Grenze International Journal of Engineering and Technology, June Issue

# Integrating Rule-based Decision-Making and Predictive Modeling for Enhanced user Engagement in Social Networks

Manoj Kumar Srivastav[1], Somsubhra Gupta[2] and Subhranil Som[3]
[1-2]School of Computer Science, Swami Vivekananda University, Barrackpore, West Bengal, India
Email: mksrivastav2015@gmail.com, gsomsubhra@gmail.com
[3]Department of Computer Science, Bhairab Ganguly College, Kolkota, West Bengal, India
Email: subhranil.som@gmail.com

***Abstract*—This paper explores a rule-based decision-making approach to guide user behavior on social network platforms. By applying predefined rules based on user activity data, such as likes, followers, posts, and videos, personalized recommendations are generated to improve engagement and content interaction. The thresholds for these rules are determined using statistical methods, specifically the median of user data attributes. Predictive models, encompassing Logistic Regression (LR), Random Forest (RF), as well as Gradient Boosting, complement rule-based approach to identify high-engagement users. Results demonstrate that complex models such as RF, along with Gradient Boosting, outperform simpler models in predicting user engagement. This framework provides a structured way to analyze and enhance engagement strategies on social media platforms. The novelty of this work is the combination of a rule-based decision system with machine learning models for user engagement prediction. Most existing studies mainly focus on prediction only. The proposed framework not only predicts user engagement but also provides useful recommendations for improving user activity and interaction.**

***Keywords*— Rule-based decision-making, engagement, recommendations, Logistic Regression, Random Forest, Gradient Boosting, predictive model.**

## I. INTRODUCTION

Social networks can be effectively modeled using graph theory, where individuals or entities are demonstrated as nodes, along relationships or interactions have been depicted as edges [1]. Rule-based decision-making in social networks involves using predefined rules or algorithms to guide actions or responses within the network [2]. These rules are typically based on certain criteria or conditions and are applied automatically or semi-automatically to make decisions. Rules can be formulated based on specific conditions or thresholds of the user attributes. Social networks have changed how people interact and share information. These platforms, characterized by their vast networks of interconnected users, generate an immense amount of data through activities like sharing, liking, commenting, as well as following. Analyzing this data offers valuable insights into user behavior along with engagement patterns, critical for enhancing user experiences as well as optimizing platform functionalities.

*Grenze ID: 01.GIJET.12.2.566_26*

Rule-based decision-making, a structured approach to generating recommendations, is crucial in understanding user roles within social networks. These recommendations are designed to boost user engagement and interaction. These rules are typically derived from user data attributes such as likes, followers, posts, and content shares, enabling targeted strategies to encourage specific user behaviors.

The objective of current investigation is to explore efficiency of rule-based decision-making approach for generating recommendations that enhance user engagement on social networks. The approach uses statistical methods to establish thresholds for user activity metrics and applies predefined rules to provide actionable suggestions. Additionally, machine learning (ML) models encompassing RF, LR, as well as Gradient Boosting have been employed for predicting high engagement users, thereby complementing the rule-based system with a predictive framework. Most existing studies in social network analysis use either machine learning models or rule-based systems. Very few studies combine both methods together. The present study combines rule-based recommendations with machine learning models for better prediction and understanding of user engagement.

This paper presents a comprehensive methodology that integrates rule-based logic with predictive modeling to analyze user engagement. By combining insights from user activity data with advanced ML techniques, investigation provides robust framework for understanding user behavior as well as enhancing their experience on social media platforms. This work pertains to enhancing field of social network analysis by demonstrating the way data-driven approaches can improve decision-making and engagement strategies. The objective of the approach is to use predefined rules based on user activity metrics. These rules are applied to provide personalized recommendations. The goal is to improve user engagement over social media platforms. Current research aims to examine efficacy of rule-based decision-making approach in generating recommendations to enhance user engagement. By leveraging user activity data such as likes, followers, posts, and videos and integrating ML models, the research offers robust framework for analyzing and improving user engagement over social media platforms.

The main contribution of this paper is the development of a hybrid framework that combines rule-based recommendations with machine learning models for user engagement prediction. The proposed approach provides both understandable recommendations and better prediction performance.

## II. Literature Review

### *A. Social Network Analysis (SNA)*

Social network analysis (SNA) examines relationships between entities in a network, offering insights into interaction dynamics and information flow. According to Wasserman and Faust (1994), SNA enables researchers to study connections between individuals, groups, or organizations, which are critical in understanding online social platforms [1]. Boyd and Ellison (2007) reported social network sites (SNS) as web-based services allowing people for creating profiles, connect with others, as well as share content [3]. Such platforms serve as environments for individuals to cultivate and maintain relationships, which can influence their behavior both online and offline. Ellison et al. (2014) further expanded on social capital concept within SNS, emphasizing how users leverage these platforms to accrue social resources such as support, trust, and information, which in turn affect their interactions and engagement [4].

### *B. Threshold-Based Approaches*

Thresholds derived from statistical methods, such as the median or mean, are used to categorize user activity. This categorization helps in identifying users who are highly engaged. These thresholds ensure recommendations are based on typical user behavior. For example, user activity metrics like likes, comments, and reposts are compared against these thresholds to determine whether a user is highly engaged or less active.

### *C. Factors Influencing User Engagement*

User engagement over social media platforms is effected by several factors, encompassing demographic characteristics, user features, as well as decision-making processes.

- Demographics and User Features: Toraman et al. (2022) highlighted the significance of user demographics, textual content, and interaction patterns in predicting social engagements [5]. Their comparative analysis of Twitter data showed that both user and textual features are crucial in understanding engagement dynamics. Sorensen et al. (2023) examined consumer behavior on social networks, identifying factors such as perceived risk, trust, as well as social support as key drivers of online engagement along with consumer satisfaction [6]. These outcomes highlight significance of understanding demographic as well as contextual features in shaping user behavior.

- Decision-Making in Social Networks: The decision-making processes that drive user interactions with content are another key area of research. Burke et al. (2010) explored how personal interests, social norms, and network characteristics influence user decisions to engage with content on Facebook [7]. They found that these factors played significant roles in determining the level of user engagement. Similarly, Lampe et al. (2010) discussed how privacy concerns and platform features affect users' decisions regarding the visibility of their communications, highlighting the trade-offs users make between sharing content and protecting their privacy[8].

*D. Machine Learning Approaches for Engagement Prediction*

Recent advancements in ML have provided valuable insights into predicting user engagement over social media platforms. Various techniques have been applied to analyze user activity and interaction patterns.

- Techniques and Models: Zhu et al. (2020) compared ML models, encompassing RFs as well as Gradient Boosting Machines, for predicting user interactions on WeChat[9]. They incorporated user demographics, activity data, and textual content to achieve high prediction accuracy. Cheng et al. (2018) employed deep learning (DL) methods, particularly Long Short-Term Memory (LSTM) networks, to analyze and predict user engagement patterns on Twitter. Their work demonstrated the power of neural networks in processing large and complex datasets, offering a more effective means of predicting engagement compared to traditional models [10].
- Specific Engagement Types: Ku et al. (2019) focused on predicting Boolean comments (yes/no engagement) on social media posts [11]. Their study identified content relevance and the relationship between the poster and the user as key factors influencing engagement. Similarly, Kim and Lee (2021) utilized decision tree models to predict Boolean comments on Instagram, emphasizing the importance of user attributes such as gender, followers, and posting frequency in enhancing prediction accuracy[12].

*E. Impact of Demographics and Activity*

Both demographic factors and user activity patterns play significant roles in influencing engagement levels on social media.

- Demographic Influences: Brandtzaeg and Heim (2011) examined how gender affects user engagement, finding that gender differences influenced the likelihood of users commenting on or liking posts [13]. Liu and Guo (2022) explored the impact of demographic factors on user participation in Chinese online social networks. Their research highlighted cultural and regional differences in social media usage patterns, showing that these factors significantly shape engagement dynamics [14].
- Activity-Based Attributes: Gao et al. (2016) analyzed user activity metrics, such as posting frequency, comments, and likes, to predict engagement with new content. Their investigation suggested that active users were likely to engage with as well as share new posts, thus influencing overall engagement dynamics on the platform. This supports the notion that users who consistently engage with content are more likely to contribute to future interactions, thereby shaping the overall engagement landscape [15].

The literature on user engagement in social networks highlights the importance of understanding various factors that influence user behavior and interactions. Demographics, user features, and decision-making processes all play key roles in shaping engagement patterns. ML approaches, particularly DL techniques, have proven effective in predicting engagement, offering valuable insights into user behavior. Additionally, demographic and activity-based attributes have been shown to significantly impact user participation and engagement. As field is emerging, future research might concentrate on integrating these factors into more sophisticated models to improve the accuracy of engagement predictions and enhance the user experience over social media platforms. Previous studies have shown that machine learning models such as Random Forest and Gradient Boosting are useful for engagement prediction. However, these models are difficult to interpret. Rule-based systems are easier to understand, but their prediction performance is lower. Therefore, the present study combines both approaches in a single framework.

## III. METHODOLOGY

Methodology for this investigation includes analyzing user activity data from a social media platform to generate personalized recommendations aimed at improving user engagement. Following steps highlights process employed to derive insights as well as create recommendations:

*A. Data Collection: The dataset is collected through Kaggle[16]*

https://www.kaggle.com/datasets/alextar04/vk-social-network-dataset?select=newDataBaseVk.csv

The dataset, newDataBaseVk.csv, contains user-specific activity metrics, encompassing:
- countLikesPhotoes: Number of likes a user receives on their photos.
- countFollowers: Number of followers a user has.
- countOwnerReposts: Number of reposts made by user.
- countVideos: Number of videos uploaded by user.
- countFriends: Number of friends a user has.
- boolComments: A binary variable indicating whether the user comments on their own posts.

### *B. Threshold Calculation*

Median values for each of the above-mentioned attributes are calculated to serve as thresholds for generating recommendations. The median values represent the central tendency of each attribute in the dataset, ensuring that recommendations are based on typical user behavior:
- Likes on Photos: The median value of countLikesPhotoes.
- Followers: The median value of countFollowers.
- Reposts: The median value of countOwnerReposts.
- Videos: The median value of countVideos.
- Friends: The median value of countFriends.

These thresholds help determine if a user's activity is above or below average in each category.
Let $X = \{x1, x2, ..., xn\}$ represent user activity data.
For each feature j, the threshold Tj is defined as:
$Tj = \text{median}(Xj)$
A binary indicator is defined as:
$Rij = 1$ if $xij > Tj$, otherwise 0

### *C. Recommendation Logic*

A set of six rules is applied to each user based on their activity levels in comparison to the calculated thresholds. The rules are as follows:
Rules for User Engagement Recommendations
Rule 1: High Likes on Photos
- Condition: If the number of likes on the user's photos (countLikesPhotoes) is greater than the median threshold for likes on photos.
- Recommendation: Engage with more photo posts.

Rule 2: Low Followers
- Condition: If the number of followers (countFollowers) is less than the median threshold for followers.
- Recommendation: Increase social interactions.

Rule 3: Commenting on Own Posts
- Condition: If the user comments on their own posts, indicated by the boolComments column (True/False).
- Recommendation: Reply to comments to increase conversion.

Rule 4: High Reposts
- Condition: If the number of reposts (countOwnerReposts) is greater than the median threshold for reposts.
- Recommendation: Create more repostable content.

Rule 5: High Video Content
- Condition: If the number of videos posted (countVideos) is greater than the median threshold for videos.
- Recommendation: Share more video content.

Rule 6: Large Friend Network
- Condition: If the number of friends (countFriends) is greater than the median threshold for friends.
- Recommendation: Use your friend network to boost engagement.

For each user, the applicable recommendations are concatenated and stored in a new column, Recommendations. If no rule applies, the recommendation is marked as "No recommendation."
The median is used as a threshold due to its robustness against outliers, which are common in social media datasets characterized by skewed distributions.

### *D. Data Visualization:*

- The distribution of user activity is visualized using histograms with Kernel Density Estimation (KDE) overlays for the following attributes:
    - Likes on photos (countLikesPhotoes).

- Followers (countFollowers).
- Reposts (countOwnerReposts).
- Videos (countVideos).
- Friends (countFriends).

These visualizations allow for the observation of trends in user activity and highlight the general patterns across the dataset.

➢ The impact of each recommendation is visualized by counting how many users are affected by each rule. This is done by:
  - Splitting the recommendations in each row by commas.
  - Exploding the resulting list to count the frequency of each recommendation across the dataset.
  - Displaying the results in a bar chart to show how many users are affected by each recommendation.

### *E. Saving the Updated Dataset*

After applying the rules and generating recommendations, updated dataset has been saved to a new CSV file, updated_dataset_with_recommendations.csv. This file includes a new column, Recommendations, which contains personalized advice for each user.

### *F. Tools and Libraries*

➢ Pandas: Employed for data manipulation as well as applying rules to the dataset.

➢ Matplotlib and Seaborn: Used for visualizations, including histograms as well as bar charts, to explore and display user activity patterns and the distribution of recommendations.

This methodology provides a systematic approach to analyzing social media user behavior and generating personalized recommendations to enhance user engagement. By applying specific rules based on thresholds derived from the dataset, the study offers valuable insights into how users can improve their social media presence and interactions.

### *G. Algorithm: Hybrid Engagement Framework*

Step 1: Input user activity dataset
Step 2: Compute median thresholds for all features
Step 3: Apply rule-based conditions to generate recommendations
Step 4: Compute engagement score for each user
Step 5: Label users as high/low engagement
Step 6: Train machine learning models
Step 7: Evaluate performance using standard metrics

## IV. RESULTS

### *A. Data Distribution and Rule Impact*

The distribution of user activity attributes is illustrated in Figure 1.

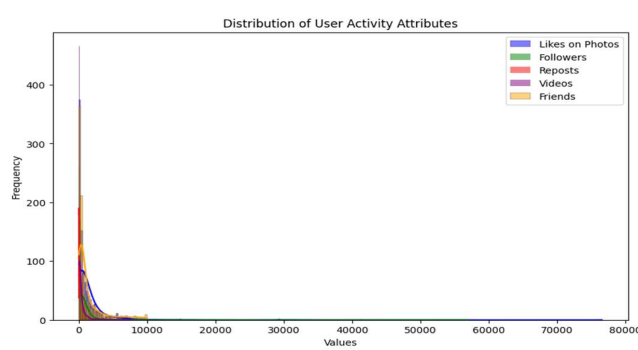


Figure 1: Distribution of User Activity Attributes

The variables such as likes on photos, followers, reposts, videos, and friends exhibit varying degrees of spread, indicating diverse user behavior patterns across the dataset. These distributions provide insight into the central tendencies and variability of user engagement metrics.

Figure 2 presents the impact of rule-based recommendations on user behavior. The frequency analysis of applied rules shows how many users are affected by each recommendation, highlighting the relative importance of different engagement strategies. Some rules are triggered more frequently, indicating common behavioral patterns among users.

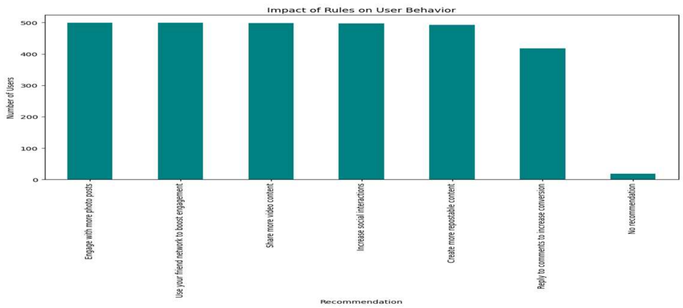


Figure 2: Impact of rules on user behavior

*B. Target Variable Definition*

To establish a ground truth for predictive modeling, users are classified as "high engagement" based on their activity levels. A user is considered highly engaged if they exceed the median threshold in at least three out of the following five metrics: likes on photos, followers, reposts, videos, and friends.

An engagement score Ei is computed as:

$$Ei = \sum Rij \quad (\text{for } j = 1 \text{ to } 5)$$

A user is classified as high engagement if:

$$Ei \geq 3$$

This binary classification (1 for high engagement, 0 otherwise) is used as the target variable for training machine learning models.

*C. Model Setup*

The dataset is divided into training and testing sets using an 80:20 ratio. The input features include user activity metrics such as countLikesPhotoes, countFollowers, countOwnerReposts, countVideos, and countFriends. Additionally, a recommendation-based prediction is considered. If a user receives at least one recommendation, the predicted engagement is set to 1; otherwise, it is set to 0. This allows evaluation of the effectiveness of the rule-based system in predicting engagement. Several machine learning classifiers [17] are employed to predict high engagement users, including Random Forest, Gradient Boosting, Logistic Regression, Naive Bayes, Support Vector Machine, and K-Nearest Neighbors.

*D. Model Performance Results*

The performance of different classification models is evaluated using accuracy, precision, recall, and F1-score.

TABLE I: COMPARATIVE PERFORMANCE ANALYSIS OF CLASSIFICATION MODELS

| Metric | Logistic Regression | Random Forest | Support Vector Machine | K-Nearest Neighbors | Gradient Boosting | Naive Bayes |
|---|---|---|---|---|---|---|
| Accuracy | 0.86 | 0.97 | 0.83 | 0.91 | 0.99 | 0.70 |
| Precision | 0.90 | 0.97 | 0.92 | 0.91 | 1.00 | 0.91 |
| Recall | 0.83 | 0.97 | 0.75 | 0.91 | 0.98 | 0.48 |
| F1-Score | 0.86 | 0.97 | 0.83 | 0.91 | 0.99 | 0.62 |

TABLE II: COMPARATIVE CONFUSION MATRIX ANALYSIS OF CLASSIFICATION MODELS

| Confusion Matrix | Logistic Regression | Random Forest | Support Vector Machine | K-Nearest Neighbors | Gradient Boosting | Naive Bayes |
|---|---|---|---|---|---|---|
| TN | 85 | 92 | 88 | 86 | 95 | 90 |
| FP | 10 | 3 | 7 | 9 | 0 | 5 |
| FN | 18 | 3 | 26 | 9 | 2 | 55 |
| TP | 87 | 102 | 79 | 96 | 103 | 50 |

*E. Comparative Analysis of Models*

Logistic Regression achieves an accuracy of 0.86, with balanced precision and recall values, indicating moderate predictive capability.

Random Forest performs significantly better, achieving an accuracy of 0.97, demonstrating its strength in handling complex feature interactions. Support Vector Machine shows relatively lower performance with an accuracy of 0.83, while K-Nearest Neighbors achieves an accuracy of 0.91, indicating good performance in capturing local patterns in the data.

Gradient Boosting provides the best performance among all models, achieving an accuracy of 0.99 along with high precision and recall values.

In contrast, Naive Bayes performs poorly with an accuracy of 0.70, mainly due to its assumption of feature independence, which does not hold in this dataset.

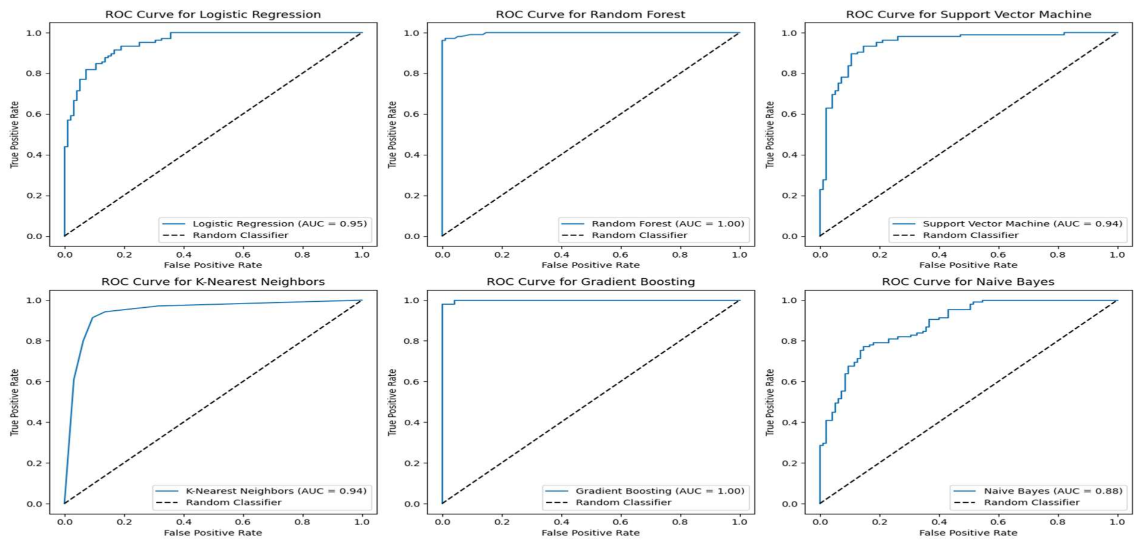


Figure 3: ROC-Curves

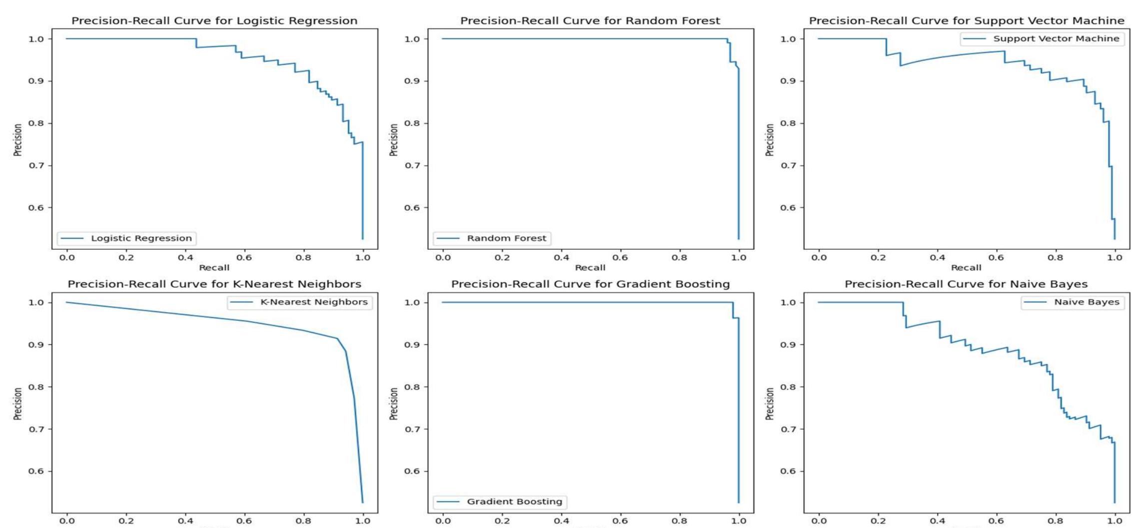


Figure 4: Precision-Recall Curve

Figures 3, 4, and 5 present the ROC curves, precision-recall curves, and comparative performance of all classification models, respectively. These visualizations provide a comprehensive comparison of model performance across multiple evaluation metrics.

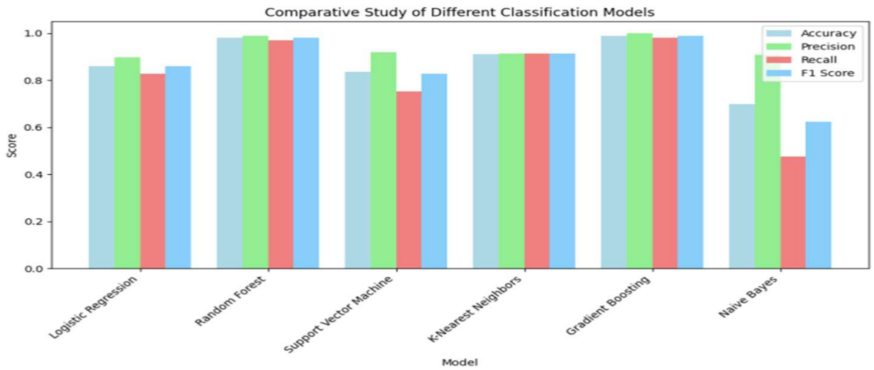


Figure 5: Comparative study of different Classification Models

The reported performance values (0.98–0.99) should be interpreted with caution, as they may not fully reflect the true predictive capability of the model. This is primarily because the engagement labels are generated using the same set of user activity features that are later used for training the machine learning models. Such overlap between feature construction and model input can simplify the classification task, leading to inflated evaluation metrics. While the results demonstrate the effectiveness of the proposed approach, they may overestimate the model's generalization ability. Therefore, further validation using independent datasets or alternative labeling strategies is necessary to ensure robustness and reliability of the findings.

The high accuracy of Random Forest and Gradient Boosting may be due to the rule formation and threshold-based labeling used in this study. Since the labels are created from the same user activity features, strong relationships between the input data and target classes are expected. The main purpose of this work is to compare different machine learning models and test the proposed framework using the available dataset. Additional testing on larger and different datasets was not included in the present study because independent labeled datasets were not available. Future work can include validation on larger datasets and other social media platforms to further examine the performance of the framework.

## IV. Discussion

The proposed framework uses both rule-based methods and machine learning models to study user engagement in social networks. Many earlier studies only use probabilistic models [18]. In this work, rule-based recommendations and machine learning models are used together. This helps improve both understanding and prediction. The rule-based system gives recommendations based on user activities such as likes, followers, reposts, videos, and friends. These recommendations are simple and easy to understand. Machine learning models are also used to identify highly engaged users.

The results show that Random Forest and Gradient Boosting perform better than Logistic Regression, Support Vector Machine, and Naive Bayes. These models give higher accuracy and better prediction results. This shows that user engagement depends on several connected factors. Ensemble models handle these factors more effectively. Another contribution of this study is the use of median-based thresholds for recommendation generation. The median is less affected by extreme values in social media data. This makes the recommendation process more reliable. The study also shows that rule-based systems and machine learning models can work together for social network analysis. The study also has a limitation. The engagement labels are created from the same features used for model training. This may increase the performance values. Therefore, future work should test the framework on other datasets and different labeling methods. Even with this limitation, the study shows that combining rule-based methods with machine learning is useful for user engagement analysis.

## V. Conclusion

The methodology successfully integrates rule-based decision-making with predictive modeling for enhancing user-engagement over social media platforms.

Rules help improve user-engagement and behavior on the platform. Here are the benefits of each rule:

- To help users with popular photos get more engagement.
- To help users with few followers get more visibility.
- To encourage users to respond to comments and start conversations.

- To motivate users to create shareable content.
- To promote sharing videos to get more attention.
- To suggest using personal connections to boost engagement.

The results demonstrate that complex models like RF, as well as Gradient Boosting outperform simpler models such as LR and SVM, providing better accuracy, precision, recall, and overall performance. These outcomes highlight significance of selecting an accurate model for predicting high engagement based on user activity metrics.

Future Scope: Future work can use larger datasets from different social media platforms. More user activity features can also be added for better analysis. Deep learning methods may be used to improve prediction results. The framework can also be tested on real-time social media data. Future studies can further check the performance of the proposed system using independent datasets.

REFERENCES


[1] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge University Press, 1994.

[2] M. K. Srivastav, S. Gupta, and S. Som, “Utilizing a decision-centric framework in social networks to predict user engagement with boolean comments,” *Indian Journal of Natural Sciences*, vol. 15, no. 88, pp. 89384–89393, Feb. 2025.

[3] D. Boyd and N. B. Ellison, “Social network sites: Definition, history, and scholarship,” *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.

[4] N. B. Ellison, J. Vitak, R. Gray, and C. Lampe, “Cultivating social resources on social network sites: Facebook relationship maintenance behaviors and their role in social capital processes,” *Journal of Computer-Mediated Communication*, vol. 19, no. 4, pp. 855–870, 2014, doi: 10.1111/jcc4.12078.

[5] F. Toraman, F. Şahinuç, E. H. Yilmaz, and I. B. Akkaya, “Understanding social engagements: A comparative analysis of user and text features in Twitter,” *Social Network Analysis and Mining*, vol. 12, no. 47, pp. 1–10, 2022, doi: 10.1007/s13278-022-00872-1.

[6] D. Sorensen, D. Vogler, S. Fürst, and M. S. Schäfer, “Platforms matter: Analyzing user engagement with social media content of Swiss higher education institutions,” *Journal of Marketing for Higher Education*, pp. 1–20, 2023, doi: 10.1080/08841241.2023.2289009.

[7] M. Burke, C. Marlow, and T. Lento, “Social network activity and social well-being,” in *Proc. 28th Int. Conf. Human Factors in Computing Systems (CHI '10)*, Atlanta, GA, USA, 2010, pp. 1909–1912, doi: 10.1145/1753326.1753613.

[8] C. Lampe, R. Wash, A. Velasquez, and E. Ozkaya, “Motivations to participate in online communities,” in *Proc. ACM Conf. Computer Supported Cooperative Work (CSCW)*, 2010, pp. 77–86, doi: 10.1145/1718918.1718931.

[9] T. Zhu, Z. Wang, and X. Zhou, “Predicting user interactions in social media: A comparative study of random forests and gradient boosting machines,” in *Proc. 24th Pacific Asia Conf. Information Systems (PACIS 2020)*, 2020.

[10] Y. Cheng, Y. Liu, and C. Dale, “Long short-term memory neural networks for user behavior prediction in Twitter,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 6, pp. 1750–1760, 2018, doi: 10.1109/TCBB.2018.2867437.

[11] W. Ku, M. C. Chen, and X. Zhang, “Factors influencing users' likelihood to comment on social media posts: An empirical study on Instagram,” *Information Systems Frontiers*, vol. 21, no. 6, pp. 1303–1316, 2019, doi: 10.1007/s10796-018-9874-3.

[12] J. Kim and J. Lee, “Predicting comment behavior on Instagram: An analysis using decision trees,” *Social Network Analysis and Mining*, vol. 11, no. 1, Art. no. 25, 2021, doi: 10.1007/s13278-021-00736-y.

[13] P. B. Brandtzaeg and J. Heim, “Exploring the relationship between online social network involvement and satisfaction with life: Preliminary results from a Norwegian sample,” *International Journal of Human-Computer Studies*, vol. 69, no. 10, pp. 785–793, 2011, doi: 10.1016/j.ijhcs.2011.07.004.

[14] Y. Liu and Y. Guo, “Influence of gender differences on online social network participation: A study from China,” *Internet Research*, vol. 32, no. 1, pp. 119–141, 2022, doi: 10.1108/INTR-10-2020-0503.

[15] X. Gao, J. Tang, and L. Hong, “Modeling user activity preference by leveraging user-generated content in social networks,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 5, pp. 1212–1225, 2016, doi: 10.1109/TKDE.2016.2516907.

[16] A. Taran, “VK Social Network Dataset,” *Kaggle*. [Online]. Available: https://www.kaggle.com/datasets/alextar04/vk-social-network-dataset

[17] S. Sridhar and M. Vijayalakshmi, *Machine Learning*. Oxford, U.K.: Oxford University Press, 2021.

[18] M. K. Srivastav, S. Gupta, and S. Som, “Modeling friendship dynamics in social networks: A graph-theoretic and computational approach,” *Journal of Information Systems Engineering and Management*, vol. 10, no. 37s, 2025, doi: 10.52783/jisem.v10i37s.6382.