Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230166

# Research on the impact of recommendation algorithms on user stickiness based on data analysis of TikTok

**Yifei Lan**

International School, Beijing University of Posts and Telecommunications, Queen Mary University of London, Beijing, China, 100876

effielan@yeah.net

**Abstract.** The rapid expansion of the Internet era provides these social media with an excellent potential. Due to its accurate location and intelligent algorithm, TikTok's daily active users distinguish it from other programs. TikTok ranked in the top ten most downloaded applications worldwide in 2022, becoming the most popular social software for short, creative music videos worldwide. TikTok's recommendation algorithm serves as a benchmark for other Internet applications in terms of customer retention. This paper, through methods of questionnaire and literature review, first analyzes the model of TikTok's recommendation algorithm and then explores the hypothesis that TikTok's recommendation algorithm has a positive effect on user stickiness in three dimensions: algorithm accuracy, perceived usefulness, and recommendation content satisfaction, based on the data from questionnaire.

**Keywords:** TikTok, recommendation algorithm, data analysis, user stickiness.

## 1. Introduction

ByteDance incubated and released TikTok in 2016 as a short-form video platform. According to the 2022 TikTok Statistics Report, the number of daily active TikTok users has surpassed 600 million, the average daily video search volume has surpassed 400 million, and the average daily usage time in China is approximately 140 minutes [1]. TikTok users spend an average of 44.16 billion minutes each day, or 73.6 minutes per person based on its 600 million daily users. In 2020, Chinese adults used their phones for an average of 100.75 minutes each day. According to this statistic, users that download TikTok spend over 70% of their time there [2]. It is evident that TikTok not only has a large user base, but also a great user retention rate. It is impossible to separate TikTok's success from its superior recommendation system. Unfortunately, there is a scarcity of research-worthy literature on the impact of recommendation algorithms on user retention. In the paper, the author will describe TikTok's recommendation algorithm mode and then analyze the questionnaire responses to see whether the algorithm has a beneficial impact on user stickiness.

DOI: 10.54254/2755-2721/8/20230166

## 2. TikTok recommendation algorithm model

### *2.1. Using big data for persona from multiple perspectives*

Collection, analysis, and processing of demographic characteristics, behavioral preference attributes, and other information about content consumers to form a series of refined feature identifiers; these identifiers are then combined into tag groups to form a user model that can be understood by people [1]. TikTok uses data about the user and user reactions to short videos to create a composite profile.

TikTok prioritizes the user's basic characteristics, including age, gender, occupation, and other demographic information. TikTok then focuses on user-specific software-based usage characteristics, such as history browsing patterns and peak usage hours. In addition, TikTok records user feedback, such as likes, comments, and favorites, about video recommendations. When using the software, the user interacts with the content. The user activity is then captured and submitted to the software's logging system, and the data is further evaluated to build a list of recommended data and a user persona. When users sign up for the app, TikTok begins collecting their information. As consumers spend more time on TikTok, they provide more information. The more precise their persona will be, the more information they supply. Consequently, the user's personal advice is closer to their actual needs.

### *2.2. Basic collaborative filtering based on user information*

On the basis of generating persona, basic collaborative filtering is the most basic algorithm in TikTok's recommendation algorithm, which is widely used in the process of video promotion.

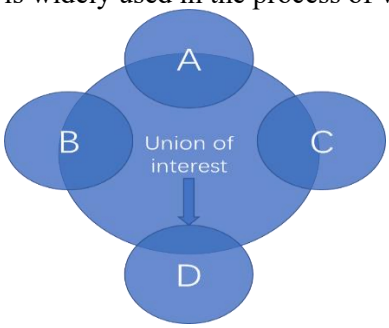


**Figure 1.** Basic collaborative filtering model [2].

If the persona of A, B, C and D is relatively similar, the contents which A, B and C are all interested in in the early stage will be recommended to D preferentially, as shown in figure 1.

### *2.3. Accurate distribution based on decentralization*

Decentralization is the most important transmission principle of social media. Each user is a node and can produce content independently. In this decentralized accurate distribution, content and social relationships become the main basis for accurate distribution of information, so as to realize peer-to-peer transmission.

The recommendation algorithms under decentralization can be divided into two categories. One is the recommendation based on the dimension of users' interest, and the short videos are recommended and matched according to the keywords in the persona. The other is recommendation based on the dimension of social circle. For example, TikTok will distribute short videos published or interested by its friends to users.

## 3. Methodology

This study collects data from TikTok users using a questionnaire and then analyzes the data statistically. On the web platform Wenjuanxing.com, a total of 101 surveys were issued and gathered. This study examines TikTok's recommendation system from three perspectives: the algorithm's accuracy, the perceived usefulness of the recommendation, and the satisfaction with the recommendation content.

DOI: 10.54254/2755-2721/8/20230166

Among them, the accuracy of the recommendation algorithm is described by the items: "TikTok can distribute a large number of created videos to the target audience", "the content recommended by TikTok is in line with my preferences" and "the recommendation system of TikTok can mine my potential preferences".

The perceived usefulness of the recommendation was described as "I can get useful information through TikTok's recommendation", "videos recommended by TikTok can help me learn knowledge or skills" and "TikTok's recommendation can help me meet new people".

The satisfaction of recommended content is described by "the content of TikTok's short videos is rich and diverse", "the content of TikTok's transmission is very timely and rapid" and "the content of TikTok's short videos is beautifully produced".

TikTok user stickiness is described by "how often you use TikTok" and "how long you use TikTok every day".

SPSS 26.0 was then used to analyze the data. The reliability and validity test, correlation and regression analysis are carried out on the data.

## 4. Results and analysis

### *4.1. Respondents analysis*

**Table 1.** Statistical results of respondents' gender [3].

| Gender | | | | | |
|---|---|---|---|---|---|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | male | 41 | 40.6 | 40.6 | 40.6 |
| | female | 60 | 59.4 | 59.4 | 100.0 |
| | Total | 101 | 100.0 | 100.0 | |

As can be seen in table 1, there are a total of 101 people who responded to this questionnaire survey. Of these, 41 are male, making up 40.6% of the total, and 60 are female, making up 59.4% of the total. There are only tiny deviations between the gender ratios of the response group, with somewhat more females than boys.

**Table 2.** Statistical results of respondents' age [3].

| Age | | | | | |
|---|---|---|---|---|---|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | 16 ~ 20 | 13 | 12.9 | 12.9 | 12.9 |
| | 21 ~ 25 | 32 | 31.7 | 31.7 | 44.6 |
| | 26 ~ 30 | 22 | 21.8 | 21.8 | 66.3 |
| | 31 ~ 35 | 24 | 23.8 | 23.8 | 90.1 |
| | above35 | 10 | 9.9 | 9.9 | 100.0 |
| | Total | 101 | 100.0 | 100.0 | |

According to table 2, in terms of the age structure, there are 13 people aged 16-20, which accounts for 12.9% of the total population; 32 people aged 21-25, which accounts for 31.7% of the total population; 22 people aged 26-30, which accounts for 21.8%; 24 people aged 31-35, which accounts for 23.8%; and 10 people aged over 35, which accounts for 9.9%.

DOI: 10.54254/2755-2721/8/20230166

**Table 3.** Statistical results of respondents' educational background [3].

| The highest degree studied or obtained | | | | | |
|---|---|---|---|---|---|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | high school degree or below | 1 | 1.0 | 1.0 | 1.0 |
| | junior college degree | 39 | 38.6 | 38.6 | 39.6 |
| | bachelor degree | 54 | 53.5 | 53.5 | 93.1 |
| | master degree or above | 7 | 6.9 | 6.9 | 100.0 |
| | Total | 101 | 100.0 | 100.0 | |

As can be seen in table 3, there are a total of 7 master's degrees or above, which account for 6.9% of the total, 54 bachelor's degrees, which account for 53.5% of the total, and 39 junior college degrees, which account for 38.6%. The educational structure primarily involves a bachelor's degree or higher.

*4.2. Reliability and validity test*

**Table 4.** Results of KMO and Bartlett's Test [3].

| KMO and Bartlett's Test | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | | .953 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 766.802 |
| | df | 55 |
| | Sig. | .000 |

The validity analysis was performed using SPSS 26.0. A standard validity test criterion is KMO. It is highly appropriate for factor analysis because KMO is more than 0.90. 0.8–0.9 is adequate; 0.7–0.8 is suitable; 0.6–0.7 is not suitable; 0.5–0.6 is just barely enough; Less than 0.5 is not suitable. In the table above, you can see the results of the KMO validity test applied to this questionnaire. According to the findings, this survey has a KMO validity score of 0.953 ($p<0.05$). This demonstrates the excellent validity and reliability of the questionnaire in representing the true state of affairs.

**Table 5.** Results of Reliability Statistics [3].

| Reliability Statistics | |
|---|---|
| Cronbach's Alpha | N of Items |
| .944 | 11 |

Indicated in the table above is the reliability analysis of the questionnaire, which measures how consistent the findings of the measurements are. The less the scale's measuring error, the higher its accuracy. In this research, the author employs the widely-used Cronbach's model in SPSS 26.0 to determine the reliability of the scale, and use the size of the Cronbach's coefficient to infer the reliability of the questionnaire. In most cases, a Cronbach's coefficient above 0.6 is considered to be respectable. With a Cronbach's coefficient of 0.944, the questionnaire has an excellent reliability index and the stability of the questionnaire test is well ensured.

DOI: 10.54254/2755-2721/8/20230166

### *4.3. Correlation analysis*

**Table 6.** The result of correlation analysis [3].

*Correlations*

| | | The accuracy of algorithm | Perceived usefulness of recommendation | The satisfaction of recommended content | User stickiness |
|---|---|---|---|---|---|
| The accuracy of algorithm | Pearson Correlation | 1 | .822** | .829** | .791** |
| | Sig. (2-tailed) | | .000 | .000 | .000 |
| | N | 101 | 101 | 101 | 101 |
| Perceived usefulness of recommendation | Pearson Correlation | .822** | 1 | .827** | .734** |
| | Sig. (2-tailed) | .000 | | .000 | .000 |
| | N | 101 | 101 | 101 | 101 |
| The satisfaction of recommended content | Pearson Correlation | .829** | .827** | 1 | .803** |
| | Sig. (2-tailed) | .000 | .000 | | .000 |
| | N | 101 | 101 | 101 | 101 |
| | | The accuracy of algorithm | Perceived usefulness of recommendation | The satisfaction of recommended content | User stickiness |
| User stickiness | Pearson Correlation | .791** | .734** | .803** | 1 |
| | Sig. (2-tailed) | .000 | .000 | .000 | |
| | N | 101 | 101 | 101 | 101 |

**. Correlation is significant at the 0.01 level (2-tailed).

To investigate the link and degree of association between the accuracy of a recommendation algorithm, the perceived utility of a recommendation, the satisfaction of a recommendation's content, and software usage persistence. The data correlation test reveals a correlation coefficient of 0.791% ($P<0.05$) between the accuracy of the recommendation algorithm and the software's user stickiness. This indicates that there is a positive association between the accuracy of a recommendation algorithm and the user retention rate of a software. The frequency and duration of TikTok usage increases proportionally with the accuracy of the recommendation system [4-5]. The correlation coefficient between perceived usefulness of recommendation and user stickiness was 0.734 ($P<0.05$), demonstrating a positive association between perceived usefulness of recommendation and user stickiness. The greater the perceived utility of the recommended material, the more frequently consumers utilize TikTok. The correlation between recommended content satisfaction and user stickiness was 0.803 ($P<0.05$), showing a favorable link. The greater the content satisfaction of users, the more likely they are to utilize TikTok [6].

DOI: 10.54254/2755-2721/8/20230166

### *4.4. Regression analysis*

This study uses regression analysis to investigate the effect of the recommendation algorithm on user stickiness, with the accuracy of the algorithm, the perceived usefulness of the recommendation, and the satisfaction with the recommended material serving as independent variables.

**Table 7.** model summary.

*Model Summary*

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .834[a] | .695 | .686 | .46465 |

a.Predictors: (Constant), the satisfaction of recommended content, perceived usefulness of recommendation, the accuracy of recommendation algorithm

**Table 8.** ANOVA Test.

*ANOVA[a]*

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 47.785 | 3 | 15.928 | 73.777 | .000[b] |
| | Residual | 20.942 | 97 | .216 | | |
| | Total | 68.728 | 100 | | | |

a. Dependent Variable: user stickiness

b. Predictors: (Constant), the satisfaction of recommended content, perceived usefulness of recommendation, the accuracy of recommendation algorithm

**Table 9.** Regression coefficient.

*Coefficients[a]*

| | | Unstandardized Coefficients | | Standardized Coefficients | | | Collinearity Statistics | |
|---|---|---|---|---|---|---|---|---|
| Model | | B | Std. Error | Beta | t | Sig. | Tolerance | VIF |
| 1 | (Constant) | .241 | .158 | | 1.520 | .132 | | |
| | the accuracy of recommendation algorithm | .280 | .083 | .373 | 3.352 | .001 | .254 | 3.942 |
| | perceived usefulness of recommendation | .046 | .080 | .063 | .573 | .568 | .257 | 3.886 |
| | the satisfaction of recommended content | .335 | .086 | .441 | 3.912 | .000 | .247 | 4.043 |

a. Dependent Variable: user stickiness

Table 7's model summary reveals that the adjusted R square of the established regression model is 0.686, indicating that the explanatory power of the independent variable to the dependent variable is 68.6% and that the independent variable can adequately explain the variation of the dependent variable. The F value of the ANOVA test for the regression model was 73.777 ($P<0.05$), showing that at least one independent variable adequately explains the dependent variable, and the model findings are acceptable, as shown in Table 8.

The VIF values are all less than 10 as evidenced by the model coefficient table, and there is no collinearity issue with the regression model presented in table 9. The regression influence coefficient of the accuracy of the recommendation algorithm on user stickiness is 0.373, and the corresponding

DOI: 10.54254/2755-2721/8/20230166

significance P value is less than 0.05, indicating that the accuracy of the recommendation algorithm has a statistically significant positive influence on user stickiness. The regression influence coefficient of perceived usefulness of suggestion on user stickiness is 0.063, and its associated significant P value is larger than 0.05. The regression influence coefficient of suggested content satisfaction on user engagement is 0.441, and the associated significant P value is less than 0.05, showing that recommended content satisfaction has a substantial positive effect on user engagement.

## 5. Conclusion

Up until now, TikTok's algorithm for recommending material has been a mature system capable of providing customised content based on user characteristics. Creating personas from many perspectives using big data, collaborative filtering based on user knowledge, and decentralized correct distribution are the three most fundamental recommendation algorithm models. According to the findings of this paper, the recommendation system has a significant impact on user retention, which is a significant factor in TikTok's ability to distinguish itself from other similar products. The success of TikTok is replicable. TikTok's great algorithm mode can be utilized as a reference by other Internet platforms to increase their user retention [7].

Even though the ratio of males to women in this study has attained a certain equilibrium, the gathered questionnaires do not adequately represent groups with varying educational levels. In addition, this research is founded on the individual experiences of consumers. There will be some differences between this investigation and the actual scenario, as it is not based on the actual data collected in the background of TikTok. In a later stage, this line of inquiry can undertake additional research on the situation of users with varying levels of education, collecting the same number of questionnaires from each educational group. In addition, the research can attempt to mine the background data of TikTok and re-analyze it using more objective information.

## References

[1] Xu Li-ping, HE Dan, LU Yuan-wen. Research on intelligent recommendation based on User portrait -- Taking TikTok APP as an example [J]. Media,2022(12):53-56. (in Chinese) DOI:10.3969/j.issn.1009-9263.2022.12.020.

[2] Zhao Chen-wei, LIU Tao, DU Hai-hong. Research on video recommendation mode of TikTok short video platform from the perspective of algorithm [J]. Publishing Wide Angle,2019(18):76-78. (in Chinese) DOI:10.16491/j.cnki.cn45-1216/g2.2019.18.023.

[3] Research on the Influence of TikTok’s Recommendation Algorithm on User Stickiness 2023 https://www.wjx.cn/

[4] [Zhang C Y. How Products make users addicted -- Analysis of Marketing Strategies based on TikTok.] National Circulation Economics,2019(7):42-44. (in Chinese) DOI:10.3969/j.issn.1009-5292.2019.07.019.

[5] On the realization of personalized recommendation of e-commerce platform [J]. National Circulation Economy,2021(5):9-11. (in Chinese) DOI:10.3969/j.issn.1009-5292.2021.05.003.

[6] [Hu J J. Empirical research on the Continuous use intention and influencing factors of short music video APP -- taking Douyin as an example [C]. // Proceedings of 2018 New Media International Forum. 2018.75-84.]

[7] Dai Debao, Gu Xiaohui. User engagement behavior, perceived value and loyalty: an analysis based on mobile short video social applications [J]. Journal of Consumer Economics,2017,33(2):58-65.