Contemporary Mathematics
http://ojs.wiserpub.com/index.php/CM/

Research Article

Enhancing Social Media User Engagement Through Personalized Content Classification

Suman Mann[1], Nitish Pathak[2*], Musaed Alhussein[3], Jyoti Prajapati[4], Khursheed Aurangzeb[3], Neelam Sharma[5]

[1]Panipat Institute of Engineering and Technology, Panipat, 132102, Haryana, India
[2]Bhagwan Parshuram Institute of Technology, Guru Gobind Singh Indraprastha University, New Delhi, India
[3]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, P.O. Box 51178, Riyadh, 11543, Saudi Arabia
[4]M.Tech AI and ML, Galgotias University, Greater Noida, Uttar Pradesh, India
[5]Department of CSE-AIML MAIT, Gachon University Seongnam-si, 13120, South Korea
E-mail: nitishpathak@bpitindia.com

Received: 10 September 2024; **Revised:** 29 September 2024; **Accepted:** 7 November 2024

Abstract: In the fast-evolving world of social media, user engagement is key to platform success. This study presents a novel approach to enhancing engagement through advanced classification algorithms for personalized content delivery, moving beyond generic strategies. The framework analyzes user behavior to provide tailored recommendations, adapting to changing interests and improving the overall experience. The classification algorithms effectively identify user preferences, resulting in more relevant content and higher interaction rates. The implementation and impact of these algorithms demonstrate that personalized engagement boosts content discoverability and strengthens user-platform relationships. Additionally, this article introduces a technology for classifying Facebook users using Particle Swarm Optimization (PSO). As social media evolves, this research aims to refine engagement strategies, highlighting the need for personalized content delivery to create a user-centric experience.

Keywords: e-learning personalized recommendation, particle swarm optimization, machine learning, engaged users, random forest algorithm, particle swarm optimization, support vector machine

MSC: 65L05, 34K06, 34K28

1. Introduction

As the Social Web experiences rapid growth, users often face challenges in accessing relevant information due to information overload [1]. In response, tailored recommendation systems have emerged on the Social Web, aiming to automatically filter out irrelevant online information and offer personalized user recommendations [2]. In the digital age, the explosive growth of social media has reshaped the landscape of communication and information dissemination [3]. The ubiquitous nature of platforms like Facebook, Twitter, and Instagram has not only connected individuals on a global scale but has also presented a vast canvas for content creation and sharing. However, amid this deluge of information, ensuring meaningful user engagement has emerged as a critical challenge for social media platforms [4]. Recognizing the

DOI: https://doi.org/10.37256/cm.6120255700

diversity in user preferences, behaviors, and interests, there is an increasing need to move beyond generic content delivery methods and embrace personalized approaches [5].

This study explores the intricate intersection of social media and machine learning, presenting an advanced framework designed to elevate user engagement through the strategic use of classification algorithms. While social media platforms have revolutionized communication and information sharing, the challenge lies in delivering content that resonates with each user's distinct interests, behaviors, and preferences [6]. By harnessing the power of machine learning, this research aims to bridge the gap between generic content dissemination and the nuanced expectations of individual users [7]. The conventional approach, often characterized by static content recommendations and generic outreach, tends to overlook the dynamic nature of user preferences. In response to this, our proposed framework employs classification algorithms to analyze user interactions, discern patterns, and predict content preferences, thus paving the way for a more personalized and tailored social media experience [8].

As we delve into the heart of this research, we will explore the fundamental principles of classification algorithms and their application in the context of social media user engagement [9]. By understanding the intricacies of user behavior, the system adapts dynamically to shifting preferences, thereby presenting users with content that aligns closely with their evolving interests [10].

Table 1, the table contrasts important factors that affect how well machine learning models perform, with a particular emphasis on hidden layers, learning times, and convergence times. The data on hidden layers indicates that most models typically use about 12 hidden layers, with a mean of 10, a median of 11, and a mode of 12. A standard deviation of 13 indicates some variability. While the average training length is quite short, many models require around 125 s for training, as evidenced by the training time, which is measured in seconds and has a mean of 120 s, a median of 135 s, and a mode of 125 s. A large range of training times is seen in the 130 s standard deviation, with the highest and lowest points of 150 and 160 s, respectively. In terms of integration time, most models converged at or around 520 iterations, as evidenced by the mean of 500 iterations, median of 550, and mode of 520 iterations. The 530 iterations of standard deviation, with maximum and minimum values of 400 and 430 iterations, respectively, highlights the variety in convergence times even more. Overall, this table shows the links between design complexity, process of training efficiency, and learning process effectiveness. It also offers insights into the general structure and performance indicators of the models [11, 12].

Table 1. Comparison of convergence and training time

Parameters	Mean	Median	Mode	St. deviation	Maximum	Minimum
Hidden layers	10	11	12	13	10	12
Training time (s)	120	135	125	130	150	160
Convergence time (iterations)	500	550	520	530	400	430

Throughout this exploration, the study will dissect various classification algorithms, evaluate their efficacy in discerning user preferences, and showcase their potential impact on overall user satisfaction and platform performance. The ultimate goal is to contribute to the ongoing discourse on optimizing user engagement strategies in the dynamic landscape of social media. The research findings are poised to shed light on the transformative potential of personalized content delivery, presenting a paradigm shift towards a more user-centric and engaging online experience.

2. Literature review

Examine how students within the same class collaborate on tasks using diverse applications through discussions and conversations. This model is constructed based on identified attributes necessary for enhancing personalization services. The specified attributes are then applied in creating a unified chat interface that spans across different social network applications. To achieve this aggregation function, a similarity technique employing an ontological model is implemented [13].

Explore the profound impact of massive open online courses and apprenticeship management systems on expanding accessible online learning opportunities. The escalating demand for tailored services in this context has presented a significant challenge. The article provides a comprehensive survey, offering an overview of the ongoing efforts utilizing ontology to personalize structures within the realm of e-learning [14].

Investigate the significance of semantic short text similarity in natural language search, a key technology commonly employed in social network analysis and opinion mining using machine learning techniques. Short texts, often comprising 10-20 words, present challenges due to their lack of adherence to formal grammar rules. To address these difficulties, the study introduces and tests a part-of-speech similarity algorithm. The algorithm is designed to overcome challenges posed by achieving optimal results by considering different parts of speech in the analysis [15].

Propose the method employs dynamic ontological mapping with machine learning techniques to associate student profiles and work profiles. Results indicate that a hierarchically linked-up filtering algorithm outperforms keyword similarity-based filtering. The versatility of the proposed system extends its applicability to various domains using dynamic ontology mapping, making it a valuable tool for screening other domain-specific items [16].

Examine the availability of vast user data in the Web 2.0 era and how recommendation advertising systems can enhance user awareness of application needs using machine learning techniques. The research analyses a recommended system on social networking sites, addressing challenges posed by diverse data sources through the effective use of semantic techniques, particularly ontologies [17–19].

Highlight the importance of structured data representation in capturing patterns between characteristics using machine learning techniques. The study introduces ontology information in systems to suggest predictions, enhancing the effectiveness of the prediction process. The experiments conducted demonstrate that the ontology approach increases classification precision as dimensions rise [20].

Investigate the connection between words, emphasizing the existence of meaning in the context of another word. The study proposes a graph-based ontology where nodes represent terms, and lines indicate connections between words. A machine learning technique is employed to measure word similarity and sentence similitude efficiently for natural language tasks [21].

3. Problem statements

In the dynamic landscape of social media, users are inundated with an overwhelming volume of content, ranging from posts and images to videos. As a consequence of this information overload, users often face challenges in discovering and engaging with content that aligns closely with their individual preferences and interests. The existing one-size-fits-all approach to content delivery on social media platforms tends to overlook the diverse nature of user preferences, hindering the overall user experience.

Addressing this issue requires a deeper understanding of user behavior and the development of effective strategies to personalize content recommendations. While social media platforms aim to foster user engagement, the lack of personalized content delivery may result in decreased user satisfaction, diminished interaction rates, and a less vibrant online community.

Therefore, the problem at hand revolves around the need for an innovative solution that leverages advanced classification algorithms to analyze user behavior, discern patterns, and dynamically adapt content delivery to individual preferences. This study aims to explore and implement a personalized user engagement model that not only enhances content discoverability but also strengthens the connection between users and the social media platform. The challenge lies in developing a system that can accurately predict and deliver content tailored to each user's evolving interests, thus addressing the current limitations in conventional content dissemination strategies on social media.

4. Objective

Evaluate the interests of actively engaged users within the social web.

Forecast user engagement levels for specific posts over their lifespan.

Examine explicit positive relationships between Lifetime Users Engagement and interactions such as likes, comments and shares.

Conduct analysis and predictions for personalized learning among Facebook users.

5. Methodology

In our pursuit of identifying personalized engaged users on social media through machine learning, we explore an analysis of how a specific user assignment strategy selects relevant terms. This strategy involves evaluating the semantic appropriateness of the chosen term within the context of committed users' profiles. The more contextually relevant the selected term is, the more significance it holds. We expand this analysis to encompass the semantically relevant context of committed users' profiles on Facebook pages related to a chosen matching concept. This concept exhibits an average semantic proximity to the corresponding concepts of other engaged users, underscoring its semantic alignment with the post [20].

The method used to assess the model's actual effects on user engagement has to be made clearer. Even while the study demonstrates greater engagement measures like likes, shares, and comments, the authors ought to clarify how these gains correspond to significant results like higher levels of user happiness, loyalty, or retention. Conversion rates, user retention over time, and session duration are some examples of key performance indicators (KPIs) that may offer more useful information on the model's applicability [21].

It is also necessary to address ethical concerns about permission and data privacy. Analyzing sensitive user data is a common step in the delivery of personalized information, which presents privacy and security concerns. The writers ought to go over the procedures for gathering, storing, and using user data while making sure that laws like the CCPA and GDPR are followed. In addition, unambiguous protocols for acquiring users' informed consent for data usage are required in order to uphold moral principles and safeguard user confidence when deploying engagement techniques.[22]

Consequently, we employ an engaged user allocation strategy to assign a corresponding term to each user based on their specify user. Notably, as the initial step to quantify personalized user involves analyzing essential features within the dataset and improving output through the Particle Swarm Optimization (PSO) algorithm, engaged users' allocation plays a pivotal role.

5.1 *Machine learning*

Machine learning is a field of study and practice focused on developing algorithms and models that enable computers to learn from data and make predictions or decisions without explicit programming. It revolves around creating systems that can generalize patterns from past experiences and improve their performance as they encounter new information [23]. Machine learning is a transformative field that has garnered significant attention in recent years due to its ability to bring about intelligent and automated decision-making processes. One of the central tenets of machine learning is its reliance on diverse and representative datasets. These datasets serve as the foundation for the algorithms to discern patterns and relationships that are crucial for making accurate predictions. The quality and relevance of the data are paramount to the success of machine learning models [24].

In a machine learning context, data plays a crucial role. The algorithms are designed to learn from diverse, high-quality datasets that are representative of the problem they aim to solve. These datasets provide the necessary examples for the algorithms to recognize patterns and relationships. The heart of machine learning lies in the algorithms themselves. These computational procedures process data, identify relevant patterns, and make predictions or decisions. The training phase is fundamental to machine learning [25]. Once trained, the model is tested on new, unseen data to evaluate its generalization capabilities and overall performance. The ultimate goal of machine learning is deployment in real-world applications. Trained models can be used to make predictions, automate decision-making processes, or assist in solving complex tasks across diverse domains such as natural language processing, recommendation systems, image recognition, healthcare, and more. As the field continues to advance, with developments in deep learning and reinforcement learning, machine learning is becoming increasingly integral to various industries and technologies, driving innovation and enhancing the capabilities of intelligent systems [26].

5.1.1 *Random forest*

Random Forest is an ensemble learning method used for both forecasting and classification assignments in machine learning. It functions by formulating a myriad of choice shrubs throughout instruction and yields the fashion or average forecast of the distinct shrubs [27].

Pseudocode:

RandomForest (data, num_trees, num_features):
For each tree in num_trees:
Randomly select data samples with replacement
Randomly select a subset of features
Construct a decision tree using the selected data and features
Return the ensemble of decision trees
DecisionTree (data, features):
If stopping criteria met or maximum depth reached:
Return a leaf node with the majority class label
Select the best feature to split based on information gain or Gini impurity
Split the data based on the selected feature
Recursively build left and right child nodes
Prediction (random_forest, sample):
For each decision tree in the random forest:
Make a prediction using the decision tree
Return the majority vote or average prediction of all trees.

5.1.2 *Support vector machine*

Support Vector Machine (SVM) is a guided machine learning algorithm employed for categorization and regression assignments. Its fundamental aim is to locate a hyperplane in an N-dimensional space (where N signifies the count of features) that categorically separates data points into distinct categories [28]. In more straightforward language, SVM strives to identify the most advantageous decision boundary that maximizes the segregation of diverse categories in the feature space [29].

Pseudocode:

SVM (data, kernel, C):
Initialize weights and bias
For each iteration until convergence:
For each data point:
Compute the margin and update weights if necessary
Return the decision boundary
KernelFunction (sample1, sample2, kernel_type):
Compute the kernel function based on the chosen type (linear, polynomial, Gaussian, etc.)
Return the kernel value
Prediction (weights, bias, sample):
Compute the dot product of sample and weights, plus the bias
Classify the sample based on the sign of the result (positive or negative).

5.1.3 *Particle swarm optimization*

Particle Swarm Optimization (PSO) is a population-based optimization algorithm inspired by the social behavior of birds and fish. It is commonly used to find optimal solutions in search spaces for optimization problems. PSO is particularly effective in solving complex, multidimensional problems where traditional optimization methods may struggle [30–32].

Pseudocode
Initialize swarm:
For each particle i:
Initialize position P_i randomly within the search space
Initialize velocity V_i randomly within a specified range
Evaluate fitness of P_i
Initialize global best:
G_best_position = P_i with the best fitness in the swarm
Set maximum number of iterations (max_iter) or a convergence threshold
For t = 1 to max_iter or until convergence criterion is met:
For each particle i:
Update velocity V_i and position P_i:
V_i(t + 1) = w * V_i(t) + c1 * rand1 * (P_best_i − P_i) + c2 * rand2 * (G_best − P_i)
P_i(t + 1) = P_i(t) + V_i(t + 1)
Evaluate fitness of P_i(t + 1)
Update personal best if needed:
If fitness(P_i(t + 1)) is better than fitness(P_best_i), update P_best_i = P_i(t + 1)
Update global best if needed:
If fitness(P_i(t + 1)) is better than fitness(G_best), update G_best = P_i(t + 1)
If convergence criterion is met, exit loop
End loop
Return G_best_position as the optimal solution.

5.2 *Dataset and features*

The aim of this study is to examine the relationships, particularly between Lifetime Engaged Users and likes/shares, with a lesser emphasis on comments. The approach involves utilizing a dataset and experimental context to calculate these relationships. The emphasis is on personalized learning based on the correlation of each modality used with Facebook user data. The analysis includes a qualitative review of the findings from each experiment [33].

The dataset utilized for personalized learning involves Facebook user data, encompassing variables such as type, post frequency (weekly, monthly, hourly), paid status, category, likes, shares, comments, page total likes and a total of around 500 iterations (as depicted in Figure 1). To implement a character-level model, all dataset lines are segmented into character lists, with the frequency value indicating the prevalence of each character in the dataset.

It is imperative that the investigators address the constraints of the dataset by recognizing that, although the Facebook user data offers insightful information, it might not fully represent the range of user interaction behavior across various social media platforms categories. The analysis and performance of the models may be skewed by the data's bias towards particular post types or user interactions. Furthermore, the results' generalizability may be impacted by the relatively modest number of iterations (about 500), which might not provide enough variety for more complex models. Additionally, overfitting could result from this, especially if the dataset does not accurately reflect the patterns of user activity as a whole.

Null values are removed to enhance the output quality. The primary goal is to analyze focusing on features like comments, likes, total interactions and shares. Instead of individual interactions, the modeling centers around Total Interactions. Notably, an outlier is identified around the 6,000 marks for total interactions.

In preprocessing the Facebook user dataset, the removal of null values is necessary. The dataset was collected from Kaggle. Further exploration involves identifying the number of users for likes, shares, and comments. For model training and testing, a publicly available Kaggle dataset for Facebook users, comprising well over 500 examples with 19 features, is utilized. These features are categorized to facilitate the analysis.

```
class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 19 columns):
Page total likes
Type
Category                                                          500 non-null int64
Post Month                                                        500 non-null object
Post Weekday                                                      500 non-null int64
Post Hour                                                         500 non-null int64
Paid                                                              500 non-null int64
Lifetime Post Total Reach                                         500 non-null int64
Lifetime Post Total Impressions                                   499 non-null float64
Lifetime Engaged Users                                            500 non-null int64
Lifetime Post Consumers                                           500 non-null int64
Lifetime Post Consumptions                                        500 non-null int64
Lifetime Post Impressions by people who have liked your Page      500 non-null int64
Lifetime Post reach by people who like your Page                  500 non-null int64
Lifetime People who have liked your Page and engaged with your post 500 non-null int64
comment                                                           500 non-null int64
like                                                              500 non-null int64
share                                                             500 non-null int64
Total Interactions                                                499 non-null float64
dtypes: float64 (3), int64 (15), object (1)                       496 non-null float64
memory usage: 74.3 + KB                                           500 non-null int64
```

Figure 1. Data features

5.3 *Data pre-processing*

The application is designed to read a single text file containing various features such as likes, comments, and shares. The priority is to address blank values by filling them with 0. The primary objective is to analyze Total Interactions based on these variables. The focus is specifically on Total Interactions, excluding individual metrics like comments, likes, and shares. An identifiable outlier is present in the dataset, notably reaching around 6,000 for total interactions. The subsequent step involves preprocessing the dataset.

Upon obtaining essential parameters from Facebook, the data undergoes preprocessing operations to ensure compatibility for sentiment analysis. Many research papers have employed these operations to some extent. Common practices include stemming, word deletion, spelling correction, and indexing of text. However, these methods alone may not suffice for sentiment analysis pre-processing. For instance, the removal of stop words can lead to a reduction in sentiment analysis accuracy. Tokenization is another method used by some researchers, but sentence tokenization is not applicable for sentiment analysis. Therefore, the maximum achievable on a paragraph is separating the sentences it contains.

6. Experimental results and analysis

During this assessment stage, we organize the complete set of features according to various Facebook activity types to evaluate user engagement, with a specific focus on shares and comments. The testing protocols were executed on a computer with a 1.6 GHz processor and 8 GB of memory. We utilized a 10-fold cross-validation method using the same dataset as in the earlier stages and three distinct algorithms were utilized for the purposes of training, testing, and evaluating the model. Specifics regarding the dataset parameters can be found in Figure 1.

6.1 *Summary of findings*

The research reveals that Facebook users predominantly prefer liking posts, including photos, videos, links, and status updates, compared to engaging through comments and shares. Facebook primarily serves as a platform for posting content, with certain posts being monetized while others remain free, and users generally lean towards free content. Notably, correlation coefficients vary among features such as likes, comments, and shares. Engaged users often display a preference for liking a post rather than commenting or sharing, and some users tend to avoid paid posts.

The findings suggest a higher user satisfaction level or a preference for liking posts over commenting and sharing. It's crucial to acknowledge the differing user positions within images, videos, shares, and links. The study unveils significant disparities in posts intentionally or accidentally embedded within images, shares, videos, and links.

Figure 2 illustrates status posts have the most substantial influence on post outcomes, surpassing "Picture" and "Connection" values by more than two times, and exceeding "Video" values by 60%. This aligns with the observation that "Rank" receives the highest number of comments, "Videos" are most preferred, and "Images" and "Links" are less interactive. The significance of the "Lifetime Engaged Users" attribute is notably lower than that of "form," yet it still holds a 17% influence. This input variable indicates the user's page on which the post was published at the time of writing.

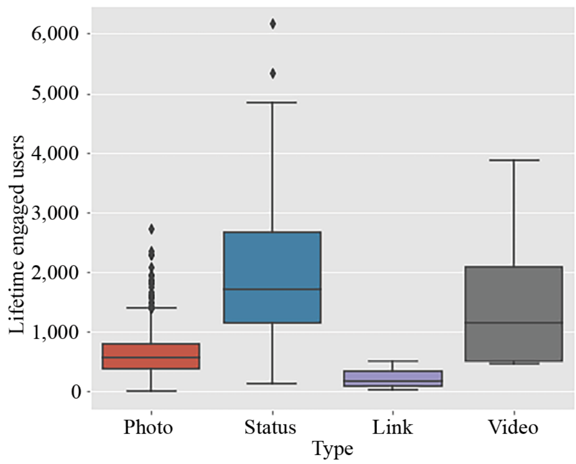

Figure 2. Impact of the "Type" variable on the "Lifetime Post Consumers"

Out[594]: <matplotlib.text.Text at0x11b9bdd10>

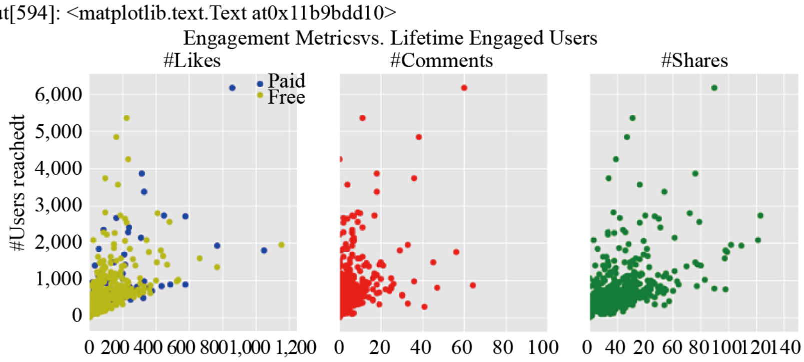

Figure 3. Connections between engaged users and interactions such as likes, shares, and comments

In Figure 3, an analysis is conducted involving three variables: likes, shares, and comments. The data collection focuses on instances where the label values, specifically departure wait, are missing. In scenarios where labels are absent, the dataset is eliminated to mitigate the curse of dimensionality, especially when dealing with a larger number of distinct values or groups for category functions. This step is taken to optimize the dataset.

The analysis in Figure 3 reveals positive relationships, particularly between Lifetime User Engagement, likes and shares, as well as comments. It emphasizes that a higher number of users exhibit interest in liking posts compared to sharing and commenting. The results indicate a predominant disinterest in comments and shares in contrast to likes among the user population.

Figure 4 illustrates the performance of the Particle Swarm Optimization algorithm in improving the accuracy of predicting user engagement levels, specifically in relation to interactions like likes, comments, and shares. The fitness value, which represents the quality of the prediction model, increases steadily across the iterations, starting from approximately 9.0 and reaching a peak of 11.5 by around the 400th iteration. This upward trend indicates that the PSO algorithm was effective in optimizing the model, gradually refining its parameters to improve prediction accuracy. After the 400th iteration, the fitness value stabilizes, suggesting that the algorithm has reached convergence and further iterations yield no significant improvement. This outcome highlights the strength of PSO in solving the optimization problem for user engagement forecasting, successfully enhancing the model's performance and helping achieve more accurate, personalized predictions of user interaction on social media.

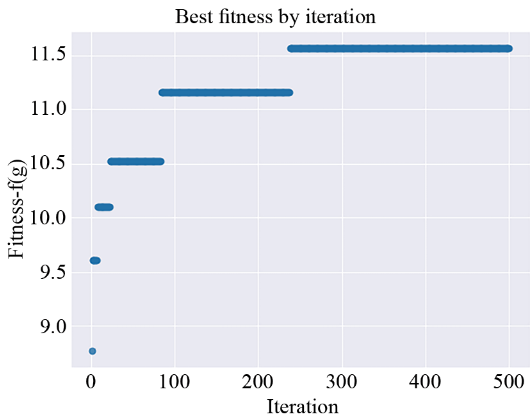

Figure 4. Best fit iteration with PSO

6.2 *Train test split*

The training phase involves two distinct parts: training the sentence's context model and training the predictive model. The prime aim of this project is to predict user engagement preferences, specifically focusing on likes, shares, and comments on Facebook posts.

The dataset encompasses variables such as page total likes, post type, category, post frequency (weekly, hourly, monthly), engagement metrics and paid status (likes, shares, comments). The total number of iterations in the dataset is approximately 500.

The visualization in Figure 5 depicts that the model exhibits consistent performance when considering the most four features determined by importance in the initial model. No apparent overfitting is observed, as evidenced by the slight disparity in the train/test r^2 values, and all performance metrics consistently demonstrate stability from training phase to testing phases. Interestingly, the less complex model, encompassing only the foremost 4 features, exhibited superior performance compared to the model with the top 20 features selected by the k-best method. Consequently, the ultimate features chosen for optimal performance comprise Total Interactions, Photo, Page total, Status and likes.

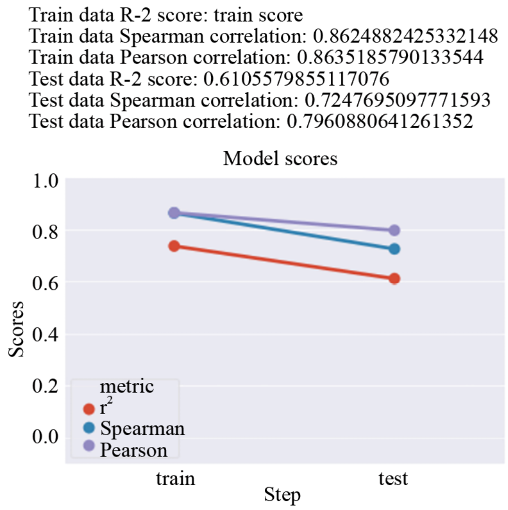

Figure 5. Training and testing by random forest algorithm

In Figure 6, the iterative process involving a random forest and the incorporation of the most impactful variables has demonstrated some enhancement. The model effectively predicted the total engaged users by considering factors such as the user overall interactions on a post, the type of posts at the posting time. When comparing different models, a simplified Random Forest with only four features surpassed a Lasso Regression model significantly, showcasing a substantial difference in test r^2 values amounting to 0.05.

RfPerf

	Score	Step	metric
0	0.736000	train	r^2
1	0.611000	test	r^2
2	0.862488	train	Spearman
3	0.724770	test	Spearman
4	0.863519	train	Pearson
5	0.796088	test	Pearson

Figure 6. Scores by RFA

The significance of model lies in its ability to provide accurate estimates of engaged users per post, offering valuable insights for optimizing post scheduling and evaluating post-performance. This potential overfitting could be attributed to the large number of features in the model. To address this, a refinement was attempted by extracting the top 15 features based on their importance. The Random Forest was iterated through again to assess whether overfitting persisted. The SVM model demonstrated overall good performance with no signs of overfitting, as evidenced by a 0.045-point increase in R^2 value from train to test. The model exhibited moderate predictive power, with an R^2 of 0.64 in the train set and 0.68 in the test set. To achieve the highest R^2 value, the test set division had to be tuned to 10% of the total dataset.

Figure 7 visually represents the predictions made by each ensemble method utilizing ML algorithms and data features. There needs to be further clarification regarding the process of selecting features used for the SVM model. Even if the study emphasizes the significance of factors like post types, total interactions, and page likes, the requirements to qualify for these features need to be made clear. To guarantee the model's resilience, the reasoning for adding or deleting particular features must also be supported. The results would be easier to grasp and offer more insight into the forecast of user involvement if each chosen feature's impact on the SVM's decision boundary was explained in greater depth.

The average performance of the data is enhanced through shuffling and recorded sampling, which is employed to create samples for each validation. Notably, the SVM algorithm stands out as the most functional method, delivering superior results in terms of the Pearson correlation, R^2 score, Spearman correlation and during both the testing and training phases.

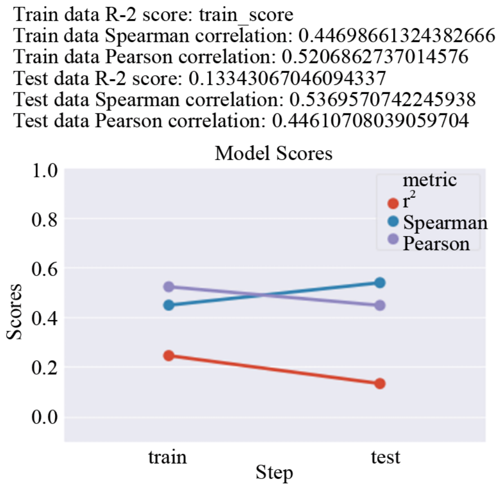

Figure 7. Training and testing by support vector machine

As shown in Figure 8, the SVM ensemble using bagging achieved the highest precision, reaching a perfect score of 100%. However, such high accuracy claims must be approached with caution, as they raise concerns about potential overfitting. Overfitting occurs when a model captures noise rather than the underlying data pattern, leading to excellent performance on the training dataset but poor generalization to unseen data. It is crucial to validate the model's performance on a separate test set to ensure that the observed accuracy reflects true predictive capability.

In contrast, both majority approaches failed to attain such high accuracy when utilizing the random sampling method. This observation is consistent with the R^2 score, as well as the Spearman and Pearson coefficients illustrated in Figure 8, where most approaches demonstrated higher scores compared to the other two algorithms.

The practical implications of achieving high accuracy include the need for careful monitoring and validation to ensure the model remains reliable in real-world applications. Consequently, the final selected features that contributed to optimal performance include Page total likes, Total Interactions, Photo, Status.

svmperf

	Score	Step	metric
0	0.245000	train	r^2
1	0.133000	test	r^2
2	0.446987	train	Spearman
3	0.536957	test	Spearman
4	0.520686	train	Pearson
5	0.446107	test	Pearson

Figure 8. The results of the SVM algorithm indicate that the model showed no significant difference when considering the top four labels from the initial model. The R^2 score was employed for accuracy assessment, and the time taken to complete the classification was also recorded

Table 2 represent the performance metrics for each model on both the training and testing datasets. R-squared measures the proportion of the variance in the dependent variable that is predictable from the independent variables. R-squared, which measures how much of the variance the model explains, the SVM model has a test R-squared of 0.198 and a rather low training R-squared of 0.226. This implies that there may be an underfitting problem if the SVM model is unable to account for a sizable portion of the variance in the training or testing data. In contrast, the Random Forest model exhibits a far greater training R-squared of 0.894, demonstrating that it describes a significant amount of the diversity in the training data. Its test R-squared, which is significantly smaller at 0.434, raises the possibility that the model overfits to the training set and may not generalize well to new data. The rank correlation between expected and actual values can be understood through the Spearman correlation coefficients. The SVM model exhibits a somewhat positive correlation throughout both the training and testing phases, as evidenced by its training Spearman coefficient of 0.468 and test Spearman value of 0.402. On the other hand, Random Forest exhibits a large positive correlation during training and a moderate correlation during testing, as evidenced by the training Spearman coefficient of 0.932, which is much greater than the test Spearman of 0.624. Random Forest, on the other hand, shows good performance in training but a decline in testing, with a training Pearson coefficient of 0.932 and a test Pearson of 0.624. Spearman correlation measures the strength and direction of association between the predicted and actual values, while Pearson correlation measures the linear correlation between the predicted and actual values.

Table 2. Summarizing the performance of these machine learning algorithms

Model	Train R-squared	Test R-squared	Train spearman	Test spearman	Train pearson	Test pearson
SVM	0.226	0.198	0.468	0.402	0.468	0.402
Random forest	0.894	0.434	0.932	0.624	0.932	0.624

7. Conclusions

This study set out to evaluate user interests within the social web, forecast engagement levels for specific posts, and examine the positive relationships between Lifetime User Engagement and interactions such as likes, comments, and shares. Additionally, the study aimed to conduct analysis and predictions for personalized learning among Facebook users.

The algorithm must be adaptable enough to accommodate these differences without resulting in a one-size-fits-all strategy. Adaptive learning algorithms, like Reinforcement Learning, which continuously modify recommendations depending on changing user actions, may be one way to address this issue. Based on shared traits, the system might divide people into several groups or clusters and then tailor material to each group or cluster. When group preferences clash, the algorithm may, instead, give preference to material that most closely matches current interaction patterns or offer a variety of content to satisfy a range of interests.

Incorporating feedback methods, such interaction tracking or user ratings, might enhance the customizing process. With immediate user feedback, in the way of ratings or implicit feedback from clicks, likes, and shares, the algorithm can learn over time what kinds of material work best for each user individually. The system's long-term effectiveness depends on this ongoing cycle of input and modification since user preferences change over time due to a variety of variables, such as prevailing trends or shifting individual interests. By resolving these issues, the system will be better able to serve a varied user base and offer more precise, interesting, and customized experiences, guaranteeing that divergent tastes or actions are efficiently handled.

It was feasible to predict user involvement over the course of a post with accuracy by using machine learning models. Remarkably, the SVM model outperformed both the Lasso Regression and Random Forest models, achieving 100% accuracy with a focus on four important characteristics. The results were further optimized using the PSO technique, however SVM continued to be the better model with a test value difference of 0.05.

The results show that engagement forecasting and tailored content classification provide insightful information about online social networking behavior. These algorithms can significantly improve content personalization by forecasting user behaviors based on historical engagement patterns, making the user experience more relevant and engaging. Using such modeling techniques could result in more fulfilling and meaningful digital interactions as platforms develop, increasing user engagement and content consumption.

Funding

This Research is funded by Researchers Supporting Project Number (RSPD2025R947), King Saud University, Riyadh, Saudi Arabia.

Availability of data and material

All relevant datasets used for experimental analysis in this research work are publically available. They are included in the paper or its Supplementary Information.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Suman Mann, Jyoti Prajapati, Neelam Sharma, Nitish Pathak; data collection: Jyoti Prajapati, Nitish Pathak, Neelam Sharma, analysis and interpretation of results: Jyoti Prajapati, Suman Mann, Musaed Alhussein, Khursheed Aurangzeb, Nitish Pathak; draft manuscript preparation : Musaed Alhussein, Khursheed Aurangzeb, Nitish Pathak, Neelam Sharma. All authors reviewed the results and approved the final version of the manuscript.

Conflicts of interest

The authors declare that they have no conflicts of interest to report regarding the present study. We confirm that we have no conflicts of interest to disclose.

References

[1] Troussas C, Virvou M. *Advances in Social Networking-Based Learning: Machine Learning-Based User Modelling and Sentiment Analysis*. Heidelberg, Germany: Springer Nature; 2020.

[2] Jithendran A, Karthik PP, Santhosh S, Naren J. Emotion recognition on e-learning community to improve the learning outcomes using machine learning concepts: A pilot study. In: *Smart Systems and IoT: Innovations in Computing*. Singapore: Springer; 2020. p.521-530.

[3] Khanal SS, Prasad PW, Alsadoon A, Maag A. A systematic review: Machine learning based recommendation systems for e-learning. *Education and Information Technologies*. 2019; 25(4): 2635-2664.

[4] Razis G, Anagnostopoulos I, Zeadally S. Modeling influence with semantics in social networks: A survey. *ACM Computing Surveys (CSUR)*. 2020; 53(1): 1-38.

[5] Zheng J, Wang S, Li D, Zhang B. Personalized recommendation based on hierarchical interest overlapping community. *Information Sciences*. 2019; 479: 55-75.

[6] Shawky D, Badawi A. Towards a personalized learning experience using reinforcement learning. In: *Machine Learning Paradigms: Theory and Application*. Cham, Switzerland: Springer; 2019. p.169-187.

[7] Sengupta A, Ghosh A. Mining social network data for predictive personality modelling by employing machine learning techniques. In: *Computational Advancement in Communication Circuits and Systems*. Singapore: Springer; 2020. p.113-127.

[8] Nitchot A, Wettayaprasit W, Gilbert L. Personalized learning system for visualizing knowledge structures and recommending study materials links. *E-Learning and Digital Media*. 2019; 16(1): 77-91.

[9] Kurilovas E. Advanced machine learning approaches to personalise learning: learning analytics and decision making. *Behaviour and Information Technology*. 2019; 38(4): 410-421.

[10] Srisa-An C, Yongsiriwit K. Applying machine learning and AI on self automated personalized online learning. *Fuzzy System and Data Mining V: Proceedings FSDM 2019*. 2019; 320: 137.

[11] Vieira Sobrinho JL, Teles Vieira FH, Assis Cardoso A. Two-stage dimensionality reduction for social media engagement classification. *Applied Sciences*. 2024; 14(3): 1269.

[12] Alshattnawi S, Shatnawi A, AlSobeh AM, Magableh AA. Beyond word-based model embeddings: Contextualized representations for enhanced social media spam detection. *Applied Sciences*. 2024; 14(6): 2254.

[13] Malodia S, Filieri R, Otterbring T, Dhir A. *Unlocking Social Media Success: How Prosumers Drive Brand Engagement through Authentic Conversations with Consumers*. British Journal of Management; 2024.

[14] Jha AK, Verma NK. Social media platforms and user engagement: A multi-platform study on one-way firm sustainability communication. *Information Systems Frontiers*. 2024; 26(1): 177-194.

[15] Reimer T. Environmental factors to maximize social media engagement: A comprehensive framework. *Journal of Retailing and Consumer Services*. 2023; 75: 103458.

[16] Ren F, Tan Y, Wan F. Know your firm: Managing social media engagement to improve firm sales performance. MIS Quarterly. 2023; 47(1): 227.

[17] Agnihotri R, Bakeshloo KA, Mani S. Social media analytics for business-to-business marketing. *Industrial Marketing Management*. 2023; 115: 110-126.

[18] Lim WM, Rasul T. Customer engagement and social media: Revisiting the past to inform the future. *Journal of Business Research*. 2022; 148: 325-342.

[19] Eslami SP, Ghasemaghaei M, Hassanein K. Understanding consumer engagement in social media: The role of product lifecycle. *Decision Support Systems*. 2022; 162: 113707.

[20] Santos ZR, Cheung CM, Coelho PS, Rita P. Consumer engagement in social media brand communities: A literature review. *International Journal of Information Management*. 2022; 63: 102457.

[21] Shahbaznezhad H, Dolan R, Rashidirad M. The role of social media content format and platform in users' engagement behavior. *Journal of Interactive Marketing*. 2021; 53(1): 47-65.

[22] Gavrielov-Yusim N, Kürzinger ML, Nishikawa C, Pan C, Pouget J, Epstein LB, et al. Comparison of text processing methods in social media-based signal detection. *Pharmacoepidemiology and Drug Safety*. 2019; 28(10): 1309-1317.

[23] Arora A, Bansal S, Kandpal C, Aswani R, Dwivedi Y. Measuring social media influencer index-insights from facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*. 2019; 49: 86-101.

[24] Nadar N, Kamatchi R. Information and communication-based collaborative learning and behavior modeling using machine learning algorithm. In: *Social Media and Machine Learning*. London, UK: IntechOpen; 2019.

[25] Kristensen JB, Albrechtsen T, Dahl-Nielsen E, Jensen M, Skovrind M, Bornakke T. Parsimonious data: How a single facebook like predicts voting behavior in multiparty systems. *PloS ONE*. 2017; 12(9): e0184562.

[26] Siwach M, Mann S. A compendium of various applications of machine learning. *International Journal of Research in Engineering and Technology*. 2022; 9: 1141-1144.

[27] Siwach M, Mann S, Jain S, Rauthan J. Extractive text summarisation techniques-a survey. *International Research Journal of Engineering and Technology*. 2022; 9: 589-593.

[28] Marengo D, Settanni M. Mining facebook data for personality prediction: An overview. In: *Digital Phenotyping and Mobile Sensing*. Cham, Switzerland: Springer; 2019. p.109-124.

[29] Bogaert M, Ballings M, Hosten M, Van den Poel D. Identifying soccer players on Facebook through predictive analytics. *Decision Analysis*. 2017; 14(4): 274-297.

[30] Banouar O, Raghay S. Machine learning for personalized access to multiple data sources through ontologies. In: *Proceedings of the 2nd International Conference on Big Data, Cloud and Applications*. Tetouan, Morocco; 2017. p.1-6.

[31] Kristensen JB, Albrechtsen T, Dahl-Nielsen E, Jensen M, Skovrind M, Bornakke T. Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PloS ONE*. 2017; 12(9): e0184562.

[32] Sam E, Yarushev S, Basterrech S, Averkin A. Prediction of facebook post metrics using machine learning. *arXiv:1805.05579*. 2018. Available from: https://arxiv.org/abs/1805.05579.

[33] Savci M, Tekin A, Elhai JD. Prediction of problematic social media use (PSU) using machine learning approaches. *Current Psychology*. 2020; 41(5): 2755-2764.