

A Comprehensive Analysis of UI/UX Metrics for Bangladeshi E-commerce Mobile Applications

Tasfia Barshat
190042126

BSc in Software Engineering
tasfiabarshat@iut-dhaka.edu

Suraiya Hasan
190042107

BSc in Software Engineering
suraiyahasan@iut-dhaka.edu

Abstract—This paper presents a comprehensive analysis of UI/UX metrics for optimizing Bangladeshi mobile e-commerce applications, focusing on usability, responsiveness, visual aesthetics, and user engagement. We move beyond traditional click-through and conversion rate metrics to explore a multi-faceted approach for evaluating and enhancing the user experience. Bangladeshi e-commerce is skyrocketing, with Daraz, Chaldal, Bikroy, and Rokomari leading a wave of user growth that’s reshaping retail and empowering lives. The exponential growth in E-commerce app users in Bangladesh underscores the burgeoning demand for robust user experience (UX) and app usability within the online business domain. Studies have demonstrated a strong correlation between positive UX and continued application usage, resulting in increased customer satisfaction. Similarly, app usability has been shown to directly contribute to enhanced user satisfaction.

Index Terms—User experience, Usability, E-commerce, User experience questionnaire, System usability scale, Sentiment analysis, E-commerce application Reviews, Text analysis, RoBERTa, DistilBERT

I. INTRODUCTION

In tandem with the exponential growth of digitalization in Bangladesh, where approximately 50 million people access the internet daily [8], the trajectory of e-commerce has been nothing short of remarkable. This surge has propelled a substantial portion of the population toward the convenience of mobile applications for their shopping needs. The dynamic landscape of e-commerce, marked by the prominence of platforms like Daraz [4], Chaldal [7], Rokomari [5], and Bikroy [6], underscores the transformative role of technology in reshaping consumer interactions with businesses [8].

Our research delves into the intricate realm of Bangladeshi e-commerce mobile applications, recognizing the essential role played by User Interface (UI) and User Experience (UX) in shaping user behavior, satisfaction, and ultimately, the success of these ventures.

User Interface (UI) and User Experience (UX) metrics play a crucial role in shaping the success and acceptance of applications, websites, and other interactive platforms. UI focuses on the visual elements and design aspects, ensuring an aesthetically pleasing and intuitively navigable interface. On the other hand, UX extends beyond the surface, encompassing the overall experience users have while interacting with a product, emphasizing usability, efficiency, and user satisfaction.

As we navigate through this comprehensive analysis, our focus converges on four essential metrics—Usability, Responsiveness, Visual Aesthetics, and User Engagement of Bangladeshi E-commerce mobile applications. Usability takes center stage, evaluating the efficiency of application navigation to ensure a user-friendly and intuitive experience. The metric of Responsiveness scrutinizes the adaptability of these applications across a number of devices and network conditions, acknowledging the diverse technological landscape of the user base. Visual Aesthetics examines the overarching design elements, recognizing the profound impact of aesthetics on user perception and engagement. Finally, User Engagement becomes a focal point, assessing interactive elements, feedback mechanisms, and features designed to foster a deeper connection and loyalty among users.

By dissecting these key aspects, our research seeks to unravel in-depth insights that extend beyond conventional metrics, aiming to contribute substantively to the optimization of user experiences within the ever-evolving digital commerce landscape in Bangladesh.

II. METHODOLOGY

A. Survey Design and Distribution

To comprehensively assess the user experience (UX) landscape of Bangladeshi e-commerce mobile applications, we designed a survey that leveraged the established frameworks of the User Experience Questionnaire (UEQ) and the System Usability Scale (SUS). This methodical approach ensured a standardized and multifaceted evaluation, enabling a deep dive into four crucial UX metrics: usability, responsiveness, visual aesthetics, and user engagement.

Participants, constituting a representative sample of Bangladeshi e-commerce users, engaged with meticulously crafted questions aligned with the UEQ’s six sub-scales (perceptibility, dependability, efficiency, stimulation, novelty, and attractiveness) and the SUS’s ten concise statements. This nuanced design yielded not only quantitative data but also unearthed qualitative gems through open-ended inquiries, granting users a platform to voice their unfiltered experiences and aspirations. [1]

1. How easy is it for you to navigate through different sections of the app?
2. How easy is it to understand the information about products on these pages?
3. How would you rate the checkout process in terms of simplicity and efficiency?
4. How satisfied are you with the responsiveness of the app to your actions?
5. To what extent do you find the app visually appealing?
6. To what extent does the app's design align with your expectations as an e-commerce user?
7. How satisfied are you with the visual organization and categorization of products on the app?
8. How likely are you to recommend this e-commerce app based on your overall user experience?
9. To what extent does the app provide a personalized and tailored shopping experience for you?
10. How likely are you to revisit and use the app for future purchases?
11. How satisfied are you with the customer support features provided within the app?
12. How satisfied are you with the security measures implemented in the app, especially during the checkout process?
13. How would you rate the overall application?

Fig. 1: Questions asked for survey



Fig. 2: Sacle

B. Webscraping Google Playstore Reviews

Complementing the insights generated by our meticulously designed survey, the second step of our methodology involved mining another valuable data source: user reviews and rating scores on the Google Playstore. This endeavor aimed to not only validate the findings of our survey but also to unveil additional aspects and perspectives not captured by closed-ended questionnaire responses.

Adhering to ethical scraping practices and respecting platform terms of service, we utilized an established library called Google Play Scraper to navigate the Google Playstore ecosystem. Through this process, we meticulously curated a robust dataset encompassing diverse viewpoints, experiences, and sentiments expressed by users, alongside corresponding rating scores providing quantitative signals of overall satisfaction.

By seamlessly bridging the gap between the structured survey data and the unfiltered real-world expressions captured through webscraping, this comprehensive strategy promises a detailed description of the UX environment in Bangladeshi e-commerce.

C. Data Collection and Preprocessing

1) *Data Collection:* Our exploration of Bangladeshi e-commerce user experiences involved collecting a sizeable dataset of 37,245 reviews from four prominent apps: Daraz, Chaldal, Rokomari, and Bikroy. Leveraging the Google Play Scraper library, we focused on publicly available reviews and corresponding rating scores for each app. While Daraz and Chaldal contributed 3,429 and 3,997 reviews respectively, Bikroy significantly boosted the dataset with 24,952 reviews, followed by Rokomari with 4,867. This data, sourced exclusively from public user accounts, was collected in separate CSV files for each app and subsequently analyzed.

The CSV files comprehensively captured key review aspects, including unique identifiers, reviewer names and images, review text, thumbs-up counts, app versions, reply content, and rating values on a 1-5 scale, as shown in Figure 3. Based on this rich dataset, Figure 7 was generated to provide a visual overview of how review distribution aligns with different rating categories.

Fig. 3: Scraped Google Playstore Reviews

2) *Data Preprocessing:* Following the data collection stage, a meticulous data cleaning and preprocessing phase was undertaken to ensure the integrity and analytical readiness of the webscraped user reviews and rating scores. This crucial step aimed to establish a cohesive and language-agnostic corpus for subsequent analysis. To achieve this, translation techniques were employed to render all user reviews into English, facilitating consistent and standardized interpretation. Additionally, all emojis were systematically removed to mitigate potential noise and ensure textual homogeneity, thereby minimizing confounding factors and enhancing the validity and replicability of subsequent analyses. Through this rigorous data refinement approach, we established a robust foundation for extracting meaningful insights from the collected data.

reviewId	score	replyContent	userEngagement	Usability	processedContent
0	5		0	Positive	Alhamdulillah best app
1	5	Thank you for providing your valuable feedback...	1	Positive	best app for books...
2	5	It is delightful to hear such positive words a...	1	Positive	Trustful institution.
3	5	Thank you for your review. Enjoy reading & keep...	1	Positive	Very helpful for book lovers
4	5	We would like to thank you for using our online...	1	Positive	Great App's

TABLE I: After Processing Data

D. Data Exploration and Visualization

To facilitate a quantitative exploration of usability and user engagement within the Bangladeshi e-commerce UX landscape, we systematically enriched the dataset with two novel metrics during the data exploration and visualization stage. These metrics were meticulously crafted to capture key dimensions of user experience, defined as follows:

1) Usability

- Ratings scores between 0-2 were classified as representing "negative usability."
- A rating of 3 was designated as "neutral usability."
- Ratings scores between 4-5 signified "positive usability."

2) User Engagement

- The presence of reply content within a review was indicative of active user engagement, resulting in a value of 1 for this metric.
- Absence of reply content, on the other hand, mapped to a value of 0, suggesting a more passive user experience.

To illuminate the interplay of survey responses and webscraped reviews within the Bangladeshi e-commerce UX landscape, we crafted a comprehensive suite of visual graphs

and charts. These visualizations encompassed two key UX metrics: Usability and User engagement. For survey responses, the visualizations encompassed an additional two key UX metrics: Responsiveness and Visual Aesthetics. Through these visual representations, we offer a multi-dimensional perspective on user experiences, fostering a deeper understanding of the factors shaping user satisfaction and engagement with e-commerce applications in Bangladesh. This visually rich presentation offers a multi-dimensional perspective on user experiences, revealing the interplay between functional aspects, perceived efficiency, user interaction patterns, and aesthetic preferences. This deeper understanding will guide design interventions that not only address functional needs but also resonate with user aspirations and preferences in the Bangladeshi e-commerce landscape.

E. Sentiment Analysis

Before conducting our sentiment analysis, we performed several preprocessing steps on the collected data to ensure that it was in a suitable format for analysis. To ensure data integrity and consistency, we initiated a rigorous preprocessing stage. Leveraging the Pandas library, we addressed missing values and errors, rectifying inconsistencies and discarding records with missing text reviews or irrelevant API content. To standardize the text data, we performed normalization techniques, including lowercase conversion, punctuation removal, and emoji elimination. To enhance the dataset's linguistic homogeneity, we employed the Google Cloud Translation API to translate Bangla reviews and transliterate Banglish reviews into English. [2]

Investigating the sentiments expressed in Bangladeshi e-commerce apps reviews, we leveraged two powerful pre-trained models: RoBERTa and DistilBERT. Their multilingual capabilities and affinity for e-commerce app reviews made them ideal choices for this domain-specific analysis. Both rooted in the transformer architecture, these models were trained on massive text datasets. For deeper understanding, they employ tokenization, breaking down text into bite-sized pieces, and utilize a combination of linguistic and structural features like word embeddings, Masked Language Modeling (MLM), and Next Sentence Prediction (NSP) to grasp the context. In the realm of sentiment analysis, data scarcity often poses a challenge, leading to two main approaches: non-machine-learning and machine-learning-based. We now delve deeper into the specific characteristics of the chosen models. [2]

1) *RoBERTa*: RoBERTa is a transformer-based model which excels in processing input sequences and generating output sequences through self-attention mechanisms. It first transforms input tokens into fixed-length embeddings, which are then passed through multi-layered transformers for capturing inter-token dependencies. A final linear layer utilizes transformer outputs to predict task-specific outcomes, such as next word prediction for translation or class labels for text classification. Trained via masked language modeling, RoBERTa effectively learns word relationships and meanings,

demonstrating strong generalization capabilities across various tasks. [2]

processedContent	vader_neg	vader_neu	vader_pos	vader_compound	roberta_neg	roberta_neu	roberta_pos
Alhamdulillah best app	0.000	0.323	0.677	0.6369	0.002774	0.065280	0.931946
best app for books. But kinda pricy	0.234	0.594	0.172	-0.2382	0.476844	0.365017	0.158439
Trustful institution.	0.000	0.244	0.756	0.4767	0.081619	0.707130	0.211252
Very helpful for book lovers	0.000	0.316	0.684	0.7594	0.001869	0.049699	0.948432
Great app's useful	0.000	0.196	0.804	0.6249	0.004349	0.054735	0.940915

TABLE II: Sentiment Analysis Using RoBERTa

2) *DistilBERT*: DistilBERT, a compact and efficient version of BERT, excels in various natural language tasks, including sentiment analysis. It leverages pre-trained knowledge from a massive English Wikipedia dataset, offering a strong foundation for sentiment analysis models. To harness its capabilities, we fine-tune DistilBERT on labeled text data with assigned sentiments (e.g., positive, negative, neutral). This supervised learning approach trains the model to accurately predict sentiments based on textual features. Once fine-tuned, DistilBERT effectively classifies the sentiment of new, unseen text inputs, making it a valuable tool for sentiment analysis tasks. [2]

Processed Content	Pipeline Label (pipeline_label)	Pipeline Score (pipeline_score)
Alhamdulillah best app	POSITIVE	0.998885
Trustful institution.	POSITIVE	0.999769
Very helpful for book lovers	POSITIVE	0.999734
Great App's	POSITIVE	0.999828
useful	POSITIVE	0.999763

TABLE III: Sentiment Analysis Using DistilBERT

Like any model, a sentiment analysis tool's accuracy hinges on the training data it feeds on and the texts it analyzes. To ensure reliable results, we need two key ingredients: a robust, diverse dataset spanning various sentiments, and careful testing on different examples. This ensures the model not only learns effectively but also performs consistently across diverse text landscapes.

F. Interpretation

Analyzing trends in Playstore reviews revealed fascinating insights into the evolution of user perceptions regarding the app's usability and user engagement. Our findings demonstrated a clear correlation between improvements in specific UI/UX elements and an upswing in positive reviews over time. This suggests that the application's developers have successfully responded to user feedback, translating it into tangible design enhancements that resonate with the audience. Notably, positive, negative and neutral usability and user engagement of true and false values emerged as a key driver of positive sentiment, highlighting the critical role of User Experience metrics in analysis of Bangladeshi E-commerce applications. However, there remains room for further optimization, as certain areas still attract negative feedback. By leveraging these insights, the app can refine its design, prioritize future interventions, and ultimately cultivate a user experience that fosters not only increased engagement but also sustained positive feedback.

III. RESULTS AND DISCUSSION

A. Survey

The evaluation of user experiences in our study involved the administration of questionnaires through google forms,

which were completed by 43 respondents actively engaging with the e-commerce mobile applications under scrutiny. The collected data was subsequently processed utilizing two widely recognized metrics, the User Experience Questionnaire (UEQ), Likert Scale and the System Usability Scale (SUS). [1] Among

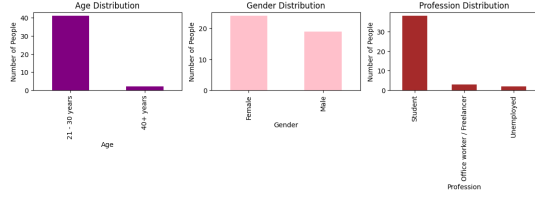


Fig. 4: Respondent Summary

43 respondents, a summary was taken as to how many of them have used the platform they are reviewing in the survey. Only then the review was taken on the metrics from them on that particular platform. The figure 5 shows the usage summary of the platforms: The questions and the System Usability Scale

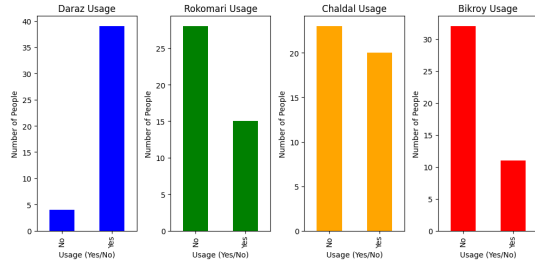


Fig. 5: Usage Summary

used for the survey are shown in the figure 1 and figure 2 respectively. All the answers of the questions were taken on a scale of 1-5 where 1 was extremely dissatisfying to 5 being extremely satisfying experience.

The interpretation from the survey shows that except user engagement, the other three metrics are mostly positive for the four kinds of e-commerce platform that were evaluated. The figure 6 shows the average usability, responsiveness, visual aesthetics and user engagement for the four platforms.

B. Webscraping Google Playstore Reviews

We conducted web scraping on four prominent platforms—Daraz, Chaldal, Rokomari, and Bikroy—to collect and analyze user reviews from Google Play Store. The data encompassed overall ratings, user comments, and insights into whether the platforms responded to user feedback.

The overall ratings across all four platforms were analyzed to understand the distribution of user sentiment. Figure 7 illustrates the distribution of ratings, providing a visual representation of user satisfaction on each platform.

The user engagement analysis involved scrutinizing user comments and identifying whether each comment received a reply from the respective platforms. This allowed us to evaluate the responsiveness of Daraz, Chaldal, Rokomari, and Bikroy to user feedback. The results found are shown in the

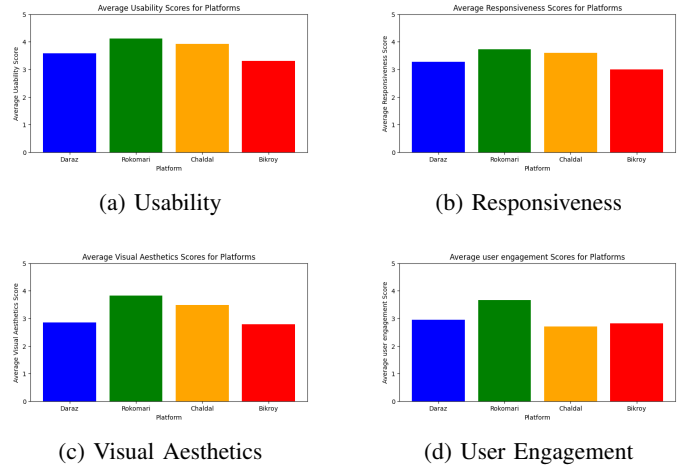


Fig. 6: Average calculated metric of each platform

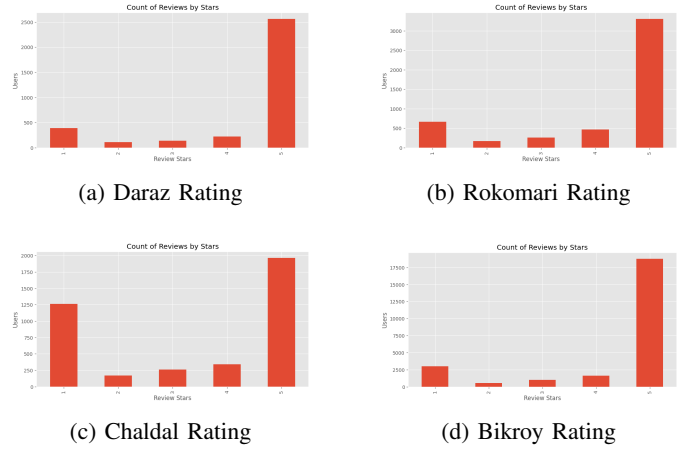


Fig. 7: Rating obtained from Google Play Store

figure 8. Assigning a user engagement score based on platform responses provided a metric to evaluate each platform's active engagement with its user base. Our analysis aimed to identify the platforms that fostered robust user engagement.

The value 1 represents positive user engagement of the application with the customer and the value 0 represents negative user engagement. From the survey, we can observe in figure 7a that Rokomari and Daraz has more average positive outcome than Bikroy and Chaldal. This data slightly defers from the reviews by the fact that bikroy engages more in google playstore review.

C. Sentiment-analysis

After preprocessing the data i.e translating the Bengali reviews to English and removing emoticons from the reviews using Google Sheets formulas, we applied state-of-the-art natural language processing models, DistilBERT and RoBERTa, for sentiment analysis on user reviews collected from Daraz, Chaldal, Rokomari, and Bikroy. The primary objective was aspect categorization to identify specific metrics reflecting

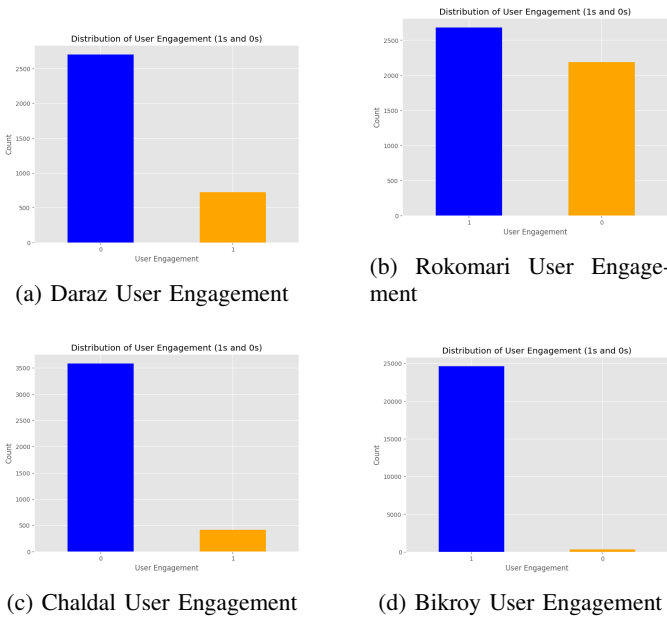


Fig. 8: User Engagement obtained from Google Play Store

user sentiment. However, despite leveraging advanced models, achieving precise aspect categorization proved challenging.

The richness and complexity of user language introduced challenges in accurately categorizing aspects. Ambiguous expressions, varying tones, and contextual nuances in the reviews posed difficulties in associating sentiments with specific metrics.

The absence of consistent patterns in user reviews further complicated aspect categorization. Users often expressed opinions in diverse ways, making it challenging to establish a uniform mapping between sentiments and specific metrics. While aspect categorization posed challenges, the sentiment analysis provided valuable insights into the overall emotional tone of user reviews. We observed patterns of sentiment distribution across positive, negative, and neutral sentiments for each platform.

Based on the challenges encountered in aspect categorization, we recommend refining the sentiment analysis approach. Exploring additional contextual information, incorporating domain-specific lexicons, and refining model fine-tuning processes could enhance the precision of aspect categorization.

The accuracy found in the models for each platform can be found in table IV.

IV. LIMITATIONS AND FUTURE WORK

A. Limitations

In our study, we acknowledge certain limitations that affect the scope and generalizability of our findings. Firstly, our analysis focused on a subset of key UI/UX metrics, specifically usability, responsiveness, visual aesthetics, and user engagement. The exclusion of other important metrics, such as accessibility and user satisfaction, implies that our insights offer a partial understanding of the overall user experience.

Model	Accuracy	Precision	Recall	F1 score
DistilBERT	0.774	0.878	0.897	0.887
RoBERTa	0.562	0.940	0.921	0.930

(a) Daraz Result

Model	Accuracy	Precision	Recall	F1 score
DistilBERT	0.874	0.912	0.932	0.922
RoBERTa	0.754	0.971	0.715	0.823

(b) Chaldal Result

Model	Accuracy	Precision	Recall	F1 score
DistilBERT	0.865	0.928	0.929	0.928
RoBERTa	0.744	0.962	0.839	0.897

(c) Rokomari Result

Model	Accuracy	Precision	Recall	F1 score
DistilBERT	0.851	0.920	0.927	0.923
RoBERTa	0.697	0.959	0.900	0.929

(d) Bikroy Result

TABLE IV: Comparison of Results

Secondly, our examination was confined to the Google Play Store app on Android devices, neglecting the iOS version. This platform-specific focus may limit the transferability of our conclusions to the broader mobile app landscape. Additionally, the subjectivity inherent in metrics like visual aesthetics raises concerns about the varied interpretations among users with diverse preferences and backgrounds.

Moreover, our reliance on data solely from the Google Play Store app's reviews as the single data source introduces limitations. While these reviews provide valuable user feedback, the absence of other channels, such as direct user interviews, surveys, or usability testing, could restrict the depth of our understanding of the user experience.

Furthermore, the study was conducted on a specific version of the Google Play Store app available during the study period. Subsequent updates to the app may introduce changes that could impact the identified metrics and conclusions. Despite these limitations, our study serves as an initial exploration into specific facets of the user experience on the Google Play Store app. Addressing these limitations and expanding the scope in future research endeavors could contribute to a more comprehensive understanding of UI/UX in mobile app environments.

B. Future Work

This study can be extended by completing aspect categorization from the Google Play Store review data, a task that remained unfinished in our current study. Aspect categorization involves identifying specific themes or topics within user reviews, providing a more detailed insight into user sentiments. We plan to leverage advanced natural language processing techniques and machine learning models to enhance the accuracy of our analysis and gain deeper insights into the features that influence user satisfaction. This effort aims to contribute to a more comprehensive understanding of user experiences on the Google Play Store.

V. RELATED WORKS

- 1) The Comparison of Evaluation on User Experience and Usability of Mobile Banking Applications Using User Experience Questionnaire and System Usability Scale [1]
- 2) Sentiment Analysis of Restaurant Reviews from Bangladeshi Food Delivery Apps [2]
- 3) User-centred Design and Evaluation of Web and Mobile based Travelling Applications [3]

REFERENCES

- [1] Setiyawati, N.; Bangkalang, D.H. The Comparison of Evaluation on User Experience and Usability of Mobile Banking Applications Using User Experience Questionnaire and System Usability Scale. *Proceedings* 2022, 82, 87. <https://doi.org/10.3390/proceedings2022082087>
- [2] E. R. Rhythm, R. A. Shuvo, M. S. Hossain, M. F. Islam and A. A. Rasel, "Sentiment Analysis of Restaurant Reviews from Bangladeshi Food Delivery Apps," 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 2023, pp. 1-5, doi: 10.1109/ESCI56872.2023.10100214.
- [3] Ismail, Nor Azman & Nizam, Fatimah & Yuen, Simon & Hasan, Layla & Mohamed, Su & Wong, Yee & Leng, Khalid & Krayz Allah, Khalid. (2021). User-centred Design and Evaluation of Web and Mobile based Travelling Applications. *International Journal of Advanced Computer Science and Applications*. 12. 2021. 10.14569/IJACSA.2021.0120854.
- [4] Daraz Online Shopping App - apps on Google Play. (n.d.). <https://play.google.com/store/apps/details?id=com.daraz.android>
- [5] Rokomari: Trusted Online Store - apps on Google Play. (n.d.). <https://play.google.com/store/apps/details?id=com.rokomari>
- [6] Bikroy - Everything Sells - apps on Google Play. (n.d.). <https://play.google.com/store/apps/details?id=com.bikroy>
- [7] Chaldal: Online grocery - apps on Google Play. (n.d.). <https://play.google.com/store/apps/details?id=com.chaldal.poached>
- [8] Maria, A. (2023, November 22). Top 10 e-commerce websites in Bangladesh. MyBangla24. <https://mybangla24.com/ecommerce-websites-bangladesh>