Analyzing User Experience in Mobile Banking Applications Through Text Mining

Bora Lamaj (Myrto)[1], Markela Muça[2], Klodiana Bani[3]
Faculty of Information Technology "Aleksander Moisiu"-Department of Computer Sciences, University of Durres, Albania[1]
Faculty of Natural Sciences-Department of Applied Mathematics, University of Tirana, Albania[2, 3]

***Abstract*—This study attempts to offer a data-driven comprehension of the factors that influence satisfaction and dissatisfaction by examining user reviews, based on a banking mobile application in Albania, specifically the Raiffeisen Bank application. All gathered reviews from Google Play and the App Store have been classified using the multilingual BERT model. It is important to mention that mBERT is particularly suitable given that the reviews are written in both Albanian and English. BERTopic has been used to determine the main topics of classification. After classifying the reviews and creating new dimensions with BERTopic, an ordinal logistic regression model was applied to the generated dataset to assess the predictive power of these dimensions and sentiment polarity on user-assigned satisfaction ratings. This study helps developers and managers to better understand the main factors that influence customer reactions and evaluate the factors that influence the rate of the application and the different bugs they report.**

Keywords—Generated reviews; sentiment analysis; topic modeling; BERTopic; mBERT

I. Introduction

As mobile banking applications increasingly replace traditional branch-based services, user experience has become a critical determinant of customer retention and satisfaction in the banking sector. Beyond serving as transactional platforms, mobile banking applications now present a service delivery mechanism and a point through which banks communicate their value proposition to customers. In this environment, users continuously generate reviews on platforms such as Google Play and Apple Store, which are a source of feedback that indicates user experiences. Unlike controlled survey instruments, these reviews capture real perceptions and recommendations expressed in natural language. The existing literature presents a lack of studies on Albanian sentiment that integrate sentiment analysis, topic modeling, and statistical modeling in a unified framework to analyze the impact of user experience dimensions on satisfaction in our case rating. Most prior research focuses either on sentiment classification or topic extraction in isolation, without systematically linking these dimensions to user satisfaction outcomes. This represents a clear research gap in understanding how different aspects of mobile banking experiences translate into user ratings. This study aims to address this gap in the literature concerning the identification and quantification of experiential dimensions that impact mobile banking application reviews and their relationship to satisfaction levels, which is reflected in star ratings. The research focuses on a second-level commercial bank that operates in Albania. This study focuses on giving answers to the following questions: 1) How does sentiment polarity influence user satisfaction ratings? 2) Which experience dimensions have the strongest impacts on user sentiment and ratings? and 3) Does negative user sentiment exert a stronger effect on ratings than positive sentiment? Based on these questions, the following hypotheses are composed:

H1: Sentiment polarity has a positive effect on user satisfaction ratings.

H2: Technical Issues negatively affect user satisfaction ratings.

H3: Negative reviews have a more significant impact on ratings than positive reviews.

The study contributes to the existing literature in three aspects: First, it applies a state-of-the-art NLP pipeline, combining it with multilingual BERT sentiment analysis and with BER Topic modeling to a real banking dataset. Second, it quantifies the influence of extracted experience dimensions on user ratings through regression analysis [16]. Third, it demonstrates the scalability and interpretability of this approach for continuous reputation monitoring in practical banking management.

The remainder of this study is organized as follows: Section II describes the relevant literature on sentiment analysis, topic modeling, and ordinal approaches to modeling user satisfaction in mobile banking applications. Section III contains details regarding research methodology, including data collection and preprocessing, sentiment classification, BERTopic-based topic modeling, dimensional assignment, and multicollinearity diagnostics. Section IV presents the empirical results, covering sentiment distribution, the identified experience dimensions, and the ordinal logistic regression findings, including a test of H3. Section V discusses these findings in relation to prior literature and outlines the study's limitations. Finally, Section VI presents the conclusions, and Section VII outlines directions for future research.

II. Literature Review

The application of transformer-based models to mobile banking reviews has gained considerable traction in recent years. Regarding topic modeling and definition, [9] provided comparison between BERTopic, Latent Dirichlet Allocation LDA and other probabilistic models [12], [18]. Their results showed that BERTopic produces higher semantic discriminability. These findings were supported also by [2], [5] who validated BERTopic in marketing reviews [8], [10], [20].

The use of topic modeling to understand mobile banking user experience has been explored by several recent studies. The authors [19] applied BERTopic to app store reviews in health data, demonstrating that the framework successfully extracts clinical and meaningful topics from multilingual data. Collectively, these papers suggest that topic modeling pipelines applied to app reviews can produce interpretable dimensions of user experience across different service domains. The existing multilingual BERT (mBERT) model will be suitable for our study by considering that the reviews are in two different languages, English and Albanian. Also, BERTopic was preferred over LDA because BERTopic offers a superior semantic understanding using transformer-based embeddings (BERT), while LDA relies on word co-occurrences. BERTopic is more accurate, and requires less preprocessing and performs better in short text environments such as app reviews [11], [15]. Paper [17] analyzed mobile applications reviews of three countries Bosnia-Herzegovina, Croatia, and Serbia, combining BERT-based with BERTopic modelling and multiple linear regression. It demonstrated that sentiment polarity and technical issues were the strongest predictors of ratings, explaining 58% of the variance. In [9], the authors applied a pipeline, regarding an Indonesian mobile banking application, using sentiment analysis and topic modelling on over 7,000 reviews on Google Play [4]. Their study confirmed that technical reliability and ease of use consistently were the most important dimensions in shaping user perception. Another study by [2] on Canadian banking applications, where reviews were taken from App Store and Google Play, demonstrated login and update errors constituted the primary driver of negative ratings, while usability and feature richness drove positive sentiment.

The present study follows the methodological tradition established by these works but departs from them in one substantive respect. Whereas all reviewed studies treat user ratings as a continuous dependent variable and apply multiple linear regression, this study recognizes that star ratings on a five-point scale constitute an ordinal rather than an interval-level outcome. The distances between consecutive rating categories are not guaranteed to be perceptually or behaviorally equal. A move from one to two stars does not necessarily represent the same magnitude of dissatisfaction as a move from four to five stars. Treating such a scale as continuous violates the assumptions of OLS regression and can produce biased coefficient estimates and misleading inference [2]. To address this, the present study employs Ordinal Logistic Regression with the proportional odds specification, estimating four threshold parameters that separate adjacent rating categories and interpreting predictor effects as changes in the log-odds of falling in a higher rather than lower category. The validity of the proportional odds assumption is confirmed via the Brant test (all $p > .97$), and model fit is evaluated using McFadden R^2 rather than the OLS coefficient of determination. This choice constitutes a methodologically more rigorous treatment of the dependent variable and yields results that are directly interpretable in terms of the ordinal structure of user satisfaction.

III. Framework Methodology

This study integrates natural language processing techniques with regression model to examine how user experience dimensions received from BERTopic provide an authentic feedback reflecting real user experiences and perceptions.

A. Sample Definition and Data Collection

User reviews were collected from the official mobile application of Raiffeisen Bank available on Google Play and App Store. The dataset includes reviews posted from September 2021 to December 2023. A total of 2,764 reviews were retrieved where only 1,000 contained text review, while the remaining 1,764 consisted rating only, with no accompanying text, and were therefore excluded from text mining. The dataset contains the following attributes:

- Review text (raw user comment)
- User rating (1-5 stars)
- Review timestamp
- Review identifier

The collected data presents a wide range of user experiences and reflects natural variability in rating behavior, language usage, and review length characteristics that support robust text mining and sentimental analysis.

B. Data Preprocessing

Before analysis, on all the reviews the preprocessing pipeline is followed designed to reduce noise and normalize textual data [3]. The preprocessing steps include the conversion of text to lowercase, removal of punctuation and special characters, elimination of duplicate entries, and removal of empty or uninformative reviews. These procedures improve textual consistency and enhance the quality of downstream NLP tasks.

C. Sentiment Analysis

Sentiment classification was performed using multilingual BERT sentiment analysis model. The multilingual was necessary because the dataset contains reviews written in both English and Albanian. The multilingual BERT-based sentiment model nlptown/bert-base-multilingual-uncased-sentiment was applied to the review text via the Hugging Face pipeline function in its pretrained, off-the-shelf form, without additional fine-tuning. For each review, the model outputs a predicted rate on a five-point scale (1 to 5), reflecting the intensity of the sentiment expressed in the text. This five-class output was then collapsed into three sentiment categories using a rate-based threshold scheme: reviews with a predicted rate of 4 or 5 were classified as Positive, a predicted rate of 3 as Neutral, and a predicted rate of 1 or 2 as Negative.

D. Topic Modeling with BERTopic

Following sentiment classification, topic classification was conducted using BERTopic due to its high-quality, context-aware topic representations, flexibility, and out-of-the-box usability. The BERTopic framework combines sentence transformer embeddings with UMAP dimensionality reduction [14], [6] HDBSCAN clustering, and class-based TF-IDF (c-TF-IDF) [13] to identify interpretable latent topics with the review corpus [1]. Topics are interpreted through examination of the most representative keywords and qualitative inspection of review samples. Based on this analysis, three primary user experience dimensions were identified and labeled.

The BERTopic model was configured as follows: Document embeddings were generated using paraphrase-multilingual-MiniLM-L12-v2, the default sentence-transformer activated by BERTopic's language="multilingual" setting and selected for its support of both English and Albanian. Dimensionality reduction was performed with UMAP (n_components = 5, n_neighbors = 15, min_dist = 0.0, random_state = 42, metric = "euclidean", the library default), followed by clustering with HDBSCAN (min_cluster_size = 10, metric = "euclidean", prediction_data = True; min_samples and cluster_selection_method were left at their default values of min_cluster_size and "eom", respectively). Topic representations were derived using class-based TF-IDF (c-TF-IDF) over the resulting HDBSCAN clusters, with calculate_probabilities = True enabled to obtain a per-document topic-probability distribution for each review. The initial topic set was then reduced to three interpretable dimensions via BERTopic's built-in topic reduction (nr_topics = 3), followed by qualitative keyword inspection and manual labeling of each dimension, as described in Section III D. The target of three dimensions was chosen to balance interpretability with the constraints of the regression sample size (N = 1,000), and aligns with the broad usability, reliability, and satisfaction constructs recurrent in prior mobile banking UX literature [9].

E. Dimensional Assignment

The BERTopic model generates a probability distribution across topics for each review. For analytical purposes, the three topics with the highest probability values for each review were retained and labeled as primary, secondary, and tertiary topic assignments [7]. Each experience dimension was indicated as a continuous variable, presenting the degree to which a review is associated with a given topic. This probabilistic representation allows each review to contribute proportionally to multiple experience dimensions, thereby capturing the multidimensional nature of user experiences.

F. Multicollinearity Diagnostics

Prior to conducting the regression analysis, multicollinearity among the independent variables was assessed using Spearman correlation analysis and the Variance Inflation Factor (VIF) [22]. Spearman correlation measures the strength and direction of the monotonic relationship between two variables (Fig. 1).

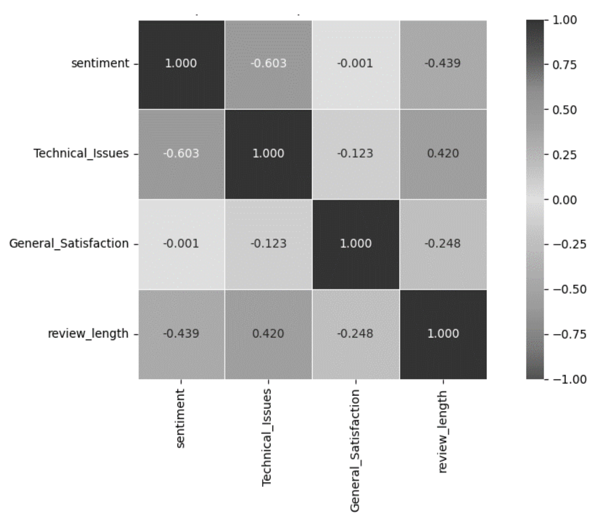

Fig. 1. Spearman's correlation matrix

Rather than operating on raw values, it converts observations to ranks and computes the correlation over those ranks. Proposed by Charles Spearman in 1904, it is the non-parametric counterpart of Pearson's r:

$$\rho = 1 - (6 \cdot \Sigma di^2) / (n \cdot (n^2 - 1)) \quad (1)$$

To further evaluate predictor independence, Variance Inflation Factor (VIF) values were calculated for all variables which are included in the regression model:

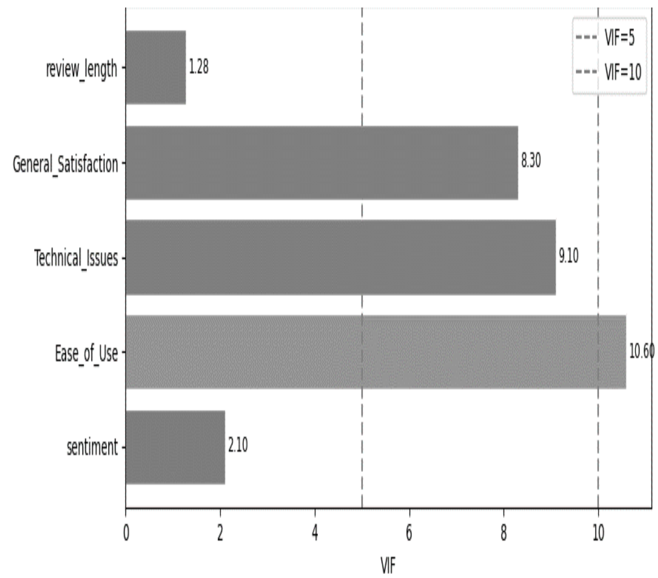

Fig. 2. VIF per predictor

$$VIF_i = \frac{1}{1-{R_i}^2} \quad (2)$$

where, R_i^2 denotes the coefficient of determination obtained by regressing predictor i on all remaining predictors.

The elevated VIF values (Fig. 2) observed for the topic probability variables (Ease_of_Use = 10.60, Technical_Issues = 9.10, General_Satisfaction = 8.30) do not represent empirical multicollinearity requiring remediation, but rather reflect a structural mathematical property of compositional data. Since BERTopic topic probabilities sum to unity for each document, their simultaneous inclusion in a regression model is formally equivalent to including a complete set of dummy variables representing a categorical variable a configuration in which high VIFs are an expected and mathematically inevitable outcome. As [2] explicitly notes, when high VIFs arise from indicator variables representing a categorical variable with three or more categories, the problem can be safely ignored: the overall model fit, the significance of other predictors, and the F-statistic are entirely unaffected. Consistent with this recommendation, Ease_of_Use the topic with the highest frequency of assignment and consequently the largest contribution to the linear constraint was designated as the reference category and excluded from the regression specification. Following this adjustment, all remaining VIF values fell below 2.20 (Fig. 3), well within the acceptable threshold of 5.0 [21], confirming that no genuine multicollinearity is present.

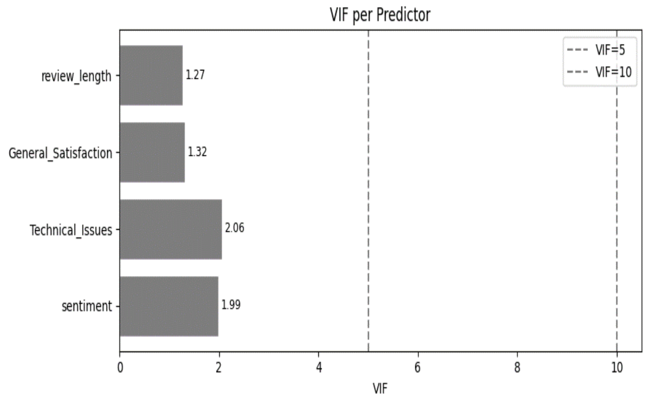

Fig. 3. Variance Inflation Factors (VIF) for the final model predictors.

Overall, the diagnostic results demonstrate that sentiment-based and topic-derived variables capture distinct yet complementary dimensions of user experience, supporting their simultaneous inclusion in the regression analysis.

IV. RESULTS

A. Dataset Overview

The final dataset comprised 1,000 user reviews collected from the Raiffeisen Bank Mobile banking application. Each review included an associated user rating ranging from 1 to 5 stars.

B. Sentiment Analysis Results

BERT-based sentiment classification indicated that 46.9% of reviews were positive, 33.4% were negative, and 19.7% neutral (Fig. 4). Overall, positive reviews represented the dominant sentiment category within the dataset. Most positive reviews were associated with ratings of 4 or 5 stars.

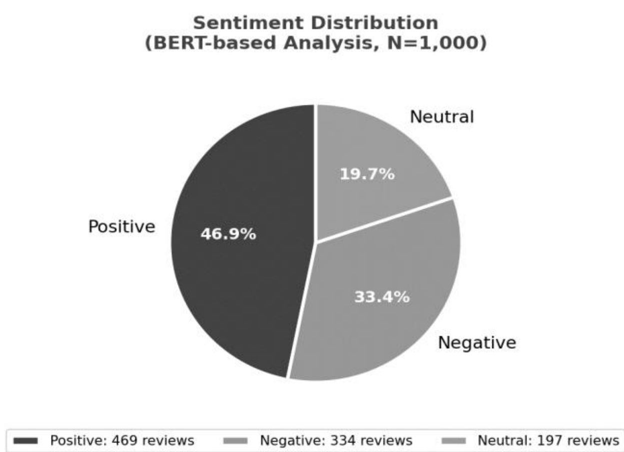

Fig. 4. Sentiment distribution based on multilingual BERT-based classification (N = 1,000).

C. Topic Modeling and Experience Dimensions

Neutral reviews present only 19.7% of the dataset (197 of 1,000), positive reviews 46.9% and 33.4% Negative reviews, leaving the model few examples on which to calibrate the neutral boundary. Users tend to express clear satisfaction or dissatisfaction, so the “3-star” rating reflects mixed rather than genuinely neutral sentiment, making it difficult to separate from Positive/Negative classes. Although the three-category sentiment scheme is retained despite the weak Neutral-class metrics, since it preserves a direct conceptual link to the five-point rating scale and underlies Sentiment Polarity, the ordinal predictor (Negative < Neutral < Positive) is used in the regression analysis. For the Positive and Negative classes, which dominate the dataset and are most relevant to the study's hypotheses, the weighted F1-score (0.775) and overall accuracy (0.742) remain acceptable (Table I).

TABLE I. CLASSIFICATION METRICS

Class	Precision	Recall	F1-score	Proxy labels (rating-based)
Negative	0.65	0.87	**0.744**	Rating 1–2
Neutral	0.03	0.12	**0.053**	Rating 3
Positive	0.97	0.71	**0.819**	Rating 4–5
Macro avg.	0.55	0.57	0.539	—
Weighted avg.	**0.85**	**0.74**	**0.775**	—
Accuracy	**0.742**	—	—	

BERTopic identified three dimensions from the review corpus. Table II summarizes the topics, their labels, and keywords.

TABLE II. USER EXPERIENCE DIMENSIONS IDENTIFIED THROUGH BERTOPIC.

Topic	Label	Description & Representative Keywords
Topic 0	Ease of Use	Reviews discussing usability, interface design, and overall app navigation. Keywords: app, easy, great, use, friendly, features, interface
Topic 1	Technical Issues	Reviews reporting application errors, login failures, and update problems. Keywords: nuk hapet, update, problem, crash, login, error, slow
Topic 2	General Satisfaction	Short affirmative reviews expressing overall positive impressions. Keywords: excellent, shume mire, good, nice, perfect, satisfied

Topic 1, Technical Issues, was dominated by Albanian-language and describes the application problem, like crashing, doesn’t open, update error, etc. The keyword 'nuk hapet' (meaning 'does not open') is highly characteristic of this cluster, indicating a post-update instability.

Fig. 5 illustrates the relationship between identified topics and sentiment categories.

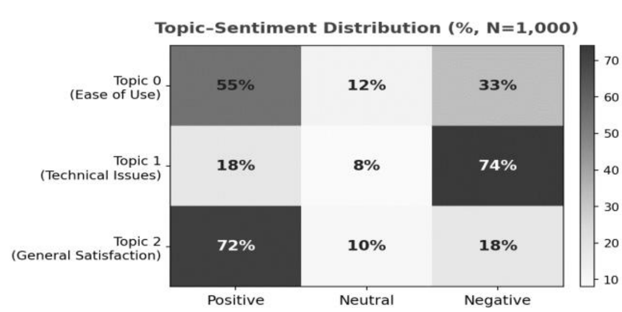

Fig. 5. Heatmap of the topic–sentiment distribution (%) across the three BERTopic-derived experience dimensions.

Topic 0, Ease of Use, is associated with positive sentiment (55%), while Topic 1, Technical Issues with negative sentiment (74%). Topic 2, General Satisfaction, was also largely linked to positive sentiment (72%).

D. Regression Analysis

The ordinal logistic regression initially included four standardized predictors: Sentiment Polarity (coded as Negative = −1, Neutral = 0, Positive = +1), Technical_Issues, General_Satisfaction, and review length, with Ease_of_Use excluded a priori as the reference category due to the compositional constraint among topic probabilities (Section III E). Backward elimination iteratively removed the predictor with the smallest absolute coefficient whenever doing so did not increase 5-fold cross-validated MAE by more than 0.01, eliminating General_Satisfaction and review length and retaining Sentiment Polarity and Technical_Issues as the final predictors (Table III). After applying the ordinal logistic regression, backward technique, the results indicate that there are only two important variables that determine the rating of each text review, especially Sentiment Polarity and Technical issues.

The regression equation is equal to:

$$\log\left(\frac{\mathrm{P}(\mathrm{Y} \leq \mathrm{k})}{\mathrm{P}(\mathrm{Y} > \mathrm{k})}\right) = \theta k + 1.108 \cdot Sentiment - 1.187 \cdot Technical\_Issues \quad (3)$$

TABLE III. REGRESSION EQUATION

Variable	Coef. (β)	SE	z-value	p-value	OR
Sentiment Polarity	+1.108	0.207	+5.360	< 0.001	3.027
Technical Issues	−1.187	0.180	−6.581	< 0.001	0.305

Table III shows the results of background elimination, indicating a p-value less than 0.001 and McFadden $R^2 = 0.350$, which indicates that the model is good.

where, θ_k is equal to (see Table IV):

TABLE IV. THRESHOLD TABLE

θ_k	Transition	θk	SE	z	p
θ_1	1| 2★	-1.97	0.23	-8.47	< .001
θ_2	2★ | 3★	-1.32	0.34	-3.82	< .001
θ_3	3★ | 4★	-1.17	0.34	-3.39	.001
θ_4	4★ | 5★	-0.83	0.29	-2.85	.004
—	5★ (derived)	—	—	—	—

Table IV presents the threshold parameters (θk) of the Ordinal Logistic Regression model, which define the boundaries between the five ordered rating categories, all statistically significant ($p < .01$), with the required monotonic ordering confirmed ($\theta_1 < \theta_2 < \theta_3 < \theta_4$) and a predicted probability of 69.8% for the highest rating category at the sample mean.

To test H3, sentiment polarity was divided into two categorical predictors, Sentiment_Negative and Sentiment_Positive, and the ordinal logistic regression was re-estimated alongside Technical_Issues. Both predictors were individually significant.

- Sentiment_Negative (β = −0.993, SE = 0.484, p = .040, OR = 0.370)
- Sentiment_Positive (β = +1.428, SE = 0.557, p = .010, OR = 4.168)

The results confirm that negative and positive sentiment shift the odds of a higher rating in the expected directions. However, a Wald test of the equality of these coefficients in magnitude (H0: |β_Negative| = |β_Positive|) was not significant, $\chi^2(1) = 0.214$, $p = .644$. Contrary to H3, the point estimates suggest a marginally larger effect for positive sentiment than for negative sentiment, though this difference is not statistically distinguishable from zero. H3 is therefore not supported. Sentiment polarity remains a strong overall predictor of ratings, but no significant asymmetry between positive and negative sentiment effects was detected.

It is also valuable to mention that reviews of 2025 result to have 90% positive sentiment (Fig. 6) and favorable evaluations of the application's usability and performance. It is slightly easier to notice that the clients ask for the integration of Apple Pay functionality, which is being integrated in many bank applications lately.

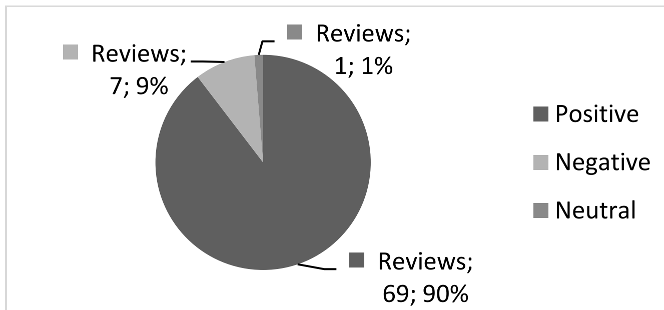

Fig. 6. Sentiment distribution (N = 77 reviews, year 2025).

V. DISCUSSION

The most closely related paper is [17], which applied BERT-based sentiment classification and BERTopic topic over 56,000 mobile banking reviews using multiple linear regression to evaluate the effect of extracted experience dimensions with BERTopic on user ratings. The paper revealed that sentiment polarity and technical performance explained 58% of the variance in ratings. This study, rather than treating the five-point star rating as continuous interval-level, employs Ordinal Logistic Regression with a proportional odds specification. This choice reflects the ordinal nature of user ratings more.

A. Limitations

A limitation of this classification is the weak performance on the Neutral category (F1 = 0.053), caused by class imbalance and the linguistic ambiguity of short, equivocal reviews. Future work could adopt a binary positive/negative classification or a dedicated classifier for neutral classes.

VI. CONCLUSION

This study demonstrates that the pipe, which implicates the integration of multilingual BERT sentiment analysis, BERTopic modeling, and regression analysis, provides an effective pipeline for identifying the main factors that have an impact on user reactions in banking applications. 1,000 user reviews from the Raiffeisen Bank mobile application identified three distinct experience dimensions: Ease of Use, Technical Issues, and General Satisfaction through BERTopic modeling. Of these, Sentiment Polarity ($\beta = +1.108$, OR = 3.027) and Technical Issues ($\beta = -1.187$, OR = 0.305) emerged as the only statistically significant predictors of user ratings ($p < 0.001$), with the model achieving a McFadden R^2 of 0.350, indicating strong explanatory power. The results suggest that banks should give priority in handling bug solutions, update stability, and real-time sentiment monitoring in order to maintain a very good service. Regarding the study's hypotheses, H1 and H2 are supported by the main model (Sentiment Polarity, Technical_Issues; Table III). H3, however, is not supported: the supplementary analysis decomposing sentiment polarity into Sentiment_Negative and Sentiment_Positive (Section IV D) found no statistically significant difference in the magnitude of their effects on ratings (Wald $\chi^2(1) = 0.214$, $p = .644$).

Related to future work, it will be the focus to extend the case studies in different countries in order to examine the client approach and bank approach to sentiment analysis.

VII. FUTURE WORK

Several directions will be the focus of the following research. First, extend the analytical framework to other commercial banks operating in the Albanian market: BKT, Credins, and OTP Bank Albania. Second, applied the same pipeline to Raiffeisen Bank reviews collected across Eastern Europe, enabling a cross-country comparative analysis that isolates market-level and linguistic factors from those attributable to the institution itself. Collectively, these directions would transform the present single-institution, single-country study into a regionally representative framework for monitoring mobile banking user experience in Southeast Europe.

REFERENCES

[1] Alrizq, M., & Alghamdi, A. (2024). Customer satisfaction analysis with Saudi Arabia mobile banking apps: A hybrid approach using text mining and predictive learning techniques. Neural Computing and Applications, 36, 6005–6023. https://doi.org/10.1007/s00521-023-09400-4

[2] Allison, P. D. (2012, September 10). When can you safely ignore multicollinearity? Statistical Horizons. https://statisticalhorizons.com/multicollinearity/

[3] Agresti, A. (2010). Analysis of ordinal categorical data (2nd ed.). John Wiley & Sons. https://doi.org/10.1002/9780470594001 ISBN: 978-0-470-08289-8

[4] Amirkhalili, Y., & Wong, H. Y. (2025). Banking on feedback: Text analysis of mobile banking iOS and Google app reviews. arXiv preprint arXiv:2503.11861. https://doi.org/10.48550/arXiv.2503.11861

[5] An, Y., Oh, H., & Lee, J. (2023). Marketing insights from reviews using topic modeling with BERTopic and deep clustering network. Applied Sciences, 13(16), 9443. https://doi.org/10.3390/app13169443

[6] Bakïasi, V., Muça, M., & Kapçiu, R. (2024). Dimensionality reduction: A comparative review using RBM, KPCA, and t-SNE for micro-expressions recognition. International Journal of Advanced Computer Science and Applications, 15(1). https://doi.org/10.14569/ijacsa.2024.0150135

[7] Basu, B., Sebastian, M. P., & Kar, A. K. (2024). What affects the promoting intention of mobile banking services? Insights from mining consumer reviews. Journal of Retailing and Consumer Services, 77, 103695. https://doi.org/10.1016/j.jretconser.2023.103695

[8] Çallı, L. (2023). Exploring mobile banking adoption and service quality features through user-generated content: The application of a topic modeling approach to Google Play Store reviews. International Journal of Bank Marketing, 41(2), 428–454. https://doi.org/10.1108/IJBM-08-2022-0351

[9] Desiraju, K., Mishra, A. N., & Sengupta, P. (2024). Customer perceptions on open banking apps: Insights using structural topic modeling. Journal of Retailing and Consumer Services, 81, 104029. https://doi.org/10.1016/j.jretconser.2024.104029

[10] Dey, M., Islam, M. Z., & Rana, T. (2023). Applying text mining to understand customer perception of mobile banking app. In T. Rana et al. (Eds.), Handbook of Big Data and Analytics in Accounting and Auditing (pp. 309–333). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-4460-4_14

[11] Edwina, & Mauritsius, T. (2024). Data-driven insights for mobile banking app improvement: A sentiment analysis and topic modelling approach for SimobiPlus user reviews. International Journal of Engineering Trends and Technology, 72(6), 347–360. https://doi.org/10.14445/22315381/IJETT-V72I6P132

[12] Egger, R., & Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. Frontiers in Sociology, 7, Article 886498. https://doi.org/10.3389/fsoc.2022.886498

[13] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794. https://doi.org/10.48550/arXiv.2203.05794

[14] Huertas-García, Á., Martín, A., Huertas-Tato, J., & Camacho, D. (2023). Exploring dimensionality reduction techniques in multilingual transformers. Cognitive Computation, 15(2), 590–612. https://doi.org/10.1007/s12559-022-10066-8

[15] Karanikola, A., Davrazos, G., Liapis, C. M., & Kotsiantis, S. (2023). Financial sentiment analysis: Classic methods vs. deep learning models. Intelligent Decision Technologies, 17(4), 1125–1148. https://doi.org/10.3233/IDT-230478

[16] Lamaj (Myrto), B., Muça, M., & Shtino, V. B. (2024). Comparative study of Naïve Bayes and Logistic Regression for text classification techniques. 5th International Eurasian Conference on Science, Engineering and Technology (EurasianSciEnTech 2024), June 26–28, 2024.

[17] Mahmutović, K. (2025). Analyzing user-generated reviews to identify experience dimensions and their impact on satisfaction with mobile banking applications. Tehnički Vjesnik / Technical Gazette, 38(2), 283–297. https://hrcak.srce.hr/file/494044

[18] Mishra, M. (2025). A holistic review of customer experience research: Topic modelling using BERTopic. Marketing Intelligence & Planning, 43(4), 802–820. https://doi.org/10.1108/MIP-09-2024-0576

[19] Rahman, M., Yee, H. P., Masud, M. A. K., & Uzir, M. U. H. (2024). Examining the dynamics of mobile banking app adoption during the COVID-19 pandemic: A digital shift in the crisis. Digital Business, 4(2), 100088. https://doi.org/10.1016/j.digbus.2024.100088

[20] Sebayang, T. E., Indrawan, D., Bakhtiar, T., & Hakim, D. B. (2023). What accelerates the choice of mobile banking for digital banks in Indonesia? Journal of Risk and Financial Management, 17(1), 6. https://doi.org/10.3390/jrfm17010006

[21] Uncovska, M., Freitag, B., Meister, S., & Fehring, L. (2023). Rating analysis and BERTopic modeling of consumer versus regulated mHealth app reviews in Germany. npj Digital Medicine, 6, 115. https://doi.org/10.1038/s41746-023-00862-3

[22] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage Learning.ISBN: 978-1-4737-5654-0