Optimizing health outcomes: a detailed comparison of features and user sentiment in popular fitness tracking applications

Optimizando los resultados de salud: una comparación detallada de las características y el sentimiento de los usuarios en aplicaciones populares de seguimiento de actividad física

Authors

Teguh Arie Sandy[1]

[1]Doctoral Student, Universitas Negeri Yogyakarta, Yogyakarta (Indonesia)

[1]Ahli Media Consultant, Malang (Indonesia)

Corresponding author:
teguharie.2021@student.uny.ac.id
teguharies@ieee.org

How to cite in APA

Sandy, T. A. (2025). Optimizing health outcomes: a detailed comparison of features and user sentiment in popular fitness tracking applications. *Retos*, *68*, 432–444. https://doi.org/10.47197/retos.v68.114746

Abstract

Objective: This study aimed to compare the user sentiment and functionality of popular fitness tracking applications, specifically Strava™, Google Fit™, and Fitbit™, to determine how their features contributed to health optimization and user engagement.
Methodology: The research utilized sentiment analysis and functionality classification based on user reviews gathered from the Google Play Store. The analysis employed Naive Bayes and Logistic Regression methods to assess user sentiment and application performance.
Results: The analysis revealed that Strava™ demonstrated superior emotional and functional engagement, although it faced concerns regarding privacy. Google Fit™ was recognized for its usability, but it showed limitations in tracking accurate data. Fitbit™ exhibited a balanced performance but lacked significant innovation compared to the other two platforms.
Discussion: The findings of this research were consistent with existing studies on user engagement, highlighting the importance of emotional connection in fitness applications. However, unlike previous studies, the current research also emphasized the role of data accuracy, which was a limitation in Google Fit™. Furthermore, the comparison among the three applications provided new insights into how emotional and functional features impact user satisfaction.
Conclusions: Emotional engagement and data accuracy were found to be critical factors in user satisfaction and the success of fitness applications. Developers are encouraged to strike a balance between technical features and social elements to enhance user experience and support healthier lifestyles.

Keywords

Fitness applications; user sentiment; health optimization; Strava; Fitbit; Google Fit.

Resumen

Objetivo: Este estudio tuvo como objetivo comparar el sentimiento de los usuarios y la funcionalidad de aplicaciones populares de seguimiento de fitness, específicamente Strava™, Google Fit™ y Fitbit™, para determinar cómo sus características contribuyen a la optimización de la salud y el compromiso de los usuarios.
Metodología: La investigación utilizó análisis de sentimientos y clasificación funcional basados en reseñas de usuarios recopiladas de Google Play Store. El análisis empleó los métodos de Naive Bayes y Regresión Logística para evaluar el sentimiento de los usuarios y el rendimiento de las aplicaciones.
Resultados: El análisis reveló que Strava™ demostró un mayor compromiso emocional y funcional, aunque enfrentó preocupaciones relacionadas con la privacidad. Google Fit™ fue reconocido por su usabilidad, pero presentó limitaciones en la precisión del seguimiento de datos. Fitbit™ mostró un rendimiento equilibrado pero carecía de innovación significativa en comparación con las otras dos plataformas.
Discusión: Los hallazgos de esta investigación fueron consistentes con estudios previos sobre el compromiso de los usuarios, destacando la importancia de la conexión emocional en las aplicaciones de fitness. Sin embargo, a diferencia de estudios anteriores, la investigación actual también subrayó el papel de la precisión de los datos, que fue una limitación en Google Fit™. Además, la comparación entre las tres aplicaciones proporcionó nuevas perspectivas sobre cómo las características emocionales y funcionales afectan la satisfacción del usuario.
Conclusiones: Se encontró que el compromiso emocional y la precisión de los datos son factores críticos para la satisfacción del usuario y el éxito de las aplicaciones de fitness. Se alienta a los desarrolladores a lograr un equilibrio entre las características técnicas y los elementos sociales para mejorar la experiencia del usuario y apoyar estilos de vida más saludables.

Palabras clave

Aplicaciones de fitness; sentimiento del usuario; optimización de la salud; Strava; Fitbit, Google Fit.

Introduction

The proliferation of digital fitness applications has significantly reshaped health and wellness management, providing users with tools to track physical activity, set goals, and foster social connections (Tong et al., 2022). Platforms such as Strava™, Google Fit™, and Fitbit™ can be collectively referred to as 'fitness tracking apps' (FTA), which have emerged as key players in the field, offering unique features that cater to diverse user preferences. These applications not only enable users to monitor their physical progress but also aim to engage them emotionally, creating immersive environments that motivate sustained participation. While these tools have gained widespread adoption, there remains a gap in the literature concerning a comparative analysis of user sentiment and satisfaction across these platforms, particularly regarding how specific features contribute to health optimization and user engagement (Wang & Xu, 2023).

The digital fitness application market has experienced rapid growth in recent years, with global revenue reaching $16.6 billion in 2024, and it is projected to expand at a compound annual growth rate (CAGR) of 17.2% until 2030 (Grand View Research, 2023). This expansion has been driven by increasing health consciousness, advancements in wearable devices, and a growing preference for personalized fitness solutions (Hamza Mayora et al., 2025; Yoganathan & Kajanan, 2014). In 2023, approximately 87.4 million Americans used fitness applications, marking a 14.6% increase from pre-pandemic levels (Statista, 2020, 2024). This shift has transformed how individuals approach physical activity, moving from traditional manual tracking methods to sophisticated digital ecosystems that offer real-time feedback, social engagement, and data visualization (Al Ardha et al., 2024; Thijs et al., 2019).

The COVID-19 pandemic further accelerated the digital shift, with adoption rates of fitness platforms rising by 57% globally during 2020-2021 (Kartiko et al., 2023; Statista, 2020, 2024). This surge demonstrates not only the resilience of digital fitness platforms but also their capacity to adapt to changing user needs during periods of limited mobility and social distancing (Clark & Lupton, 2021; Tong et al., 2022). Additionally, these applications have evolved from basic step counters to comprehensive health management systems, integrating features such as sleep analysis, nutrition tracking, and mental wellness monitoring (Services, 2024).

Although individual fitness applications have been extensively studied, there is still a significant gap in the comparative analysis of user sentiment and emotional engagement across competing platforms. While previous studies have independently assessed user experiences with specific applications, there is a lack of systematic comparison using standardized emotional and functional metrics. This gap is particularly notable given the competitive nature of the fitness application market, where understanding the relative strengths and weaknesses of each platform could significantly inform design improvements and feature prioritization (Cho et al., 2020; Fietkiewicz & Ilhan, 2020; Martín et al., 2023).

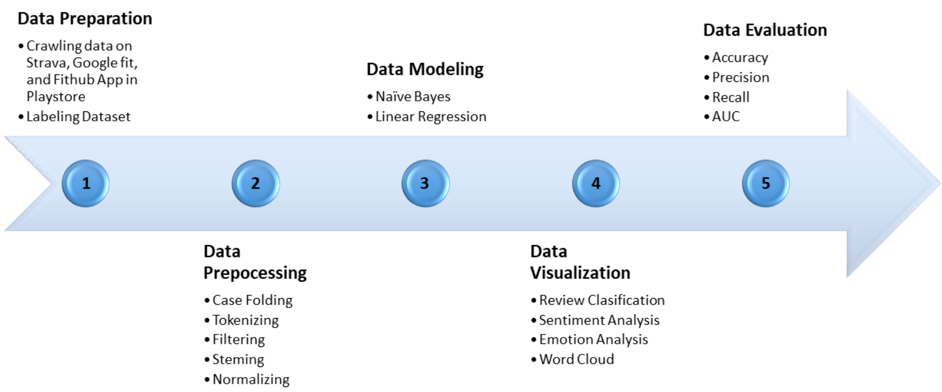
This research is significant not only for its academic contribution but also for practical applications in digital health innovation. As fitness applications increasingly integrate with broader health ecosystems and electronic medical records, understanding the emotional drivers behind user engagement is critical for developing interventions that promote sustained healthy behaviors (Sun et al., 2024). Moreover, with healthcare providers increasingly prescribing digital fitness tools as part of treatment plans, comparing platform effectiveness represents an important step toward evidence-based digital therapeutics (Cho et al., 2020; Garber et al., 2011).

To address these gaps, this study will explore how user emotional responses differ across FTA, and identify the features that drive these emotional patterns. The research will also assess which functional elements most significantly impact user satisfaction and emotional engagement within each platform. Ultimately, this study aims to develop a comparative framework for evaluating fitness applications that integrates both functional assessment and emotional response analysis. Secondary objectives include identifying specific feature-emotion relationships to inform future application development and validating methodological approaches for sentiment analysis in digital health contexts.

Method

This study employed a comprehensive five-stage methodological approach to analyze and compare user sentiment across three leading fitness applications: Strava™, Google Fit™, and Fitbit™ (Sandy et al., 2025). The research design prioritized systematic data collection, rigorous preprocessing, advanced modeling techniques, and thorough evaluation to ensure robust and reliable insights (see Figure 1 for the research workflow).

Figure 1. Research Workflow



Participants

The initial phase involved extensive data collection from multiple sources to capture a representative sample of user experiences. User reviews were systematically crawled from the Google Play Store for all three target applications, gathering textual feedback spanning a three-year period (March 2022 to February 2025). This yielded a substantial dataset of 500,000 reviews (55,031 for Strava™, 55,000 for Google Fit™, and 390,317 for Fitbit™). Each review was labeled according to its source application, timestamp, user rating (1-5 stars), and geographic region, when available. Quality control measures were implemented to identify and exclude potentially fraudulent reviews, resulting in a final dataset of 14,982 valid reviews for analysis.

Procedure

To optimize the dataset for computational analysis, a multi-step preprocessing protocol was implemented. All textual data were converted to lowercase to ensure consistency and eliminate redundancy in term recognition. Reviews were segmented into individual tokens (words and phrases) using natural language processing techniques to facilitate granular analysis. Stop words, punctuation, and non-alphanumeric characters were removed, as they contributed minimal semantic value. Words were reduced to their root forms using the Porter stemming algorithm to consolidate related terms and improve pattern recognition. Text was standardized by removing excessive spaces, correcting common misspellings, and expanding contractions to ensure consistency across the dataset.

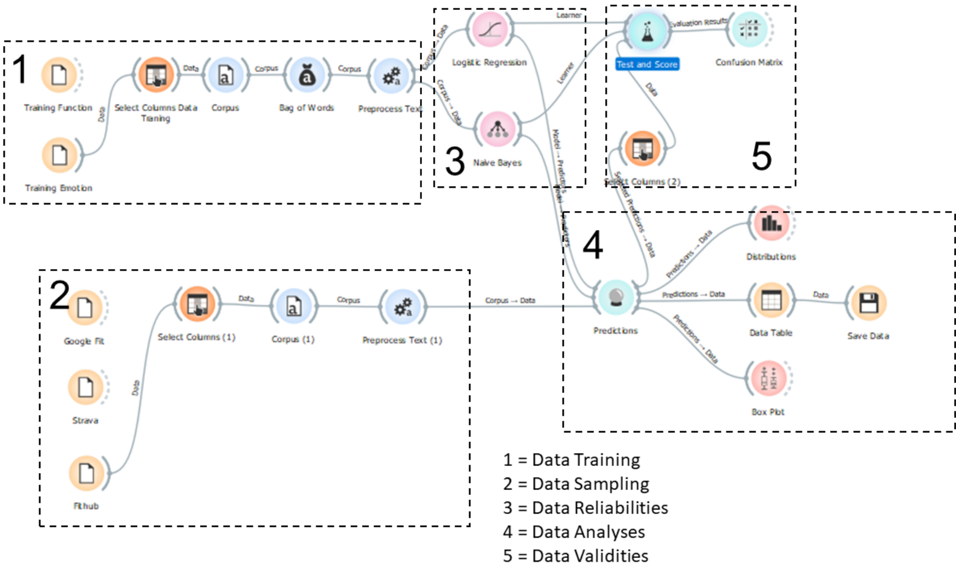
Data Modeling

The preprocessed data was subjected to multiple modeling approaches to extract meaningful patterns and relationships. A probabilistic Naïve Bayes classifier was implemented to categorize reviews based on sentiment polarity (Happiness, Disgust, Sadness, Surprise, Fear, and Anger), and to identify application-specific trends in user satisfaction (Nugroho et al., 2021). Quantitative relationships between user ratings and specific application features were analyzed using linear regression models, allowing for the identification of the features that most significantly impact overall user satisfaction across the three platforms.

Data Visualization

To translate complex analytical findings into interpretable insights, several visualization techniques were employed. Results were stratified by application, feature category, and temporal trends to identify patterns in user sentiment evolution. Emotional valence across reviews was mapped using lexicon-based approaches supplemented by machine learning algorithms to detect nuanced sentiment expressions. Beyond binary sentiment classification, the NRC Emotion Lexicon was implemented to identify specific emotional states (e.g., joy, anger, trust) associated with different application features. Frequency and co-occurrence of terms were represented visually to highlight dominant themes and concerns within user feedback (Gogula et al., 2023) (see Figure 2 for the data analysis workflow).

Figure 2. Data analysis workflow



Data Analysis

The validity and reliability of the analytical approach were assessed through multiple evaluation metrics. The overall correctness of sentiment classification was measured against a manually coded subset of reviews. The proportion of correctly identified positive/negative sentiments relative to all instances classified as such was evaluated. The sensitivity of models in detecting the full spectrum of relevant sentiments was systematically assessed. The Area Under the ROC Curve (AUC) provided a comprehensive evaluation of classification models' discriminative capacity across various threshold settings. All analyses were conducted using Python 3.10 with scikit-learn, NLTK, pandas libraries, and Orange3 Software (Ch. Kesava Manikanta et al., 2023; Sandy et al., 2025; Sitorus et al., 2024) .

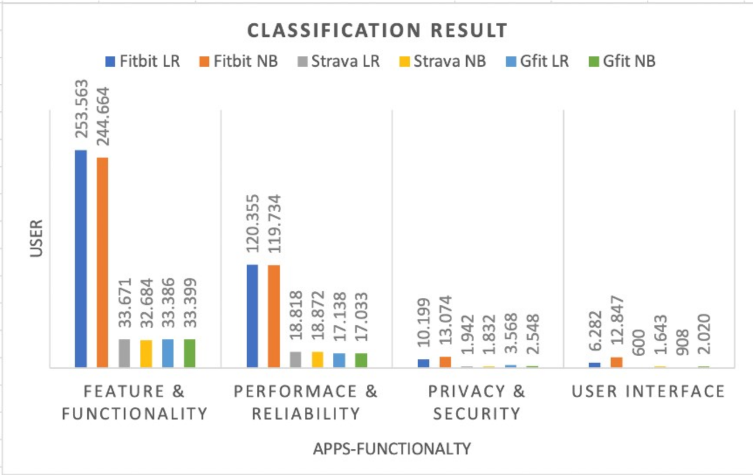
Results

The analysis of user sentiment and functionality across FTA revealed insightful patterns about how users perceive each app's features and overall user experience. The results section is categorized into three key areas: emotional sentiment, functionality assessment, and user rating distribution. These findings contribute to a deeper understanding of what drives user satisfaction and dissatisfaction in the realm of fitness applications.

Functional Features Analysis

The first aspect of the analysis focused on the functional features of the three fitness applications. Users were asked to evaluate the apps based on four core functionalities: User Interface, Privacy & Security, Performance & Reliability, and Features & Functionality. This assessment, as shown in Figure 3 (Apps-Functionality Classification Result), highlights key areas of strength and weakness for each application.

Figure 3. Apps-Functionality Classification Result



- Feature & Functionality: Fitbit™ LR (253,563 reviews) and Fitbit™ NB (244,664 reviews) lead in this category by a significant margin. Strava™ LR and Strava™ NB have lower counts, with 33,671 and 32,684 reviews, respectively. Google Fit™ has similar numbers across both models (33,386 and 33,399 reviews). This indicates that Fitbit™ stands out in offering a more extensive and diverse set of features.
- Performance & Reliability: Again, Fitbit™ performs well in this category with 120,355 reviews for LR and 119,734 reviews for NB. Strava™ follows closely with 18,818 reviews (LR) and 18,872 reviews (NB). Google Fit™ shows lower numbers in this area (17,138 and 17,033 reviews), indicating some concerns about performance and reliability in comparison.
- Privacy & Security: Fitbit™ continues to show a strong performance in Privacy & Security, with 10,199 reviews for LR and 13,074 reviews for NB, compared to Strava™'s 1,942 (LR) and 1,832 (NB), and Google Fit™'s 3,568 and 2,548 reviews. These figures suggest that users are more satisfied with Fitbit™'s security measures compared to Strava™ and Google Fit™.
- User Interface: In terms of User Interface, Fitbit™ LR had 6,282 reviews, while Fitbit™ NB had 12,847 reviews. Strava™'s reviews in this category were significantly lower (600 for LR and 1,643 for NB), while Google Fit™ received 908 (LR) and 2,020 (NB) reviews. This shows that Fitbit™, while not excelling in this area, has managed to maintain a relatively positive user experience regarding UI compared to Strava™ and Google Fit™.

To classify the data, Logistic Regression (LR) and Naïve Bayes (NB) algorithms were employed. Logistic Regression (LR) is widely used for binary classification tasks and performed particularly well in assessing the functionality and sentiment of user reviews. On the other hand, Naïve Bayes (NB), a classification algorithm based on Bayes' Theorem, assumes independence between features and performed admirably in the sentiment classification tasks. Both algorithms were applied across the platforms to classify the reviews into different emotional categories (such as "Happiness," "Anger," and "Sadness") and functionality assessments.

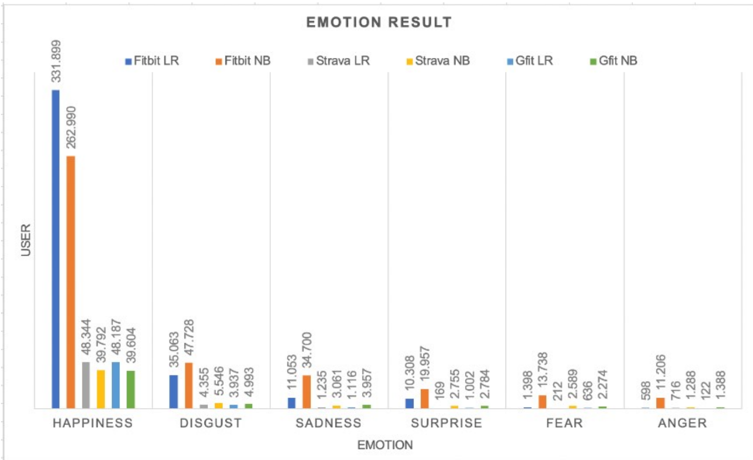
These findings reflect how user satisfaction and dissatisfaction are influenced by each application's functionality and design. While Fitbit™ dominates in functionality and privacy security, Strava™ excels in competitive features and social engagement, and Google Fit™ focuses on ease of use and integration with other Google services. Understanding these distinctions can guide future developments in digital health platforms, ensuring they meet the diverse needs of their user base.

In conclusion, the results demonstrate that each fitness application caters to different user preferences. Fitbit™ offers the most comprehensive features but struggles with UI and innovation, while Strava™ excels in competitive features but has privacy and technical issues. Google Fit™, while easy to use, lags behind in advanced features and performance. These insights are crucial for developers looking to enhance user satisfaction and engagement in the rapidly evolving fitness app market.

Emotional Sentiment Analysis

The second key finding involved the emotional responses of users across the three platforms, as captured in the sentiment analysis. Using Ekman's emotion classification model, the emotional tone of reviews was classified into six distinct categories: Happiness, Disgust, Sadness, Surprise, Fear, and Anger. The classification results, shown in Figure 4 (Ekman's Emotion Classification Result), revealed notable differences in emotional engagement between the applications.

Figure 4. Ekman's Emotion Classification Result



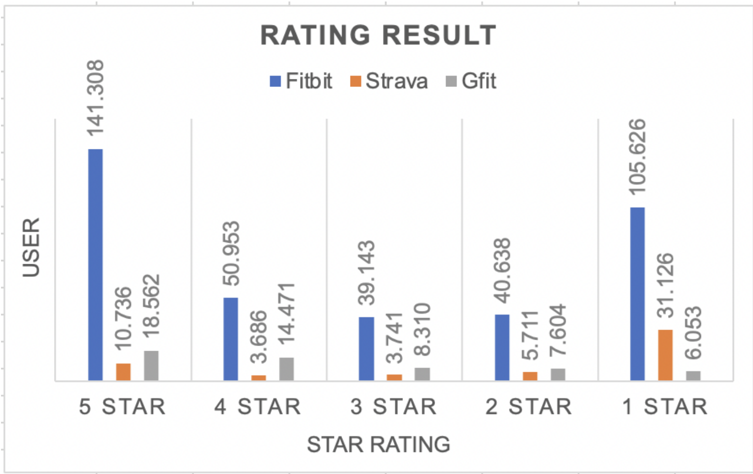
- Strava™, in particular, exhibits a dominant positive emotional response, with 48.34% of the reviews categorized as "Happiness," indicating that the majority of users have positive sentiments toward the app's features and performance. However, negative emotions are also present, with 5.5% of the reviews classified as "Disgust," and 11.05% categorized as "Sadness." Additionally, 3.95% of users expressed "Anger," indicating frustration, likely due to issues such as the app's subscription model and syncing problems. The surprise and fear emotions are less pronounced, with lower percentages for these categories.
- Google Fit™ shows a somewhat balanced emotional distribution, with 39.79% of reviews reflecting "Happiness." However, there are notable levels of negative emotions, such as "Sadness" at 3.97% and "Fear" at 1.39%. Interestingly, "Disgust" and "Surprise" also appear, with 5.5% and 2.75%, respectively. These mixed emotions can be attributed to the app's simplicity and integration with the Google ecosystem, which some users appreciate, while others express dissatisfaction due to the lack of advanced features.
- Fitbit™'s emotional sentiment also leans towards positivity, with 46.18% of the reviews classified as "Happiness." Nevertheless, "Sadness" emerges as a significant emotion, representing 10.53% of the reviews, reflecting dissatisfaction with the app's outdated design and limited features. Negative emotions like "Anger" (5.98%) and "Disgust" (4.35%) are also prevalent, indicating frustration from users who expect more from the app in terms of advanced functionality and user experience.

The emotional sentiment analysis confirms that while the majority of users for each app report positive emotions, there are notable issues related to user experience, technical difficulties, and feature limitations, especially for Fitbit™. These insights are crucial for understanding the emotional drivers behind user engagement and satisfaction, and they suggest that addressing user concerns, such as privacy issues and functionality limitations, could enhance user retention and satisfaction.

User Rating Distribution

The star rating distribution for the three fitness applications—Fitbit™, Strava™, and Google Fit™—provides a clear picture of overall user satisfaction and dissatisfaction, as shown in Figure 5 (App Rating Distribution). The distribution is categorized across five rating levels: 5-star, 4-star, 3-star, 2-star, and 1-star.

Figure 5. Apps Rating Distribution



- Fitbit™ stands out with the highest number of 5-star ratings, totaling 141,308 (45%) of the total ratings, indicating strong user satisfaction. However, it also experiences a significant number of 1-star ratings, with 105,626 (34%) reviews. This suggests that while many users are highly satisfied with the app, others are deeply dissatisfied, likely due to limitations in advanced features and design issues.
- Strava™'s ratings are more evenly distributed across different levels. The app garnered 50,953 (16%) 4-star ratings and 39,143 (13%) 3-star ratings, with a considerable 18,562 (6%) 5-star ratings. Although Strava™ has a relatively high proportion of 4- and 3-star ratings, it also faced dissatisfaction, with 8,310 (3%) 2-star ratings and 3,741 (1%) 1-star ratings. This indicates that users appreciate the app's competitive features but experience dissatisfaction due to issues like pricing, data accuracy, and syncing problems.
- Google Fit™, on the other hand, shows a more polarized rating distribution. The app received 18,562 (6%) 5-star ratings but also faced a substantial number of low ratings, with 7,604 (2%) 2-star ratings and 6,053 (2%) 1-star ratings. This suggests that while users appreciate the app's simplicity and integration with the Google ecosystem, its limitations in performance tracking and motivational features contribute to lower satisfaction for many users.

In summary, Fitbit™ leads with the highest number of 5-star ratings but also faces significant criticism, especially from users seeking more advanced features. Strava™ exhibits a more balanced distribution, with mixed reviews reflecting both positive sentiment and dissatisfaction due to specific functionality issues. Google Fit™, although appreciated for its ease of use, struggles to achieve higher ratings, with a significant proportion of users expressing dissatisfaction due to its limited features and performance inconsistencies.

Model Performance for Emotion

The performance of the models in emotion classification was evaluated for Fitbit™, Strava™, and Google Fit™ using two classifiers: Logistic Regression (LR) and Naïve Bayes (NB). The results across key metrics—such as AUC, CA, F1, Precision, Recall, and MCC—are shown in Table 1.

Table 1. Model Performance for Emotion Classification

Aspect	Fitbit™ LR	Fitbit™ NB	Strava™ LR	Strava™ NB	Gfit™ LR	Gfit™ NB
AUC	-0,921	0,079	0,946	0,947	0,886	0,887
CA	0,795	0,795	0,919	0,919	0,811	0,811
F1	0,748	0,748	0,901	0,901	0,765	0,765
Prec	0,802	0,802	0,885	0,885	0,804	0,804
Recall	0,795	0,795	0,919	0,919	0,811	0,811
MCC	0,556	0,556	0,606	0,606	0,528	0,528

For Fitbit™, the emotion classification models showed weak performance, particularly with the Logistic Regression (LR) model, which had an AUC of -0.921. The Naïve Bayes (NB) model performed slightly better, but still showed weak discriminatory power with an AUC of 0.079. The performance metrics indicated that Fitbit™'s emotion classification model was not as reliable as the others.

In contrast, Strava™ demonstrated significantly better performance in emotion classification, with both LR and NB models achieving high AUC scores (0.946 and 0.947, respectively). Google Fit™ performed moderately well, with AUC scores of 0.886 for LR and 0.887 for NB, suggesting that while its models performed better than Fitbit™, they did not reach the same level of precision as Strava™.

Model Performance for Function Classification

The results of the functionality classification analysis across the three fitness applications—Strava™, Google Fit™, and Fitbit™—reveal distinct patterns in their performance. Table 2 presents the detailed performance metrics for each application, including AUC (Area Under the Curve), Classification Accuracy (CA), F1 Score, Precision, Recall, and Matthews Correlation Coefficient (MCC), evaluated using both Logistic Regression (LR) and Naïve Bayes (NB) classifiers.

Table 2. Model Performance for Function Classification

Aspect	Gfit™ LR	Gfit™ NB	Strava™ LR	Strava™ NB	Fitbit™ LR	Fitbit™ NB
AUC	0,670	0,669	0,715	0,721	0,833	0,833
CA	0,917	0,917	0,949	0,949	0,948	0,948
F1	0,916	0,916	0,944	0,944	0,943	0,943
Prec	0,925	0,925	0,938	0,938	0,946	0,946
Recall	0,917	0,917	0,949	0,949	0,948	0,948
MCC	0,844	0,844	0,899	0,899	0,896	0,896

Table 2 provides a comparative overview of the model performance for all three applications across the different classifiers and performance metrics. It highlights Fitbit™ as the top performer in functionality classification, followed by Strava™, and Google Fit™, which showed the least effective performance.

- Fitbit™ showed the highest performance in classifying functionalities, with an AUC score of 0.833 for both classifiers (LR and NB), indicating strong discriminatory power. The Classification Accuracy (CA) was 0.948, reflecting the model's ability to accurately categorize the app's functionalities. The F1 Score was 0.943, which demonstrates a strong balance between precision and recall. The Precision and Recall values were also high at 0.946 and 0.948, respectively, while the MCC of 0.896 indicates a strong correlation between the predicted and actual functionality classifications.
- Strava™, while not as strong as Fitbit™, still performed well in functionality classification. The AUC for Strava™ was 0.715 for LR and 0.721 for NB, indicating solid discriminatory power but not as robust as Fitbit™. The CA was 0.949 for both classifiers, showing high classification accuracy. The F1 Score of 0.944 suggests that Strava™ was effective at balancing precision and recall. The Precision and Recall values were 0.938 and 0.949, respectively, indicating that Strava™ was highly effective in recalling relevant functionalities. The MCC for Strava™ was 0.899, showing a positive correlation between predicted and actual classifications.
- Google Fit™ exhibited the weakest performance in functionality classification. The AUC scores were 0.67 for LR and 0.669 for NB, which were the lowest among the three applications. However, the CA for Google Fit™ was still relatively high at 0.917, although it was lower than that of Strava™ and Fitbit™. The F1 Score of 0.916 suggests that the models for Google Fit™ were effective at balancing precision and recall, but slightly less effective than the other two applications.

The Precision and Recall values were 0.925 and 0.917, respectively, indicating that while Google Fit™ was effective at correctly identifying functionalities, it was less reliable in recalling positive instances. The MCC for Google Fit™ was 0.844, demonstrating a strong correlation between predicted and actual functionality classifications, but still lower than Strava™ and Fitbit™.

Discussion

This study provides a comprehensive comparative analysis of three major fitness applications—Strava™, Google Fit™, and Fitbit™—focusing on both their functional features and emotional responses. By employing advanced sentiment analysis and functionality classification, this research reveals key patterns in user engagement, shedding light on how application design influences user sentiment and overall ratings. The findings are discussed in relation to each application's functionality, emotional responses, user ratings, and model performance, along with the broader implications for digital health platforms.

Interpretation of Application Functionality Findings

The analysis of functionality across Strava™, Google Fit™, and Fitbit™ provides valuable insights into how their design philosophies impact user experiences. Each platform demonstrates distinct strengths and weaknesses, shaped by their core design principles and target user demographics.

Strava™ excels in "Features & Functionality," which aligns with its focus on creating a competitive social environment. Users appreciate its robust set of features for tracking, sharing, and competing in activities such as running and cycling. However, Strava™ faces challenges in "Privacy & Security," where concerns about user data exposure are prevalent. These privacy vulnerabilities compromise user trust, especially among users who prioritize data security (Nwaimo et al., 2024).

Google Fit™ stands out for its user-friendly interface, scoring highly in "User Interface," which contributes to its appeal among casual users. However, its weaker performance in "Performance & Reliability" highlights limitations in tracking accuracy, particularly in outdoor activities. The trade-off between simplicity and performance places Google Fit™ as an accessible choice for casual fitness users but limits its appeal to more advanced trackers (Lewis & Sauro, 2021).

Fitbit™ performs moderately across all categories but lacks specialized strengths. While its balanced approach appeals to users seeking an all-around solution, its outdated design and lack of innovation place it at a disadvantage compared to Strava™ and Google Fit™. This finding underscores the challenge of catering to both casual users and fitness enthusiasts without specializing in either group (Li et al., 2018; Yan et al., 2021).

Analysis of Emotional Response Patterns

The relationship between functionality strengths and emotional responses reveals important insights into how users interact with the apps on an emotional level.

Strava™ evokes the most positive sentiment, particularly "Happiness," with users appreciating its competitive features and social engagement. However, a significant portion of negative emotions, such as "Anger" and "Disgust," arise due to issues like the subscription model and technical glitches. These findings illustrate how a feature-rich app can drive positive emotions while simultaneously generating frustration when performance issues arise (Dirin et al., 2022).

Google Fit™ exhibits a more balanced emotional distribution, with a mix of positive and negative sentiments. While the app's simplicity is appreciated, its lack of motivational features and occasional performance issues contribute to the more even emotional split. This balanced emotional engagement suggests that while users find Google Fit™ useful, it does not foster the deep emotional connection seen in more specialized apps like Strava™.

Fitbit™ generates the least positive sentiment, with many users expressing "Sadness" and "Anger." Complaints about its limited features and lack of innovation likely contribute to these negative emotions. This pattern emphasizes how an app that does not offer standout features may struggle to maintain positive emotional engagement (Tahir et al., 2024).

Correlation Between Ratings and Sentiment

The correlation between emotional sentiment and user ratings reveals significant insights into user satisfaction (Nurmi et al., 2020; Rehman et al., 2023; Taylor, 2024). Fitbit™ demonstrates high user ratings, with a large proportion of 5-star reviews, suggesting strong user loyalty and satisfaction. Despite its moderate performance in sentiment and functionality, Fitbit™'s brand loyalty may play a significant role in these high ratings.

Strava™, while generating high levels of "Happiness," has a more mixed rating distribution, with a significant proportion of lower ratings (1 and 2 stars). This indicates that while users are emotionally engaged with Strava™'s features, issues like pricing, syncing problems, and data accuracy contribute to dissatisfaction.

Google Fit™ struggles with lower ratings, correlating with its functionality and emotional sentiment findings. The app's limitations in performance and functionality are reflected in the higher number of 1-star reviews, indicating user frustration despite its simplicity and integration with other Google services.

Model Performance and Methodological Implications

The performance of models across applications reveals important insights into the effectiveness of sentiment classification and functionality assessment. Strava™ demonstrated strong performance in both emotion and functionality classification, while Fitbit™ showed weaker results, especially in emotion classification. This discrepancy could be due to the utilitarian user base of Fitbit™, which prioritizes functionality over emotional engagement.

Naïve Bayes performed well across all applications, whereas Logistic Regression struggled with some cases, particularly emotion classification for Fitbit™. These findings suggest the need for continued exploration of classification models to better capture the complexities of user sentiment in digital health platforms.

Theoretical and Practical Implications

The results underscore the importance of emotional engagement in user retention and satisfaction in digital health applications. Emotional responses, such as happiness from social interaction or frustration from technical issues, play a significant role in shaping user perceptions. Developers should focus on emotional engagement alongside technical performance to foster long-term user satisfaction (Barbosa et al., 2022).

From a practical standpoint, the study highlights the need for developers to balance functionality and privacy concerns, particularly for social platforms like Strava™. Improving performance and reliability, as seen in Google Fit™, and offering specialized features, as in Strava™, can enhance user satisfaction. Fitbit™, however, demonstrates the risks of offering a generalized experience without a clear focus on user needs.

Future Research Directions

Future studies should focus on gaps in understanding user engagement by incorporating diverse user demographics and longitudinal behavior studies. Additionally, exploring other features, such as sleep tracking or nutrition integration, would provide insights into their impact on user sentiment and engagement.

The novelty of this study lies in combining functionality and emotional analysis. The use of sentiment analysis and classification models provides a nuanced understanding of user engagement in fitness apps, opening avenues for improving user-centered design in digital health applications.

Conclusions

This study provides a comprehensive comparison of three major fitness applications—Strava™, Google Fit™, and Fitbit™—analyzing both their functional features and the emotional responses they elicit from users. Through sentiment analysis and functionality classification, the findings highlight how each app's design philosophy influences user engagement, satisfaction, and overall experience. Strava™ stands out for its superior performance in both emotional and functional classifications, driven by its rich feature set and strong community-oriented approach. However, its struggles with privacy and security issues suggest that the balance between social engagement and user protection remains a challenge. Google Fit™, with its focus on simplicity and integration within the Google ecosystem, performs well in terms of usability but falls short in terms of advanced tracking features and emotional engagement. Fitbit™, while offering a more balanced feature set, fails to capture user attention in the same way, struggling with both interface design and functionality, leading to lower emotional and user ratings.

These results underscore the importance of emotional engagement in shaping the success of digital health platforms. The emotional responses of users—ranging from happiness and motivation to frustration and anger—directly correlate with functionality, highlighting the need for apps to not only deliver accurate and reliable fitness tracking but also foster a positive, user-friendly experience. By addressing issues such as privacy concerns, enhancing feature innovation, and improving user interfaces, developers can create more engaging and effective platforms that meet the diverse needs of their users. Future research should continue exploring the evolving dynamics of user sentiment over time and across different app categories to better understand how to enhance user engagement and satisfaction in the digital health space.

References

Al Ardha, M. A., Nurhasan, N., Nur, L., Chaeroni, A., Bikalawan, S. S., & Yang, C. B. (2024). Analysis of Android-Based Applications in Physical Education and Sports: Systematic Review. Retos, 57, 390–398. https://doi.org/10.47197/retos.v57.107158

Barbosa, H. F., García-Fernández, J., Pedragosa, V., & Cepeda-Carrion, G. (2022). The use of fitness centre apps and its relation to customer satisfaction: A UTAUT2 perspective. International Journal of Sports Marketing and Sponsorship, 23(5), 966–985. https://doi.org/10.1108/IJSMS-01-2021-0010

Ch. Kesava Manikanta, A. Gowtham, Ch. Prasanth Kumar, Ch. Sai Sundhar Raghuram, B. Sai Mahesh, & B. Sai Jyothi. (2023). YOUTUBE COMMENT ANALYSIS USING MACHINE LEARNING. EPRA International Journal of Research & Development (IJRD), 7–11. https://doi.org/10.36713/epra14774

Cho, H., Chi, C., & Chiu, W. (2020). Understanding sustained usage of health and fitness apps: Incorporating the technology acceptance model with the investment model. Technology in Society, 63, 101429. https://doi.org/10.1016/j.techsoc.2020.101429

Clark, M., & Lupton, D. (2021). Pandemic fitness assemblages: The sociomaterialities and affective dimensions of exercising at home during the COVID-19 crisis. Convergence: The International Journal of Research into New Media Technologies, 27(5), 1222–1237. https://doi.org/10.1177/13548565211042460

Dirin, A., Laine, T., & Nieminen, M. (2022). Feelings of Being for Mobile User Experience Design. International Journal of Human-Computer Interaction, 39. https://doi.org/10.1080/10447318.2022.2108964

Fietkiewicz, K., & Ilhan, A. (2020, January). Fitness Tracking Technologies: Data Privacy Doesn't Matter? The (Un)Concerns of Users, Former Users, and Non-Users. https://doi.org/10.24251/HICSS.2020.421

Garber, C. E., Blissmer, B., Deschenes, M. R., Franklin, B. A., Lamonte, M. J., Lee, I.-M., Nieman, D. C., & Swain, D. P. (2011). Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory, Musculoskeletal, and Neuromotor Fitness in Apparently Healthy Adults: Guidance for Prescribing Exercise. Medicine & Science in Sports & Exercise, 43(7), 1334–1359. https://doi.org/10.1249/MSS.0b013e318213fefb

Gogula, S. D., Rahouti, M., Gogula, S. K., Jalamuri, A., & Jagatheesaperumal, S. K. (2023). An Emotion-Based Rating System for Books Using Sentiment Analysis and Machine Learning in the Cloud. Applied Sciences, 13(2), 773. https://doi.org/10.3390/app13020773

Grand View Research. (2023). Fitness App Market Size, Share And Growth Report, 2030. https://www.grandviewresearch.com/industry-analysis/fitness-app-market

Hamza Mayora, S., Terrades Daroqui, J., Fernández Piqueras, R., Dorado, V., & Farías Torbidoni, E. I. (2025). Digitalización del senderismo en España. Uso de aplicaciones móviles, preferencias y barreras percibidas. Retos, 67, 624–642. https://doi.org/10.47197/retos.v67.114425

Kartiko, D. C., Siantoro, G., Callixte, C., Lesmana, H. S., Aljunaid, M., Komaini, A., & Ayubi, N. (2023). Application of a Healthy Lifestyle Through Sports Science Knowledge to Correct Bad Habits After the COVID-19 Outbreak: Systematic Review. Retos, 50, 511–515. https://doi.org/10.47197/retos.v50.99124

Lewis, J., & Sauro, J. (2021). USABILITY AND USER EXPERIENCE: DESIGN AND EVALUATION (pp. 972–1015).

Martín, F., García-Fernández, J., Valcarce-Torrente, M., Bernal-García, A., Gálvez-Ruiz, P., & Angosto-Sánchez, S. (2023). Importance-performance analysis in fitness apps. A study from the viewpoint of gender and age. Frontiers in Public Health, 11, 1226888. https://doi.org/10.3389/fpubh.2023.1226888

Nugroho, K. S., Sukmadewa, A. Y., Wuswilahaken Dw, H., Bachtiar, F. A., & Yudistira, N. (2021). BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews. 6th International Conference on Sustainable Information Engineering and Technology 2021, 258–264. https://doi.org/10.1145/3479645.3479679

Nurmi, J., Knittle, K., Ginchev, T., Khattak, F., Helf, C., Zwickl, P., Castellano-Tejedor, C., Lusilla-Palacios, P., Costa-Requena, J., Ravaja, N., & Haukkala, A. (2020). Engaging Users in the Behavior Change Process With Digitalized Motivational Interviewing and Gamification: Development and Feasibility Testing of the Precious App. JMIR mHealth and uHealth, 8(1), e12884. https://doi.org/10.2196/12884

Nwaimo, C., Adegbola, A., & Adegbola, M. (2024). Data-driven strategies for enhancing user engagement in digital platforms. International Journal of Management & Entrepreneurship Research, 6, 1854–1868. https://doi.org/10.51594/ijmer.v6i6.1170

Rehman, U., Abbasi, A., Ting, D. H., Hassan, M., & Khair, N. (2023). Exploring the Impact of Gamified Experiences on User Engagement in Fitness Apps: A GAMEFULQUEST Perspective. IEEE Transactions on Engineering Management, PP, 1–15. https://doi.org/10.1109/TEM.2023.3347231

Sandy, T. A., Anik Ghufron, Ali Muhtadi, & Pujiriyanto. (2025). Text Classification of Duolingo Reviews on Google Play: Insights for Enhancing M-Learning Applications. International Journal of Interactive Mobile Technologies (iJIM), 19(07), 206–223. https://doi.org/10.3991/ijim.v19i07.52891

Services, C. I. (2024, April 30). From Tracking to Training: The Evolution of Fitness Apps. Medium. https://medium.com/@ciphernutz/from-tracking-to-training-the-evolution-of-fitness-apps-ba1255d16c49

Sitorus, Z., Saputra, M., Sofyan, S. N., & Susilawati. (2024). SENTIMENT ANALYSIS OF INDONESIAN COMMUNITY TOWARDS ELECTRIC MOTORCYCLES ON TWITTER USING ORANGE DATA MINING. INFOTECH Journal, 10(1). https://doi.org/10.31949/infotech.v10i1.9374

Statista. (2020, June). Global health and fitness app downloads 2020. Statista. https://www.statista.com/statistics/1127248/health-fitness-apps-downloads-worldwide/

Statista. (2024). Fitness Apps—United States | Statista Market Forecast. https://www.statista.com/outlook/hmo/digital-health/digital-fitness-well-being/health-wellness-coaching/fitness-apps/united-states

Sun, J., Ren, Y., Qian, G., Yue, S., & Szumilewicz, A. (2024). Development and application of the integration of physical activity into health care – a scoping review. Annals of Agricultural and Environmental Medicine, 31. https://doi.org/10.26444/aaem/183778

Tahir, A. H., Adnan, M., & Saeed, Z. (2024). The impact of brand image on customer satisfaction and brand loyalty: A systematic literature review. Heliyon, 10(16), e36254. https://doi.org/10.1016/j.heliyon.2024.e36254

Taylor, T. E. (2024). Users and technology: A closer look at how technology engagement affects users. Computers in Human Behavior Reports, 15, 100473. https://doi.org/10.1016/j.chbr.2024.100473

Thijs, I., Fresiello, L., Oosterlinck, W., Sinnaeve, P., & Rega, F. (2019). Assessment of Physical Activity by Wearable Technology During Rehabilitation After Cardiac Surgery: Explorative Prospective Monocentric Observational Cohort Study. JMIR mHealth and uHealth, 7(1), e9865. https://doi.org/10.2196/mhealth.9865

Tong, H. L., Maher, C., Parker, K., Pham, T. D., Neves, A. L., Riordan, B., Chow, C. K., Laranjo, L., & Quiroz, J. C. (2022). The use of mobile apps and fitness trackers to promote healthy behaviors during

COVID-19: A cross-sectional survey. PLOS Digital Health, 1(8), e0000087. https://doi.org/10.1371/journal.pdig.0000087

Wang, C., & Xu, X. (2023). Digital Media Empower The Marketing Promotion of Health Management: A Case Study of KEEP APP. BCP Business & Management. https://doi.org/10.54691/bcpbm.v45i.4906

Yoganathan, D., & Kajanan, S. (2014). What Drives Fitness Apps Usage? An Empirical Evaluation. IFIP Advances in Information and Communication Technology, 429, 179–196. https://doi.org/10.1007/978-3-662-43459-8_12

Authors' and translators' details:

Teguh Arie Sandy	teguharies@ieee.org	Author