**Trends in Social Sciences and Humanities Research**
**Print ISSN: 2959-992X**
**Online ISSN: 2959-9938**
**DOI: https://doi.org/10.61784/tsshr3208**

# PREDICTING USER ACTIVITY AND INTERACTION SEQUENCES ON SOCIAL PLATFORMS BY INTEGRATING ENSEMBLE LEARNING AND TIME SERIES DECOMPOSITION

ZhiYang Chen[1*], Wei Ding[2], GengYan Li[2]
[1]*School of Mathematics and Physics, Southwest University of Science and Technology, Mianyang 621000, Sichuan, China.*
[2]*School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang 621000, Sichuan, China.*
[*]*Corresponding Author: ZhiYang Chen*

**Abstract:** Addressing the highly discrete and complex patterns of user behavior on social platforms, this study constructs a predictive framework combining ensemble learning with dynamic time series decomposition to achieve precise modeling of user online status and interaction behaviors. Within the user online status recognition and multi-level influencer recommendation models, the study first extracts key features such as the user's past 3-day, 7-day, and historical total active days. Utilizing a random forest classifier to capture the nonlinear correlations in user behavior, it effectively determines the online status for specific dates. Subsequently, by defining an interaction score formula incorporating time-decay weights and integrating user preferences with global popularity, the model efficiently recommends interactive bloggers, achieving a recall rate of 80.48% on the validation set. In precise temporal interaction prediction based on user segmentation and time series decomposition, this study addresses the temporal periodicity of historical behavior by introducing the Facebook Prophet model. Through automatic decomposition of trend and seasonal components, it captures user activity patterns during holidays and specific periods. To enhance prediction accuracy, the study further clusters users using the K-means algorithm and configures model hyperparameters based on behavioral feature differences. Experimental results demonstrate that this approach effectively addresses the challenge of predicting interaction intensity across time intervals under sparse data conditions, providing scientific decision support for personalized platform operations.
**Keywords:** Random forest; Facebook prophet; User behavior modeling

## 1 INTRODUCTION

Against the backdrop of exponential growth in social media data, accurately capturing user activity dynamics and predicting potential interaction behaviors has become a core challenge for enhancing platform user retention and traffic conversion rates. User behavior data typically exhibits distinct characteristics of discontinuity, time preference, and susceptibility to specific holiday effects, rendering traditional static forecasting methods inadequate for complex dynamic scenarios. While previous studies have applied logistic regression or basic time series models to such contexts, they often struggle to balance the tension between uncovering nonlinear relationships and capturing long-term periodicity. The innovation of this study lies in two aspects: first, constructing a multidimensional activity decay feature system that enhances the robustness of online user identification through ensemble learning algorithms; second, adopting an innovative “cluster-then-predict” strategy. This involves segmenting user groups using the K-means algorithm and then specifically tuning the seasonal parameters of the Prophet model to achieve high-precision restoration of time-segmented interaction behaviors [1,2]. This section's research approach follows a logical progression from “all-day state determination” to “time-segmented behavior refinement.” It first employs random forests to establish preliminary profiles for online probability and influencer recommendations, then utilizes time series decomposition techniques to deeply model interaction intensity within a 24-hour window, constructing a comprehensive behavioral prediction framework [3].

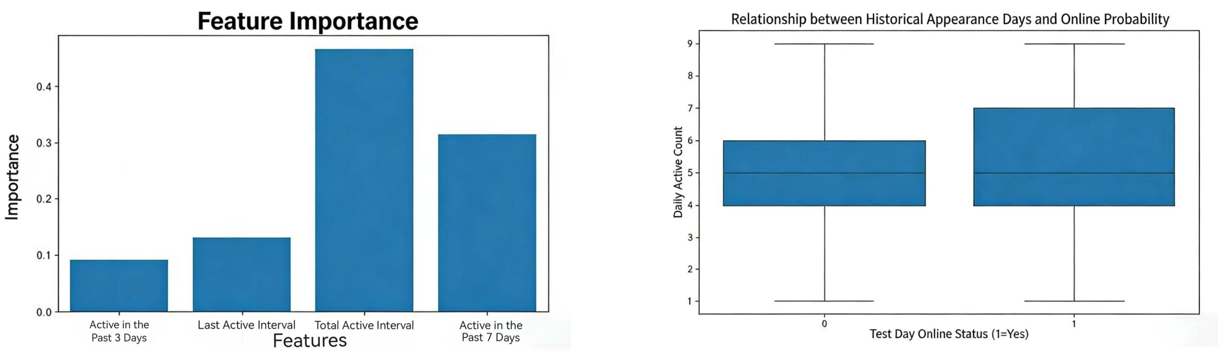
For the user online status recognition and multi-level influencer recommendation model, this section first transforms the online status determination into a binary classification task. The core approach involves feature engineering to extract users' recent (3 days), medium-to-long-term (7 days), and long-term (entire history) activity patterns, combined with the interval since last activity to characterize activity decay characteristics. By processing these complexly correlated features with a random forest classifier, the model automatically identifies feature importance, enabling precise determination of a user's online status on the target date. For predicting interactions among online users, this study designed a multi-level score evaluation model [4,5]. Interaction scores are defined by a formula incorporating a time decay coefficient wi to reflect the influence of recent interactions on current interest, with differentiated weights assigned to distinct behaviors like likes, comments, and follows. Validation results show the model achieves an F1 score of 74.25% and recall exceeding 80%, demonstrating the superior performance of the combined strategy of ensemble learning and decay weighting in user interaction recall tasks.

For the precise temporal interaction prediction component based on user segmentation and time series decomposition, a time series forecasting approach distinct from classification algorithms is adopted to address the cyclical fluctuations in user interactions within a 24-hour period. First, raw event data with irregular intervals underwent resampling and binning at 2-hour granularity to ensure continuous timestamps. Subsequently, the K-means clustering algorithm segmented users into four distinct clusters based on total behavior count, active time period characteristics, and behavior type distribution. Each cluster was configured with differentiated Prophet hyperparameters, such as the change point prior scale. Within this framework, the Facebook Prophet model employs Fourier series to model seasonality while incorporating day-of-week characteristics and weekend effects as additional regressors [6,7]. This enables the inference of interaction probabilities across specific dates and time slots. Ultimately, the model successfully pinpoints each user's peak activity periods throughout the day and their corresponding interaction intensity with bloggers, revealing the intrinsic temporal logic of user behavior through a data-driven approach.

## 2 SOLVING MULTI-LEVEL RECOMMENDATION MODELS FOR USER ONLINE STATUS RECOGNITION AND INTERACTION WITH BLOGGERS

### 2.1 Classification Model

To predict whether users are online, we convert this into a binary classification problem and use a Random Forest classifier to handle nonlinear relationships and capture feature importance [8,9].
We reflect the recent activity level through the number of active days of users in the past 3 days (i.e., the number of active days from July 18 to 20 for users), the medium and long-term activity pattern through the number of active days in the past seven days, and the long-term activity through the number of active days of users in the entire historical data. The number of days between the user's last active date and the target date— the larger the interval, the lower the online probability. Feature importance (left) and relationship between historical active days and online probability (right) are shown in Figure 1.




(a) (b)

**Figure 1** (a) Feature Importance; (b) Relationship between Historical Active Days and Online Probability
Note: Define the label as 1 if the user was online on July 20, 2024, and 0 otherwise.

### 2.2 Interactive Blogger Prediction Model

To predict the top 3 bloggers with the highest interaction count for online users, we need to combine personalized preferences and globally popular bloggers. We define the interaction score calculation as follows:

$$f_{u,b}=\sum_{i\in I_j} w_i \times z_i \qquad (1)$$

$$w_i=0.95^d \qquad (2)$$

where $f_{u,b}$ is the final score, $I_j$ is the set of interaction events, $w_i$ is the attenuation coefficient, d is the number of days difference, and $z_i$ is the weight for different interaction events. Weight table for the interactive blogger prediction model is shown in Table 1.

**Table 1** Weight Table for the Interactive Blogger Prediction Model

| Interaction Type | Weight |
|---|---|
| Watch | 1 |
| Like | 2 |
| Comment | 3 |
| Follow | 4 |

### 2.3 Model Verification

To comprehensively evaluate the model performance from different perspectives, we introduce recall rate to measure the proportion of positive samples correctly predicted as positive by the model. The F1 score is used to comprehensively consider precision and recall to balance the relationship between the two[10].

Through calculations using formulas (3) and (4), the verification recall rate is 80.48%, the verification F1 score is 74.25%, the precision rate can reach 78.34%, and the upper and lower error is 8%. Confusion matrix and ROC curve of the verification set is shown in Figure 2.

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

$$\text{F1}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \tag{4}$$

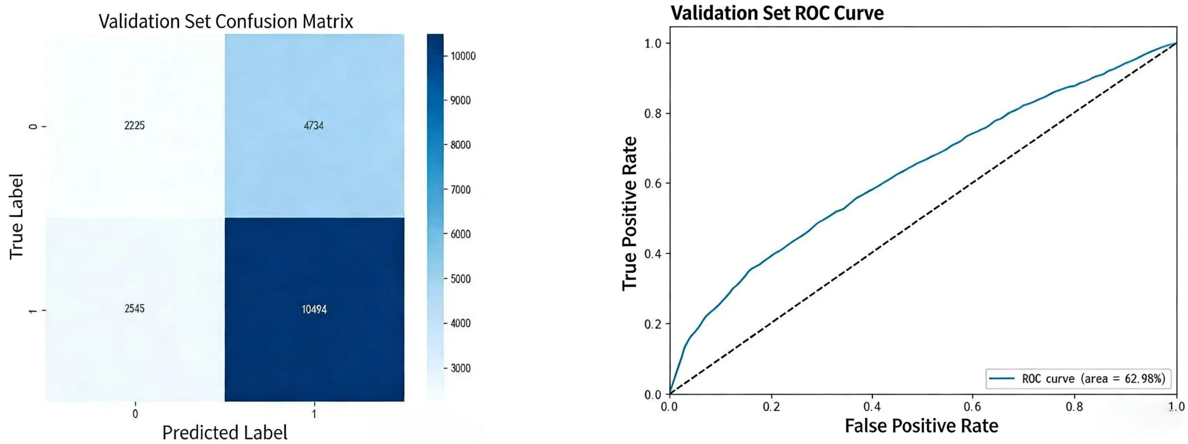



**Figure 2** Confusion Matrix and ROC Curve of the Verification Set

### 2.4 Model Solution

Through calculations using formulas (1) and (2), we obtain the following Table 2.

**Table 2** Results

| User ID | U9 | U22405 | U16 | U48420 |
|---|---|---|---|---|
| Blogger ID 1 | B3 | B47 | B23 | B5 |
| Blogger ID 2 | B13 | B59 | B24 | B21 |
| Blogger ID 3 | B24 | B42 | B4 | B2 |

## 3 SOLVING PRECISE SEQUENTIAL INTERACTION PREDICTION BASED ON USER CLUSTERING AND TIME SERIES DECOMPOSITION

### 3.1 Model Preparation

To improve the accuracy and non-accidentality of the solution results, this section adopts a different solution model from the above problems. Considering the discrete weak correlation of user behavior and the fact that user active time usually has periodicity, the data has obvious trends, seasonality, and holiday effects. We will add weekday features to the data and use the Facebook Prophet time series prediction model that can automatically capture such seasonal features. Prophet assumes that data can be decomposed into trends, seasonality, and noise, so it is necessary to ensure the continuity and uniformity of timestamps. Resampling can reduce the impact of irregular intervals. Therefore, we will perform bin aggregation on the data (resampling with a 2-hour granularity):

$$y_t=\sum_{i\in \text{bin}(t)} x_i,\ \text{bin}(t)=[t,t+2\text{hours}) \tag{5}$$

where:

$x_i$ represents the original event count;

$y_t$ represents the aggregated value;

Total number of interactions divided by hour in historical data is shown in Figure 3.

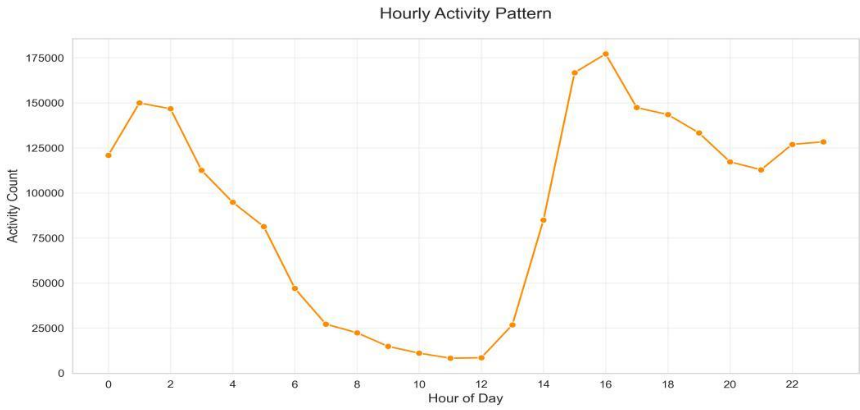



**Figure 3** Total Number of Interactions Divided by Hour in Historical Data

To meet the problem requirements of predicting specified users, we will sort the corresponding behavioral features of specified users in combination with time factors, including the daily interaction count and the interaction count divided by hourly segments over ten days, for model prediction and analysis. The data of User 10 is shown in Figure 4 as an example:


**Figure 4** Collation of Interaction Data for User 10 (User U10 Daily Activity, User U10 Hourly Activity, User U10 Behavior Distribution)

To enhance the model prediction results, we will cluster user behaviors to supplement the prediction time segments, so that users in different clusters use different Prophet hyperparameters to make the model results more accurate and reliable. First, extract the behavioral features of each user (such as active time segments, behavior type distribution, etc.) and normalize them, then select the optimal number of clusters according to the elbow method, and finally generate user groups. The Within-Cluster Sum of Squares (WCSS) is:

$$\mathrm{WCSS}=\sum_{i=1}^{k}\sum_{x\in C_i}\|x-\mu_i\|^2 \tag{6}$$

where:
x is the standardized user feature vector;
k=4, the number of clusters;
$C_i$ is all users in the i-th cluster;
$\mu_i$ is the centroid of the i-th cluster;

### 3.2 Model Establishment

Based on the above data processing, the Facebook Prophet model based on an additive model is used for time series prediction:

$$y(t)=g(t)+s(t)+h(t)+\varepsilon_t \tag{7}$$

where:
$g(t)=\frac{C}{1+e^{-k(t-m)}}$, trend term (logistic growth trend);
$s(t)=\sum_{n=1}^{N}(a_n\cos(\frac{2\pi nt}{T})+b_n\sin(\frac{2\pi nt}{T}))$, seasonality term (modeling periodicity using Fourier series);
h(t): holiday effect;
$\varepsilon_t$: noise term (follows a normal distribution);

### 3.3 Model Solution

Based on the above prediction formula (7), a mathematical model is established to solve the guessed active periods of specified users in Figure 5.

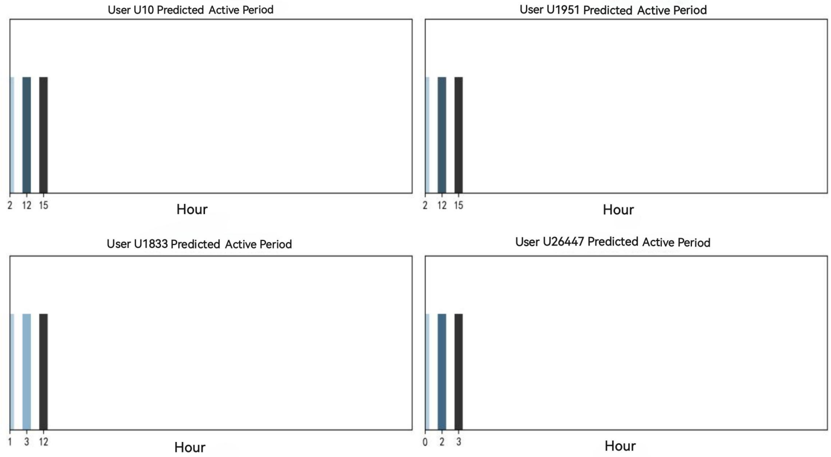



**Figure 5** Predicted active periods of specified users (User U10 Predicted Active Periods, User U1951 Predicted Active Periods, User U1833 Predicted Active Periods, User U26447 Predicted Active Periods)

Through data prediction, we finally obtain the following solution results in Table 3.

**Table 3** Results

| User ID | U10 | U1951 | U1833 | U26447 |
|---|---|---|---|---|
| Blogger ID 1 | B4 | B68 | B1 | B45 |
| Time Segment 1 | 2:00-3:00 | 2:00-3:00 | 1:00-2:00 | 0:00-1:00 |
| Blogger ID 2 | B3 | B4 | B23 | B5 |
| Time Segment 2 | 12:00-13:00 | 12:00-13:00 | 3:00-4:00 | 2:00-3:00 |
| Blogger ID 3 | B68 | B5 | B76 | B44 |
| Time Segment 3 | 15:00-16:00 | 15:00-16:00 | 12:00-13:00 | 3:00-4:00 |

## 4 CONCLUSIONS

This study systematically addresses the chain-based prediction challenge on social platforms—from “whether users are online” to “when and with whom they interact”—by integrating Random Forest ensemble learning, K-means clustering, and Facebook Prophet time series decomposition techniques. The study demonstrates that constructing multidimensional activity features significantly improves classification accuracy, while parameter optimization based on user segmentation substantially enhances the generalization capability of time series predictions. Limitations include the current time-slice prediction being constrained by a fixed 24-hour interval, making minute-level granularity challenging, and room for improvement in computational efficiency when processing extremely large datasets using the random forest model. Future research will explore incorporating deep learning sequence models like LSTM or Transformer to more flexibly capture long-range dependencies in user interests. Additionally, density-based clustering algorithms such as DBSCAN will be investigated to dynamically identify user activity windows, further enhancing prediction real-time capabilities and flexibility.

## COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

## REFERENCES

[1] Hartl T, Hutter C, Weber E. Matching for three: The search activities of workers, firms, and employment services. Economic Modelling, 2026, 155: 107434. DOI: 10.1016/J.ECONMOD.2025.107434.

[2] Vineela K, Grover J. A Comprehensive Evaluation of Health Hazard Prevention Education Programs Targeting Online Activity Risks Among Students. Journal of Educational Research and Policies, 2025, 7(11): 3-9. DOI: 10.53469/JERP.2025.07(11).02.

[3] Schutt K R, O'Donoghue L A, Hargraves L J, et al. Diminished mental well-being in the COVID-19 pandemic: The role of infection risk, social disconnection, and income loss at the individual and neighborhood levels. Wellbeing, Space and Society, 2025, 9: 100302. DOI:10.1016/J.WSS.2025.100302.

[4] Ron D A J, Whitehead J, Kaufmann J, et al. Gender and Sexual Orientation Differences in the Relationship Between Social Media Use and Disordered Eating: Results From a Serial Cross-Sectional Youth Survey From 2022 to 2024. The Journal of adolescent health: official publication of the Society for Adolescent Medicine, 2025, 78(1): 110-118. DOI: 10.1016/J.JADOHEALTH.2025.08.022.

[5] Lu Y, Quan K. Effects of doctors' participation in internet live streaming free medical consultations on their benefits: empirical evidence from a quasi-natural experiment. Journal of Business Research, 2025, 199: 115560. DOI: 10.1016/J.JBUSRES.2025.115560.

[6] Soltani M, Dara R, Poljak Z, et al. Leveraging social media and google trends to identify waves of avian influenza outbreaks in USA and Canada. Expert Systems With Applications, 2025, 291: 128482.

[7] Ghermandi A, Depietri Y, Orenstein E D. Human-nature interactions through a digital prism: Heterogeneity in user socio-demographics and content across social media platforms. Landscape and Urban Planning, 2026, 268: 105568.

[8] Aydın T, Parris A B, Arabacı G, et al. On the relationship between internet addiction and ADHD symptoms in adults: does the type of online activity matter?. BMC public health, 2025, 25(1): 2072. DOI: 10.1186/S12889-025-23040-4.

[9] Slote K, Daley K, Succar R, et al. How advocacy groups on Twitter and media coverage can drive US firearm acquisition: A causal study. PNAS nexus, 2025, 4(6): pgaf195. DOI: 10.1093/PNASNEXUS/PGAF195.

[10] Barrow G. GETTING TRUMPED. Commercial Motor, 2025, 236(6152): 38-39.