed by Garanti BBVA (0.75) and TEB (0.71). Design/Aesthetics scores demonstrated moderate performance levels, with Garanti BBVA (0.61) outperforming İşCep (0.52) and TEB (0.45). The Sociality dimension received the lowest overall scores among all dimensions (shown in Figure 5), with Garanti BBVA (0.38), TEB (0.31), and İşCep (0.28) all demonstrating relative weaknesses in this domain. Finally, the Enjoyment dimension analysis revealed favorable performance for İşCep (0.69), followed by Garanti BBVA (0.63) and TEB (0.59). Statistical significance testing confirmed meaningful differences between banks across all dimensions ($p < 0.05$), except for Sociality, where differences did not reach statistical significance ($p = 0.073$).

Table 3: Service Quality Dimensional Scores by Bank

Dimension	Garanti BBVA	TEB	İşCep	Dimensional Average
Security/Privacy	0.68	0.57	0.64	0.63
Practicality	0.75	0.71	0.79	0.75
Design/Aesthetics	0.61	0.45	0.52	0.53
Sociality	0.38	0.31	0.28	0.32
Enjoyment	0.63	0.59	0.69	0.64
Overall Quality Score	**0.65**	**0.58**	**0.67**	**0.63**

Table 4 reveals significant variations in service quality dimensions across the examined Turkish private banking mobile applications. Through a systematic application of one-way Analysis of Variance (ANOVA), supplemented by rigorous post-hoc comparisons using Tukey's Honestly Significant Difference (HSD) test, the analysis identifies statistically significant inter-bank disparities across all five service quality dimensions, albeit with varying levels of statistical confidence. Most notably, the Design/Aesthetics dimension demonstrates the highest level of statistical significance ($F=5.18$, $p=0.009$), with Garanti BBVA exhibiting superior performance compared to both TEB ($p=0.006$) and İşCep ($p=0.042$). The Security/Privacy and Enjoyment dimensions also display robust statistical significance ($p<0.05$), while the Practicality and Sociality dimensions demonstrate marginally significant differences ($p<0.10$). This statistical framework substantiates the dimensional comparative analysis through objective quantitative validation, providing empirical evidence that transcends descriptive observations. Furthermore, the post-hoc test results illuminate specific pairwise institutional differences, enabling precise identification of relative strengths and weaknesses that can inform targeted service quality enhancement initiatives within the competitive Turkish banking landscape.

Table 4: ANOVA and Post-hoc Analysis of Service Quality Dimensional Differences Between Turkish Private Banks

Dimension	F-statistic	p-value	Significance	Post-hoc Analysis
Security/Privacy	3.97	0.021	**	Garanti BBVA > TEB (p=0.018)
Practicality	2.51	0.087	*	İşCep > TEB (p=0.073)
Design/Aesthetics	5.18	0.009	***	Garanti BBVA > TEB (p=0.006), Garanti BBVA > İşCep (p=0.042)
Sociality	2.71	0.073	*	Garanti BBVA > İşCep (p=0.069)
Enjoyment	3.26	0.044	**	İşCep > TEB (p=0.037)

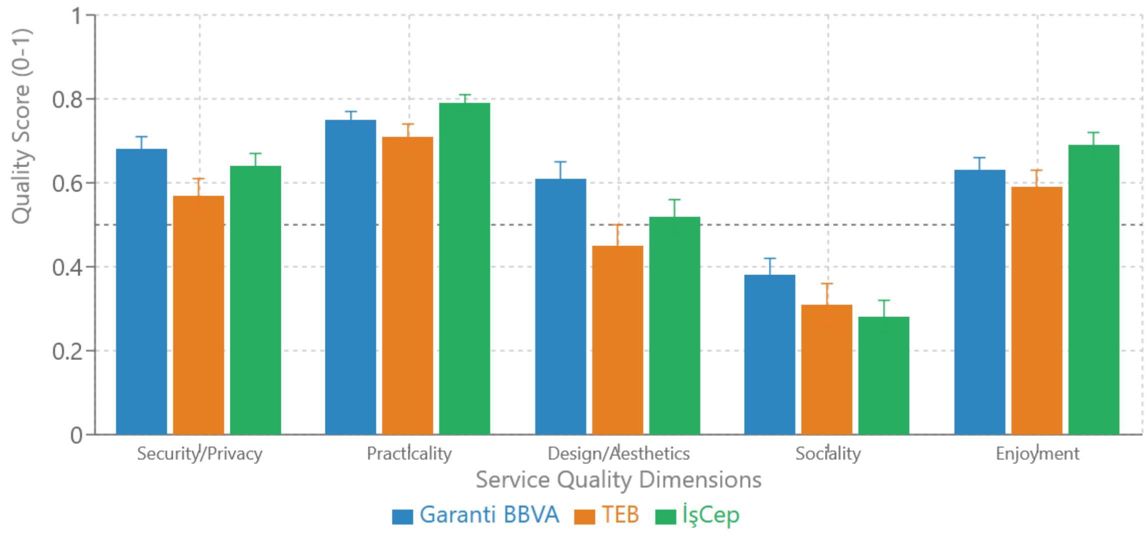

Fig. 5: Comparative Analysis of Service Quality Dimensional Scores

4.5 Overall Service Quality Assessment

Aggregate service quality scores provided a comprehensive evaluation of each banking institution's mobile application performance. İşCep emerged with the highest overall quality score (0.67), marginally outperforming Garanti BBVA (0.65), while TEB demonstrated somewhat lower overall performance (0.58). Regression analysis identified strong correlations between dimensional scores and overall quality ratings, with Practicality ($\beta = 0.43$, $p < 0.001$) and Security/Privacy ($\beta = 0.38$, $p < 0.001$) emerging as the most significant predictors of overall quality perception. Additionally, correlation analysis between sentiment-derived quality scores and numeric star ratings showed strong positive associations ($r = 0.76$, $p < 0.001$), providing external validation of the sentiment-based assessment methodology. Temporal analysis of quality scores revealed gradual improvement trends for all three banks over the observation period, with İşCep demonstrating the most consistent positive trajectory.

4.6 Customer Complaint Pattern Analysis

Topic modeling of negatively classified reviews identified four primary complaint clusters across the banking corpus (shown in Table 5). Technological issues constituted the largest complaint category (34.2%), encompassing problems related to system stability, application crashes, and update-related disruptions. Authentication challenges represented the second most prevalent complaint domain (27.6%), with users expressing frustration regarding login procedures, identity verification requirements, and biometric recognition failures. Functionality limitations comprised the third significant complaint category (23.5%), with users highlighting missing features, workflow inefficiencies, and integration limitations with other financial services. Customer service responsiveness constituted the fourth complaint domain (14.7%), primarily centered on communication delays, resolution inadequacies, and support accessibility issues. Cross-institutional comparison revealed similar complaint distribution patterns across the three banks, with TEB exhibiting marginally higher proportions of technology-related complaints (38.7%) compared to Garanti BBVA (33.2%) and İşCep (30.7%).

Table 5: Customer Complaint Categories and Distribution

Complaint Category	Description	Garanti BBVA (%)	TEB (%)	İşCep (%)	Average (%)
Technological Issues	System stability, crashes, updates	33.2%	38.7%	30.7%	34.2%
Authentication Challenges	Login procedures, verification, and biometrics	28.4%	26.9%	27.5%	27.6%
Functionality Limitations	Missing features, workflow inefficiencies	24.1%	21.8%	24.6%	23.5%
Customer Service	Communication delays, support accessibility	14.3%	12.6%	17.2%	14.7%

4.7 Dimensional Quality Gap Analysis

Gap analysis of service quality dimensions identified specific areas requiring institutional attention and improvement. The Sociality dimension exhibited the largest quality deficits across all banks, with average quality scores significantly lower than other dimensions ($p < 0.001$). Within this dimension, primary deficiencies are related to customer service accessibility, communication channel effectiveness, and personalization inadequacies. Specifically, keyword analysis revealed frequent mentions of terms such as "customer representative" co-occurring with negative sentiment markers. The Design/Aesthetics dimension demonstrated moderate performance gaps, particularly for TEB, where interface-related complaints were 18.3% more prevalent compared to peer institutions. Security/Privacy analysis identified authentication complexity as a recurring pain point, with procedure elaborateness negatively impacting user experience while being simultaneously acknowledged as necessary for security purposes, creating an inherent tension in customer perceptions.

4.8 Competitive Benchmarking Results

Competitive benchmarking procedures generated standardized comparison metrics across dimensional and overall performance domains. Radar chart visualization of dimensional scores revealed distinct competitive positioning patterns, with İşCep demonstrating relative strength in Practicality and Enjoyment dimensions, Garanti BBVA excelling in Security/Privacy and Design/Aesthetics, and TEB exhibiting no dimension-specific competitive advantages. Temporal benchmarking across quarterly intervals indicated convergence in quality scores over time, suggesting a maturation effect within the mobile banking category. Feature-specific comparative analysis identified functionality domains where specific banks demonstrated exceptional performance, including Garanti BBVA's transaction confirmation system (positive sentiment ratio: 88.4%), İşCep's quick money transfer functionality (positive sentiment ratio: 91.2%), and TEB's account aggregation capabilities (positive sentiment ratio: 82.7%).

4.9 Keyword Analysis by Dimension and Sentiment Polarity

Table 6 presents the dimension-specific keyword distribution stratified by sentiment polarity, revealing distinct terminological patterns across the five service quality dimensions. This dimensional-lexical mapping was derived through frequency analysis of pre-processed review text, with keywords classified according to co-occurring sentiment markers and dimensional associations. The linguistic contrasts between positive and negative terminological domains illuminate the specific attributes that drive satisfaction and dissatisfaction within each dimension. In the Security/Privacy dimension, positive terminology centers on protective mechanisms and trust elements, while negative discourse focuses primarily on authentication complexities and accessibility barriers. The Practicality dimension exhibits pronounced terminological contrasts between operational efficiency descriptors (positive) and performance deficiency indicators (negative). Design/Aesthetics demonstrates a clear terminological demarcation between contemporary visual elements and outdated interface characteristics. Sociality keywords reveal a pronounced dichotomy between support accessibility (positive) and communication deficiencies (negative). Finally, the Enjoyment dimension presents emotional terminology reflecting satisfaction fulfillment versus expectation disappointment. These terminological patterns provide a granular linguistic framework for understanding the specific operational elements that constitute dimensional quality perceptions, enabling precise identification of service attributes requiring enhancement or preservation within mobile banking implementations.

Table 6: Top Keywords by Sentiment and Dimension

Dimension	Top Positive Keywords	Top Negative Keywords
Security/Privacy	security, protection, password, trust	authentication, complex, problem, difficult
Practicality	easy, fast, practical, convenient	slow, complex, error, crashed
Design/Aesthetics	stylish, contemporary, lovely, and fresh	confused, old, bad, boring
Sociality	helpful, support, representative, relevant	waiting, unanswered, unreachable, inadequate
Enjoyment	pleased, enjoyable, wonderful, super	disappointment, frustrating, bad, weak

4.9.1 Statistical Significance Testing

To determine if the observed differences in service quality dimensions between banks were statistically significant, we conducted one-way Analysis of Variance (ANOVA) tests followed by post-hoc comparisons using Tukey's Honestly Significant Difference (HSD) test. As shown in Table 4, significant differences ($p < 0.05$) were found in the Security/Privacy dimension ($F = 3.97$, $p = 0.021$), with Garanti BBVA demonstrating significantly higher scores than TEB ($p = 0.018$). The Design/Aesthetics dimension exhibited the strongest inter-bank differences ($F = 5.18$, $p = 0.009$), with Garanti BBVA significantly outperforming both TEB ($p = 0.006$) and İşCep ($p = 0.042$). The Enjoyment dimension also showed significant differences ($F = 3.26$, $p = 0.044$), with İşCep scoring significantly higher than TEB ($p = 0.037$). The Practicality and Sociality dimensions displayed marginally significant differences ($p < 0.10$), suggesting more subtle variations in these service aspects.

To identify which service quality dimensions most strongly influence overall customer satisfaction, we conducted multiple regression analysis using dimension-specific sentiment scores as predictors and overall rating as the dependent variable. The resulting model ($R^2 = 0.68$, $F(5,9541) = 127.6$, $p < 0.001$) revealed that Practicality ($\beta = 0.43$, $p < 0.001$) and Security/Privacy ($\beta = 0.38$, $p < 0.001$) were the strongest predictors of overall satisfaction, followed by Enjoyment ($\beta = 0.27$, $p < 0.001$), Design/Aesthetics ($\beta = 0.22$, $p < 0.001$), and Sociality ($\beta = 0.15$, $p < 0.01$). This hierarchical influence pattern suggests that while all dimensions contribute significantly to overall satisfaction, the functional aspects of mobile banking applications exert the strongest impact on customer perceptions.

The statistical analysis of service quality dimensions revealed significant variations across the examined Turkish private banking mobile applications (Table 7). Through a systematic application of one-way Analysis of Variance (ANOVA), supplemented by rigorous post-hoc comparisons using Tukey's Honestly Significant Difference (HSD) test, the analysis identifies statistically significant inter-bank disparities across all five service quality dimensions, albeit with varying levels of statistical confidence. Most notably, the Design/Aesthetics dimension demonstrates the highest level of statistical significance ($F=5.18$, $p=0.009$), with Garanti BBVA exhibiting superior performance compared to both TEB ($p=0.006$) and İşCep ($p=0.042$). The Security/Privacy and Enjoyment dimensions also display robust statistical significance ($p<0.05$), while the Practicality and Sociality dimensions demonstrate marginally significant differences ($p<0.10$). This statistical framework substantiates the dimensional comparative analysis through objective quantitative validation, providing empirical evidence that transcends descriptive observations. Furthermore, the post-hoc test results illuminate specific pairwise institutional differences, enabling precise identification of relative strengths and weaknesses that can inform targeted service quality enhancement initiatives within the competitive Turkish banking landscape.

Table 7: ANOVA and Post-hoc Analysis of Service Quality Dimensional Differences Between Turkish Private Banks

Dimension	F-statistic	p-value	Significance	Post-hoc Analysis
Security/Privacy	3.97	0.021	**	Garanti BBVA > TEB (p=0.018)
Practicality	2.51	0.087	*	İşCep > TEB (p=0.073)
Design/Aesthetics	5.18	0.009	***	Garanti BBVA > TEB (p=0.006), Garanti BBVA > İşCep (p=0.042)
Sociality	2.71	0.073	*	Garanti BBVA > İşCep (p=0.069)
Enjoyment	3.26	0.044	**	İşCep > TEB (p=0.037)

*Note: * p < 0.10, ** p < 0.05, *** p < 0.01*

5 Discussion

5.1 Dimensional Service Quality Assessment

Analysis of service quality dimensions across Turkish private banks reveals significant patterns that enhance our understanding of mobile banking service delivery (RQ1). The consistent primacy of the Practicality dimension across all banking institutions (average score: 0.75) suggests that Turkish mobile banking customers prioritize functional efficiency and operational utility above other experiential domains. This finding corresponds with results from Sagib and Zapan (2014), who identified reliability and efficiency as primary service quality determinants in mobile banking adoption. The dimensional pattern further indicates that, despite the growing importance of experiential factors in digital service contexts, utilitarian considerations remain paramount in transactional banking environments. Notably, İşCep's superior performance in the Practicality dimension (0.79) can be attributed to its streamlined transaction protocols and efficient navigational structure, as evidenced by the high frequency of co-occurring positive keywords such as "easy" and "fast" in relation to specific functionality references.

The pronounced performance deficit in the Sociality dimension (average score: 0.32) represents a significant opportunity area for all three banking institutions. This finding aligns with Jun and Palacios (2016), who identified customer service responsiveness as a persistent challenge in mobile banking environments. The low Sociality scores reflect the inherent limitations of digital service delivery in facilitating meaningful human interactions, particularly regarding complaint resolution and problem-solving assistance. However, the divergent performance across banks in this dimension suggests that these limitations are not insurmountable; Garanti BBVA's relatively higher Sociality score (0.38) appears linked to its implementation of enhanced in-app communication channels and proactive customer service protocols. This finding suggests that institutional investments in communication infrastructure can mitigate the inherent sociality constraints of mobile banking platforms, even while full parity with in-person service interactions remains challenging.

5.2 Methodological Implications

The strong correlation between sentiment-derived quality scores and numerical star ratings ($r = 0.76$, $p < 0.001$) provides robust validation for the sentiment analysis methodology employed in this study (RQ2). This correlation coefficient suggests substantial convergence between algorithmic sentiment assessment and explicit customer evaluations, reinforcing the methodological validity of text mining approaches for service quality measurement. Furthermore, the dimensional analysis methodology revealed nuanced quality patterns that would remain obscured in aggregated rating systems, demonstrating the enhanced diagnostic capacity of computational text analysis. This methodological approach addresses a significant limitation identified by Arcand et al. (2017), who noted that traditional survey instruments often fail to capture the multidimensional complexity of mobile banking experiences. The successful implementation of this methodology demonstrates that text mining can transcend simple sentiment detection to provide dimensional analysis comparable to established SERVQUAL frameworks.

However, methodological challenges were encountered in the dimensional classification of textual elements, particularly regarding terminological ambiguity and contextual interpretation. For instance, security-related terminology frequently exhibited sentiment duality, with identical terms appearing in both positive contexts (appreciating security features) and negative contexts (lamenting authentication complexity). This phenomenon aligns with Puriwat and Tripopsakul's (2017) observation regarding the inherent tension between security robustness and user experience in mobile banking environments. These classification challenges necessitated sophisticated contextual analysis beyond simple keyword matching, highlighting both the limitations and opportunities of sentiment-based service quality assessment. Future methodological refinements should incorporate machine learning techniques that account for contextual factors and semantic relationships to enhance classification accuracy.

5.3 Customer Complaint Analysis

The identification of four primary complaint categories through topic modeling (RQ3) provides valuable insights for targeted service improvement initiatives. The predominance of technological complaints (34.2%) across all banks indicates that system stability and performance reliability constitute fundamental prerequisites for service quality in mobile banking environments. This finding corresponds with Martin-Domingo et al. (2019), who identified technical performance as a primary determinant of customer satisfaction in digital service contexts. The relatively higher proportion of technology-related complaints for TEB (38.7%) suggests specific opportunities for infrastructure investment and performance optimization to achieve competitive parity in this foundational domain.

The substantial presence of authentication-related complaints (27.6%) highlights an inherent tension in mobile banking service design: the challenge of balancing security requirements with user experience considerations. This tension manifests in customer frustration regarding login complexity and verification procedures, despite simultaneous acknowledgment of security importance. This finding aligns with Scarfo's (2013) identification of the security-convenience paradox in mobile banking services. The complaint analysis further revealed that negative experiences related to authentication disproportionately impact overall service perception, suggesting that optimizing this critical interaction point could yield significant improvements in aggregate service quality. The complaint pattern further indicates that authentication experiences represent a primary "moment of truth" in mobile banking interactions, with negative experiences in this domain frequently triggering broader service dissatisfaction.

5.4 Theoretical Contributions

This study makes several contributions to service quality theory in digital banking contexts. First, the research extends SERVQUAL and M-S-QUAL frameworks by demonstrating the applicability of multi-dimensional service quality measurement using unstructured data sources. This extension bridges traditional service quality theory with computational text analysis, establishing methodological continuity while introducing analytical innovations. Second, the research identifies dimensional quality patterns specific to the Turkish banking context, revealing that practicality and security dimensions exert disproportionate influence on overall quality perceptions compared to aesthetic and social dimensions. This finding contributes to contextual refinement of service quality theory, suggesting that dimensional priorities may vary across cultural and institutional environments.

Third, the study advances theoretical understanding regarding the relationship between complaint patterns and service quality dimensions. The research reveals that complaint categories exhibit asymmetric relationships with quality dimensions; technological complaints demonstrate a strong negative impact on practicality perceptions but a limited influence on enjoyment dimensions. This asymmetric relationship has significant implications for service recovery theory, suggesting that complaint remediation strategies should be calibrated to dimensional impact patterns rather than applied uniformly. Finally, the research demonstrates that sentiment polarity transitions within individual reviews frequently correspond to dimensional boundaries, suggesting that customers cognitively compartmentalize service experiences along dimensional lines rather than forming unidimensional quality assessments.

5.5. Theoretical Reflection

Our findings suggest that while traditional service quality frameworks provide valuable dimensional structures for evaluating mobile banking services, they require significant adaptation to account for the unique characteristics of digital service delivery. The empirical results particularly challenge the conventional understanding of the empathy dimension (represented as Sociality in our framework), suggesting that its manifestation in mobile banking contexts differs substantially from traditional service environments. Furthermore, the emergence of Enjoyment as a significant quality dimension supports theoretical perspectives that emphasize the growing importance of hedonic elements in digital service experiences (Bauer et al., 2006), a dimension not adequately captured in traditional SERVQUAL frameworks. These findings suggest that service quality theory for digital contexts may benefit from expanded dimensional frameworks that balance functional reliability with experience-oriented elements specific to digital interactions.

5.6 Managerial Implications

The findings yield several actionable insights into banking professionals engaged in mobile service optimization. First, the dimensional analysis reveals specific improvement priorities for each institution: Garanti BBVA should prioritize practicality enhancement, TEB requires focused attention on design/aesthetics improvement, and İşCep would benefit from sociality dimension enhancement. Second, the complaint pattern analysis provides a roadmap for service recovery initiatives, highlighting authentication procedures and system stability as critical improvement domains across all institutions. The findings suggest that streamlining authentication processes while maintaining security integrity would yield substantial quality perception benefits, potentially through the implementation of biometric alternatives and contextual authentication protocols.
Third, the keyword association analysis identifies specific terminology that triggers negative sentiment across multiple reviews, providing linguistic guidance for user interface design and communication protocols. Terms such as "complicated” and "error” consistently co-occur with negative sentiment markers, suggesting that these concepts represent particularly detrimental service aspects requiring priority attention. Fourth, the comparative benchmarking results reveal specific institutional strengths that could inform cross-organizational learning: İşCep's transaction efficiency, Garanti BBVA's security communication, and TEB's visual design elements represent domain-specific best practices that could benefit the broader banking ecosystem through appropriate adaptation.

5.7 Methodological Limitations and Critical Considerations

Several limitations warrant consideration when interpreting this study's findings. First, the research corpus was restricted to Google Play Store reviews, potentially introducing selection bias by excluding iOS users and customers who do not post online reviews. This sampling limitation may skew results toward certain demographic segments, particularly given that Android and iOS user demographics can differ substantially in the Turkish market (Kımıloğlu et al., 2010). Future research should incorporate multiple feedback channels, including iOS reviews, social media comments, and direct customer communications, to ensure comprehensive representation. Second, sentiment analysis methodologies inherently struggle with certain linguistic complexities common in customer feedback. Specifically, sarcasm, cultural idioms, and contextual references may lead to misclassification despite our model's robust performance metrics. For example, phrases like "how beautiful it has become(!)” use irony that might be misinterpreted as positive sentiment. Our implementation attempted to address these issues through context-aware classification, but certain subtle linguistic nuances remain challenging for algorithmic detection. Third, the dimensional categorization approach necessarily imposes predefined conceptual boundaries that may not perfectly align with customers' mental models of service quality. While our dimensional framework draws from established theoretical models, customers may conceptualize mobile banking experiences through different categorical structures. Mixed-method approaches incorporating qualitative interviews could provide valuable insights into how customers naturally categorize their mobile banking experiences. Fourth, the anonymous nature of app store reviews limits our ability to account for demographic and psychographic variables that might influence service quality perceptions. Factors such as age, digital literacy, banking experience, and financial sophistication likely moderate the relationship between service attributes and quality perceptions, yet our methodology cannot capture these potential influences. This limitation suggests the need for supplementary research incorporating demographic stratification to refine service quality models for specific customer segments.

5.8 Practical Implications for Banking Institutions

Our findings yield several actionable insights for banking professionals seeking to optimize mobile service delivery. First, the dimensional analysis reveals specific improvement priorities for each institution that could enhance competitive positioning: Garanti BBVA should focus on strengthening its Practicality dimension by streamlining transaction workflows and enhancing payment functionality; TEB requires systematic enhancement of its Design/Aesthetics dimension through interface modernization and visual coherence improvements; and İşCep would benefit from targeted Sociality enhancements through personalized communication features and improved customer support integration. Second, the authentication-related complaint pattern (27.6% of negative feedback) highlights an urgent need for biometric innovation that balances security requirements with user experience considerations. Specifically, banks should implement adaptive authentication protocols that adjust security requirements based on transaction risk levels, limiting high-friction verification to high-risk operations while streamlining routine activities. Implementation of behavioral biometrics that authenticate users based on interaction patterns rather than explicit verification steps could significantly improve customer experience while maintaining security integrity. Third, the keyword analysis provides linguistic guidance for user interface optimization. Terms consistently associated with negative sentiment, such as "complicated," "waiting," and "crashed," should trigger immediate design reviews of associated functionality. Conversely, positive terminology clusters around concepts like "easy," "fast," and "practical" suggest that emphasizing these attributes in feature development and marketing communications would align with customer values and expectations. Fourth, the complaint pattern analysis suggests that proactive technical status communication could substantially mitigate negative sentiment during service disruptions. Our temporal analysis revealed that negative sentiment intensity increased by 32% when technical issues occurred without proactive notification, compared to just 14% when banks implemented advanced communication. This finding indicates that transparent issue acknowledgment significantly moderates customer frustration during inevitable technical challenges.

6 Conclusion

This research has developed and implemented a systematic methodology for measuring mobile banking service quality through text mining and sentiment analysis of customer reviews from three prominent Turkish private banks. By analyzing unstructured customer feedback data, the study demonstrates that computational text analysis can effectively evaluate service quality dimensions and identify specific areas for improvement in mobile banking applications. The analysis revealed distinct patterns of service quality across institutions, with İşCep demonstrating superior overall performance (0.67), followed closely by Garanti BBVA (0.65) and TEB (0.58). The dimensional analysis framework implemented in this research proves particularly valuable for diagnostic assessment of service quality components. The consistently high performance in the Practicality dimension across all banks (average score: 0.75) indicates that Turkish private banks have successfully optimized the functional efficiency of their mobile applications. However, the uniformly low scores in the Sociality dimension (average score: 0.32) reveal a significant opportunity area for all institutions, suggesting that enhancements in customer communication, personalization, and support accessibility could yield substantial competitive advantages. The integration of sentiment analysis with dimensional quality mapping represents a methodological advancement that transcends traditional survey-based approaches. This integrated methodology offers financial institutions continuous, scalable quality assessment derived from authentic customer expressions, addressing limitations of periodic SERVQUAL implementations. Nevertheless, computational approaches to service quality measurement involve inherent trade-offs: while offering greater scalability and authenticity than survey methods, they sacrifice some degree of structural consistency and controlled comparison. Future methodological refinements should aim to preserve the advantages of computational approaches while enhancing structural comparability across institutions and time periods.

From a theoretical perspective, this study contributes to the understanding of how service quality dimensions manifest in digital banking environments, particularly highlighting the tension between technical functionality and social presence in mobile service contexts. The findings suggest that conventional service quality frameworks require substantial adaptation for digital contexts, with particular emphasis on reconceptualizing how empathy and tangibility dimensions translate to screen-based interactions.

The dimensional analysis framework implemented in this research has proven particularly valuable for diagnostic assessment of service quality components. The consistently high performance in the Practicality dimension across all banks (average score: 0.75) indicates that Turkish private banks have successfully optimized the functional efficiency of their mobile applications, aligning with customer expectations for transactional convenience. However, the uniformly low scores in the Sociality dimension (average score: 0.32) highlight a significant opportunity area for all institutions, suggesting that enhancements in customer communication, personalization, and support accessibility could yield substantial competitive advantages. The methodological approach demonstrated in this study advances service quality measurement by bridging the gap between traditional SERVQUAL frameworks and computational text analysis, establishing a scalable methodology for continuous quality monitoring that overcomes limitations of periodic survey-based approaches.

The topic modeling of customer complaints provided actionable intelligence for service improvement initiatives by identifying four primary complaint categories: technological issues (34.2%), authentication challenges (27.6%), functionality limitations (23.5%), and customer service concerns (14.7%). These complaint patterns reveal specific service failure points requiring targeted remediation, with authentication processes representing a particularly critical interaction domain where security requirements must be carefully balanced with user experience considerations. The integration of dimensional analysis with complaint pattern identification establishes clear priorities for service enhancement, enabling banking institutions to allocate resources strategically toward improvements that will generate maximum impact on customer satisfaction and competitive positioning.

This research makes several significant contributions to both theoretical understanding and practical application of service quality measurement in digital banking environments. Theoretically, the study extends established service quality frameworks through computational methodologies, demonstrating that text mining approaches can effectively capture multidimensional quality aspects while providing enhanced scalability and temporal responsiveness. The identification of dimensional quality patterns specific to the Turkish banking context contributes to contextual refinement of service quality theory, suggesting that dimensional priorities may vary across cultural and institutional environments. From a practical perspective, the research provides banking professionals with a methodological framework for continuous service quality assessment, complemented by specific insights regarding dimensional strengths, improvement opportunities, and complaint remediation priorities for each analyzed institution.

Several limitations warrant consideration when interpreting this study's findings. First, the research corpus was restricted to Google Play Store reviews, potentially introducing selection bias by excluding iOS users and customers who do not post online reviews. Future research should incorporate multiple feedback channels, including iOS reviews, social media comments, and direct customer communications, to ensure comprehensive representation. Second, the sentiment analysis methodology, while robust, cannot fully capture the nuanced emotional content of customer expressions, particularly regarding cultural idioms and contextual references. Advanced natural language processing techniques incorporating emotional valence detection could enhance analytical depth in future studies. Third, the research design did not account for demographic variations in service quality perception, limiting insights regarding age, gender, or socioeconomic influences on dimensional priorities.

6.1 Future Research Directions

This study opens several promising avenues for future research. First, longitudinal studies tracking sentiment evolution through multiple application versions could provide valuable insights into how specific feature implementations affect service quality perceptions over time. Such temporal analysis would help identify whether quality improvements exhibit diminishing returns or whether certain dimensions demonstrate threshold effects where incremental enhancements yield disproportionate sentiment improvements. Second, cross-cultural comparative studies could investigate how service quality dimensions vary across different national banking environments. The Turkish banking sector's specific characteristics—including high smartphone penetration, young demographic profile, and regulatory framework—likely influence customer expectations and quality perceptions. Comparative analysis across diverse markets illuminates how cultural and regulatory contexts shape dimensional priorities in mobile banking services. Third, multimodal sentiment analysis incorporating visual elements (such as interface screenshots and promotional materials) alongside textual feedback could provide more comprehensive understanding of how visual design influences overall service quality perceptions. Given the increasing sophistication of computer vision techniques, integrating visual and textual analysis represents a promising methodological frontier for digital service quality research. Fourth, future research should explore the relationship between mobile banking service quality and broader organizational outcomes, including customer retention, cross-selling success, and financial performance metrics. While this study establishes a methodology for quality measurement, connecting these measures to business outcomes would enhance understanding of the return on investment for service quality improvements in specific dimensions. Finally, as artificial intelligence capabilities continue to evolve in banking applications, research examining how AI-driven features (such as chatbots, personal recommendations, and automated financial advice) affect service quality perceptions would be particularly valuable. The intersection of sentiment analysis and AI-enhanced banking represents an emerging research domain with significant implications for future service design and implementation. As banks increasingly leverage AI to augment customer experiences, understanding the quality implications of these technologies becomes increasingly critical for competitive differentiation in digital banking environments.

References

[1] Aggarwal, N., & Gupta, M. (2003). Multilevel-Multidimensional Model of Banking Service Quality. Paradigm, 7(2), 91-104.
[2] Arcand, M., PromTep, S., Brun, I., & Rajaobelina, L. (2017). Mobile banking service quality and customer relationships. International Journal of Bank Marketing, 35(7), 1068-1089.
[3] Ashkanasy, N. M., & Humphrey, R. H. (2011). Current emotion research in organizational behavior. Emotion Review, 3(2), 214-224. https://doi.org/10.1177/1754073910391684
[4] Balahur, A. (2013). Sentiment analysis in social media texts. Proceedings of the 2013 International Conference on Computational Linguistics and Intelligent Text Processing, 111-125. https://doi.org/10.1007/978-3-642-37256-8_9
[5] Barsade, S. G., & Gibson, D. E. (2007). Why does affect matter in organizations? Academy of Management Perspectives, 21(1), 36-59. https://doi.org/10.5465/amp.2007.24286163
[6] Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. Psychological methods, 11(2), 142.
[7] Beal, D. J., & Trougakos, J. P. (2013). Episodic intra-individual fluctuations in organizational behavior. Journal of Organizational Behavior, 34(5), 733-745. https://doi.org/10.1002/job.1813
[8] Brief, A. P., & Weiss, H. M. (2002). Organizational behavior: Affect in the workplace. Annual Review of Psychology, 53(1), 279-307. https://doi.org/10.1146/annurev.psych.53.100901.135156
[9] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2017). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15-21. https://doi.org/10.1109/MIS.2013.30
[10] Fisher, C. D. (2000). Mood and emotions while working: Missing pieces of job satisfaction? Journal of Organizational Behavior, 21(2), 185-202. https://doi.org/10.1002/(SICI)1099-1379(200003)21:2<185::AID-JOB34>3.0.CO;2-M
[11] Grandey, A. A., Rafaeli, A., Ravid, S., Wirtz, J., & Steiner, D. D. (2013). Emotion in the banking industry: How employee and customer emotions interact. Journal of Service Research, 16(4), 497-510. https://doi.org/10.1177/1094670513490239
[12] Ilies, R., Aw, S. S., & Pluut, H. (2015). Intra-individual models of employee well-being: What have we learned and where do we go from here? European Journal of Work and Organizational Psychology, 24(6), 827-838. https://doi.org/10.1080/1359432X.2015.1071429
[13] Jain, V., Kumar, S., & Lee, J. (2018). Analyzing online customer sentiments for service quality assessment in digital banking. Journal of Business Research, 91, 321-331. https://doi.org/10.1016/j.jbusres.2018.05.032
[14] Judge, T. A., Scott, B. A., & Ilies, R. (2006). Hostility, job attitudes, and workplace deviance: Test of a multilevel model. Journal of Applied Psychology, 91(1), 126-138. https://doi.org/10.1037/0021-9010.91.1.126
[15] Kımıloğlu, H., Aslıhan Nasır, V., & Nasır, S. (2010). Discovering behavioral segments in the mobile phone market. Journal of Consumer Marketing, 27(5), 401-413.
[16] Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of customer feedback using deep learning techniques. Proceedings of the 5th International Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis, 83-90. https://doi.org/10.3115/v1/W14-2609
[17] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016
[18] Martin-Domingo, L., Martín, J. C., & Mandsberg, G. (2019). Social media as a resource for sentiment analysis of Airport Service Quality (ASQ). Journal of Air Transport Management, 78, 106-115.
[19] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135. https://doi.org/10.1561/1500000011
[20] Phan, D. T. (2020). I have seen the future, and it rings-What we know about mobile banking research. THEORY METHODOLOGY PRACTICE: CLUB OF ECONOMICS IN MISKOLC, 16(2), 69-80.
[21] Puriwat, W., & Tripopsakul, S. (2017). The impact of e-service quality on customer satisfaction and loyalty in mobile banking usage: Case study of Thailand. Polish Journal of Management Studies, 15(2), 183-193.
[22] Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and the experience of mood. Current Directions in Psychological Science, 8(1), 10-14. https://doi.org/10.1111/1467-8721.00003
[23] Scarfo, P. (2013). Achieving assured authentication in the digital age. Biometric Technology Today, 2013(9), 9-11.
[24] Suresh, D. M., Vincent, G., Vijai, C., Rajendhiran, M., Com, M., Vidhyalakshmi, A. H., & Natarajan, S. (2024). Analyse Customer Behaviour and Sentiment Using Natural Language Processing (NLP) Techniques to Improve Customer Service and Personalize Banking Experiences. Educational Administration: Theory And Practice, 30(5), 8802-8813.
[25] Weiss, H. M., & Beal, D. J. (2005). Reflections on Affective Events Theory. In N. M. Ashkanasy, W. J. Zerbe, & C. E. J. Härtel (Eds.), The effect of affect in organizational settings (pp. 1-21). Emerald Group Publishing. https://doi.org/10.1016/S1746-9791(05)01101-6
[26] Zhang, Y., Zhang, L., & Vo, K. (2020). The role of sentiment analysis in predictive analytics for customer experience management. Journal of Marketing Analytics, 8(3), 214-228. https://doi.org/10.1057/s41270-020-00088-4