Influencer sales on TikTok: Forecasting prominent factors

Jiahao Liang[1], Simon Fong[1[0000-0002-1848-7246]], Sandra Méndez-Muros[2[0000-0003-4312-9798]] and Antonio J. Tallón-Ballesteros[3[0000-0002-9699-1894]]

[1] University of Macau, Macau, China
[2] University of Seville, Seville, Spain
[3] University of Huelva, Huelva, Spain
antonio.tallon@diesia.uhu.es

Abstract. The rapid growth of e-commerce, particularly on platforms that merge socializing and shopping functions like TikTok, has led to the rise of the influencer economy, where influencers drive significant sales through their online presence. This paper aims to forecast the prominent factors influencing influencer sales on TikTok, motivated by the need to better understand this evolving economic model. We establish an original dataset containing data from 100 influencers over a three-month period on TikTok's e-commerce platform. A comprehensive descriptive analysis is conducted to identify variations among influencers, followed by a data mining process to extract key characteristics based on their behavior across different levels, including daily activities, marketing strategies, and basic demographic information. A predictive model is developed to assess influencers' sales levels, revealing that the root mean square error (RMSE) of the model is close to 13. In particular, the study identifies the top 10 most influential sales features. These findings contribute to a deeper understanding of the factors driving influencer sales, offering valuable insights for both influencers and marketers in optimizing their strategies for success on TikTok.

Keywords: Influencer, Machine learning, TikTok, Data Mining, E-commerce

1 Introduction

The influencer role has emerged and consumer decisions are based on social media and currently prominent social network users are the reference model for other users [3]. In recent years, online shopping has surged in popularity, driven by the rapid advancements in information technology and the expansion of the Internet economy [10]. This trend has been further accelerated by the global spread of COVID-19, which led to the closure of physical stores and an increased demand for online shopping. Consequently, more companies, brands, and new retail platforms have entered the e-commerce market, contributing to its continuous growth. Between 2014 and

2022, the global retail e-commerce market nearly tripled in size. In 2021, global retail e-commerce sales reached approximately $5.2 trillion, and this figure is projected to grow by 56%, reaching around $8.1 trillion by 2026 (https://www.helpscout.com/blog/ecommerce-statistics/). According to the U.S. Department of Commerce, U.S. e-commerce sales increased from $1.040 trillion in 2022 to approximately $1.119 trillion in 2023, a 7.6% rise, while total retail sales grew from $4.904 trillion in 2022 to about $5.088 trillion in 2023, an increase of 3.8%. As Generation Z's purchasing power continues to grow, they are expected to become the dominant online consumers. The leading platforms in the global e-commerce market include Alibaba, Amazon, Jingdong, Pinduoduo, Apple, eBay, Samsung, Xiaomi, Coupang, and Walmart. In the U.S. and Canadian e-commerce markets, Amazon dominates in total visits. As of August 2023, Amazon's total visits in the U.S. and Canada reached 2.5 billion and 182 million, respectively. In second place is eBay, with 722 million total visits in the United States and Canada. Additionally, Walmart, Home Depot, and other e-commerce platforms are highly active in the North American market, with a comparable number of visits. The rise of China's "four little dragons" has intensified competition in the North American e-commerce market [9]. Temu, for example, quickly captured consumer interest with ultra-low prices after entering the overseas market; TikTok recently launched its TikTok Shop service for U.S. users; and SHEIN introduced its third-party platform model, SHEIN Marketplace, earlier this year.

As the market and user base expand, major social and e-commerce platforms are exploring new business models, continually refining their operations to grow their content ecosystems, and striving to build a diverse and high-quality ecosystem. Facebook, Instagram, and TikTok are the three most prominent social platforms globally, and for years, they have driven e-commerce sales by directing users to shopping sites via branded advertisements. Recently, however, these platforms have begun integrating stores and checkout features to allow consumers to complete transactions directly within the platform. In April 2023, Meta announced that stores on Facebook and Instagram would require in-platform checkout through Facebook. In August, TikTok announced plans to shut down its semi-closed-loop model in preparation for the rollout of TikTok Shop. TikTok officially launched TikTok Shop in the U.S. in September and celebrated its first Black Friday, offering discounts of up to 50%. By the end of the year, some of TikTok Shop's best-selling items had sold over 100,000 units per month. This new business model is largely fueled by the influencer economy. The influencer economy [7] refers to the phenomenon where influencers use the Internet and social platforms to independently select and promote products, guiding their followers toward consumption, thereby generating revenue and creating business opportunities. As live broadcasts, short videos, and other social platforms continue to evolve, the forms of Internet content consumption are becoming more diverse. Millions of original content creators on major online media platforms actively create content and participate in its dissemination. Many of these ordinary Internet users have attracted large followings through continuous content creation and sharing, transforming into influencers and giving rise to the influencer economy. For e-commerce platforms, combining shopping functions with live broadcasting, short videos, and self-media platforms can increase user engagement, improve order conversion and repurchase rates, and effectively maintain user loyalty. For brands and sellers, the conversion rate of orders generated by an influencer's private domain traffic is often higher than that of the platform's public domain traffic. According to Statista, the global influencer marketing market has grown significantly, from $6.5 billion in 2019 to $16.4 billion in 2022, more than doubling in size in under three years [2]. In 2023, influencer marketing was valued at over $21 billion, with more than 75% of brands investing in influencer marketing and over 11% spending more than $500,000 on it.

To address these developments, this paper examines the factors influencing influencer sales on TikTok. Our objectives are threefold: first, to gather primary data on influencers and their sales; second, to analyze the marketing behaviors of these influencers to uncover patterns and principles; and third, to build a machine learning model that combines influencers' behavioral features with basic characteristics to explore the relationship between these factors and sales levels. These studies offer insights into the new economic model of influencer marketing and assist influencers and their brokers in achieving precise market positioning. The model, which combines influencers and e-commerce, holds significant research value in the current context. Therefore, this paper begins by examining the rise and development of influencers, analyzing the current state of influencer marketing, and identifying the characteristics exhibited by influencer representatives over time. Finally, it provides rational suggestions aimed at offering theoretical guidance for influencer operations and assisting merchants in selecting the right partners. The remainder of this paper is structured as follows: Section 2 reviews key concepts related to the influencer economy and associated issues; Section 3 describes the data acquisition process; Section 4 introduces the proposed methodology; Section 5 presents the experimental results; and Section 6 concludes the paper.

2 Foundations

Influencers are individuals who gain fame either through a specific event or behavior that captures the attention of netizens in real or online life, or through the consistent and long-term sharing of their expertise. Their popularity arises from certain qualities that the Internet amplifies, resonating with the public's psychological preferences for aesthetics, entertainment, excitement, voyeurism, and other such traits. These qualities make them the focal point of online attention, leading them to become influencers, whether intentionally or unintentionally [8]. The modern concept of influencers did not originate with social media platforms but rather with mommy bloggers. The first wave of mommy bloggers began in 2002 when Melinda Roberts created The Mommy Blog.com, a site where she shared the highs and lows of motherhood [11]. Through her tips, product recommendations, and parenting experiences, she influenced how mothers around the world raised their children. The launch of Instagram in 2010 marked a significant shift in influencer culture. Instagram enabled users to connect, post pictures, and share their favorite products. Influencers quickly began using the platform to connect with followers on a deeper level, sharing their daily lives and the products they used. In 2013, Instagram introduced a paid advertising feature, making it easier for brands to collaborate with influencers, streamlining the process of selling products and enabling influencers to monetize their recommendations with a simple click. Twitch, founded in 2011, introduced a new dimension to social media through live gaming. It allowed gamers to connect, live stream their games, and interact with viewers in real time, giving rise to a new category of social media influencers known as gaming influencers. These gamers could now earn income by streaming their content and sponsoring various games. TikTok, launched in 2016, brought another unique approach to social media. Unlike other platforms, TikTok's content is tailored to individual users, offering a personalized and highly engaging experience through its "For You" page. The platform's constant stream of diverse content keeps users entertained, while influencers use it to grow their following and connect with audiences in innovative ways. TikTok also allows influencers to participate in trends, such as popular songs and dance challenges, and to create sponsored content. As influencer marketing continues to evolve, the forms of influencers have diversified. What started as individual content creation has expanded into various development models, including capital investment, platform-based operations, and the involvement of MCN (Multi-Channel Network) agencies.

E-commerce, a broad term for trading goods and services over the Internet, was defined by the International Chamber of Commerce (ICC) in 1997 as the process of digitally realizing entire transactions through technology. It transforms traditional brick-and-mortar stores into online establishments on the information superhighway. E-commerce can be categorized into three types based on the transaction platform: the first category includes franchised e-commerce platforms like Amazon, Taobao, and Jingdong; the second involves social media platforms, such as Facebook and Weibo, which promote e-commerce by attracting fans to shop on these platforms; and the third combines socializing and shopping functions, exemplified by platforms like Xiaohongshu and TikTok. Influencer marketing [6], which leverages individuals with significant social media followings, has become increasingly important for brands. This form of marketing involves partnering with influencers to promote a brand and drive consumers to purchase through the social media channels the influencers operate on. Unlike traditional marketing, where the seller communicates directly with the consumer, influencer marketing entrusts the influencer to deliver the brand's message, building trust with the audience by sharing their expertise and insights. It has the following four main features: a) Powerful Influence: Influencers usually have a large fan base and a solid social presence. Each of their tweets and videos can attract thousands of attention and interactions, which makes them ideal candidates to promote their products; b) Precise audience targeting: Different influencers have different fan bases and audience targeting. By choosing the right influencers to work with relevant to the merchant's product or brand, they can precisely deliver ads to their target audience and increase marketing efficiency; c) Highly interactive: Influencer marketing is usually presented in video, live streaming, social media, etc. This multimedia format is highly interactive. Consumers can interact with influencers by commenting, liking, sharing, etc., increasing user participation and brand attention and d) Image spokesperson effect: As influencers have unique personal images and styles, products cooperating with them can use their image spokesperson effect to give products a unique brand image and increase product recognition and reputation.

3 Data acquisition

TikTok is the world's largest short-format video platform, with 1.56 billion active users in 2024. The platform is predominantly youthful, with 60% of its monthly active users aged between 16-24, and those aged 10-29 accounting for over 60% of all active users in the U.S. Generation Z (Gen Z) represents TikTok's most significant potential user base. Today, TikTok has become the preferred search engine for Gen Z, even surpassing older search engines. In early 2021, TikTok launched its e-commerce platform, TikTok Shop, in Indonesia and the U.K., and within just one year, TikTok Shop generated approximately $5 billion in Gross Merchandise Value (GMV).

Unlike traditional shelf e-commerce platforms such as Amazon, Shein, and Temu, TikTok employs an interest-based e-commerce model with a one-stop shopping experience within the app. This dual-driven model combines content and shelf e-commerce. Traditional shelf e-commerce operates on the principle of users searching for specific products when they have a clear shopping need. In contrast, interest-based e-commerce involves actively presenting graphic or short video content to users through recommendation algorithms. TikTok, with its vast pool of highly engaged users, has successfully tested this interest-based e-commerce model through its domestic Douyin store. This emerging model is expected to become the mainstream approach in future e-commerce.

Given the significance of this model, this study focuses on a specific market segment, namely the TikTok platform. From an academic perspective, focusing on a specific segment can reduce the number of variables missed during analysis. By concentrating on a single platform, it is possible to effectively assess the commercial value of influencers within that segment, aiding merchants in quickly identifying the right influencers for their needs.

Due to TikTok's regulations, it is not possible to directly obtain influencers' marketing data from the platform. Therefore, we selected 110 influencers in the apparel category in the U.K. using the EchoTik platform to gather their marketing data and personal information over last three months of the previous year (01 October 2023 till 31 December 2023), resulting in a dataset of 2,750 records. In any case, EchoTik is a TikTok shop data analytics platform.

The data preprocessing in this study involved several key steps: a) removing records for influencers with missing marketing data, and b) standardizing the data to thousands of units. After cleaning the data, we obtained a final dataset of 2,610 records for 100 influencers.

3.1 Basic information about the influencers

In this study, we analyze the personal attributes of influencers to identify common characteristics, which will serve as the foundation for the subsequent quantitative analysis. On the TikTok platform, influencers can select up to five labels that help users quickly find influencers of similar types or receive recommendations tailored to their browsing preferences. Therefore, these labels can be considered representative of the influencers' attributes. Among the 100 influencers analyzed, 88% (88 influencers) chose the label "Professional Services." Other frequently selected labels include Government, Health, and Fashion & Style, all of which were chosen by more than half of the influencers. These attributes will be used as variables in the following study to explore the relationship between influencer attributes and sales performance.

Influencers capture the attention of their fans by posting content on TikTok, and they typically begin to monetize their influence once their follower count reaches a certain threshold. Thus, the role of followers is crucial for influencers. In this paper, we obtained follower data for 100 influencers, which were the top 100 over the year, and analyzed their follower counts. The average number of followers among these influencers is approximately 218,000, with the highest follower count being 2.9 million and the lowest being just 1,300. Statistically, 75% of the influencers have fewer than the average number of followers, with only a few top-tier influencers having significantly larger followings. Given the importance of followers, the frequency with which influencers publish content plays a vital role in attracting and retaining followers. Therefore, this paper examines the number of posts made by 100 influencers. The average number of posts per influencer is 1,081, though there is considerable variation. The influencer with the fewest posts has published 59, while the most prolific influencer has published 5,400. The relationship between the number of posts and sales performance will be analyzed in the following sections. The number of likes on an influencer's content is an indicator of popularity and content quality. However, since the number of posts varies between influencers, relying solely on the total number of likes to assess popularity can be misleading. To address this, we introduce two indirect metrics: the average number of likes per video and the likes-to-fans ratio. The average number of likes per video is calculated by dividing the total number of likes by the number of posts for each influencer. The likes-to-fans ratio is obtained by dividing the total number of likes by the number of followers. This ratio reflects the engagement level and content quality relative to the influencer's follower base.

In our analysis, we found significant variation in the quality of content and follower engagement among influencers. The likes-to-fans ratio, for instance, ranges from as low as 500 likes per fan to as high as 161,000 likes per fan, with an average of about 28,000 likes per influencer. Regarding the average number of likes per video, the influencers analyzed received an average of 10 million likes per post. The influencer with the lowest average likes per post garnered just 22,000 likes per post. These variations suggest that the quality of content differs greatly among influencers, and in this paper, we will use the number of likes and the likes-to-fans ratio as dependent variables to investigate their correlation with sales performance. In addition to likes, the number of views also serves as an indicator of an influencer's popularity. The number of views reflects the reach and appeal of the content, with higher view counts indicating broader attention and, potentially, higher popularity. In this paper, we analyze the total view counts for the content posted by 100 influencers. The analysis reveals a wide range of total views, with the lowest being 5,500 and the highest reaching 655 million, while the average is approximately 47 million views.

3.2 Influencer’s marketing message description

On the TikTok platform, influencers primarily earn money through selling goods. The typical process involves requesting free samples from sellers on the platform, awaiting the merchant's approval to ship the products, and then creating at least one short video featuring the samples to complete the transaction. In this study, we collected data on the video promotions of 100 TikTok influencers over the past three months, along with statistics on the retweets and likes of these videos. Given that fewer TikTok influencers opt for live streaming to promote products, video-based marketing is the sole marketing behavior examined in this paper. We analyzed the number of marketing videos posted by influencers within a three-month period. The minimum number of videos posted per month was two, while the maximum reached 564. On average, 60% of the videos posted by these influencers were marketing-related, with a few influencers dedicating all their videos to marketing. The number of items an influencer sells is often indicative of their popularity among merchants; only those who are highly reputable and effective at selling are consistently chosen to promote products for various stores. This paper examines the number of items sold by 100 influencers over three months. The analysis reveals that the average number of items sold is 138, with significant variation among influencers—ranging from as few as five items to as many as 979. However, a high number of items sold does not necessarily correlate with high sales volume. Therefore, the following section introduces the number of items sold by influencers over three months as a dependent variable to explore whether there is a correlation between this metric and sales volume. The average price of items promoted by influencers can reflect the purchasing power of their fan base to some extent, providing merchants with insight into whether their product prices align with the spending capacity of the influencers' followers. This paper examines the average price of items promoted by 100 influencers over three months. The analysis shows that the price range is relatively narrow, with the lowest price at £12.5, the highest at £49.2, and an average of £20. Approximately 60% of the influencers' average item prices hover around £20. Sales volume is frequently considered one of the most critical metrics for assessing an influencer's ability to sell products, as it often directly impacts the success of product promotions. This paper analyzes the sales volume of items sold by 100 influencers over three months. The findings indicate significant disparities in sales volume among influencers, with the lowest being 16 pieces and the highest reaching 100,000 pieces. The average sales volume for influencers stands at 7,246 pieces. In the subsequent section, we will explore whether sales volume is indeed the most important factor influencing overall sales performance.

4 Proposal

Selling products is a crucial aspect for influencers aiming to monetize their presence on TikTok, where most influencers opt to promote items through short videos. The number of marketing videos refers to the quantity of videos posted by influencers that include shopping cart links. The proportion of marketing videos is defined as the ratio of such videos containing shopping carts to the total number of videos posted by the influencer over the past three months. Additionally, variables like the average number of comments, likes, and retweets on marketing videos are also considered.

Before constructing the model, a correlation analysis was conducted to determine the relationship between each feature and sales. The correlation coefficient between an influencer's marketing features and their sales volume was calculated. Notably, the correlation coefficient between the number of sales and total sales volume is 0.731, indicating a strong relationship, which suggests that sales volume is a significant factor in influencing overall sales. The average number of likes and the average duration of marketing videos show moderate correlations with sales, with coefficients of 0.305 and 0.310, respectively, while other features exhibit weaker correlations.

Everyday videos, which do not include shopping carts, typically involve influencers sharing their daily lives or creating more narrative-driven content. These behaviors, classified as non-marketing behavioral characteristics, are important because they help influencers build trust and deepen the connection with their audience. This enhanced trust can increase the likelihood that followers will purchase the products recommended by the influencer. Therefore, variables representing these daily behaviors are included in the model construction.

The correlation analysis between influencers' daily characteristics and sales, as shown in Table 1, reveals that most of these characteristics are weakly correlated with sales. However, the average number of likes and views on daily videos are exceptions, showing moderate correlations with sales.

Table 1: Output of correlation coefficients

Characteristic classification	**Eigenvalue**	**Correlation coefficient**
Non-marketing Behavioral Characteristics	Number of daily videos	0.082
	Average number of comments on daily videos	0.256
	Average number of retweets on daily videos	0.077
	Average number of likes on daily videos	0.307
	Average number of daily video views	0.334

User fan volume is a key indicator of an influencer's reach and influence. Most influencers have a fan base ranging from 500,000 to 2 million. To analyze how different levels of fan volume impact sales, influencers are categorized into five levels: fans lev1, fans lev2, fans lev3, fans lev4, and fans lev5. The criteria for these divisions and their correlation coefficients with sales are detailed in the following table. The correlation coefficients for fans lev1 and fans lev3 with sales are 0.354 and 0.383, respectively, indicating a moderate correlation. In contrast, the correlation coefficients for the other three levels show only a weak correlation with sales.

The number of videos published by an influencer is another important metric, reflecting their level of activity. This study categorizes influencers based on the number of videos they have posted, with the classification criteria and correlation test results compiled in the table below. However, it was found that all these characteristic quantities are only weakly correlated with sales.

Video views are indicative of an influencer's traffic and the popularity of their content. The views were divided into specific levels, and their division criteria along with the correlation test results are presented in the table below. The correlation coefficients for Views_lev1, Views_lev2, and Views_lev4 with sales are 0.344, 0.430, and -0.314, respectively, suggesting a moderate correlation, while the other two levels show a weak correlation with sales.

Additionally, other features such as influencer likes, the like-to-fan ratio, and the total number of items sold are also included as statistical quantities in this model evaluation, as shown in Table 2.

Table 2: Output of correlation coefficients

Eigenvalue	**Definition**	**Correlation coefficient**
Likes	Likes of Influencer Accounts	0.179
Likes to Fans Ratio	Likes-to-fans ratio of the influencer's account	0.125
Total Products	Total number of items of the influencer's account	0.093

5 Experimental results

We will now construct a machine learning model to explore further the relationship between influencers' behavioural characteristics and sales. This paper makes use of a visual tool [4] to cover tasks from Data Analytics (DA) [5] within Data Engineering. This section mainly introduces the grouping of the data and the division of the comparison experiments, which will be used to make the prerequisite assumptions for constructing the model in the following two sections. This paper will use decision trees, random forests, and k-NN models to conduct comparative experiments and predict the influencers' sales levels. Therefore, this section will divide the data into test and training sets. The ratio of the model training set and test set is 8:2, i.e., 80% of the original data is used to train the model, which is represented by the feature dataset X_train and the dependent variable set Y_train, respectively, and 20% of the data is used to test the model, which X_test and Y_test, respectively represent. The model predicts the features of the test data X_test to produce the prediction data set Y_pred and then evaluates the classification effect of the model by comparing the test result Y_pred with the actual data Y_test.

We explore the relationship between influencer sales and behavioural characteristics. The experiments in this paper are conducted through Rapidminer [1] and are categorized into the following five sets of comparison experiments: i) full-time full-feature model: All the features of influencers are put into the model classifier to investigate the influence of the overall features of influencers on the sales level, ii) personal attribute information feature model: Train the classification model using personal attribute information features to explore how personal attribute information features affect influencer sales level prediction, iii) non-marketing behavioural features model: Input influencers' non-marketing behavioural features and basic attribute information into the training classification model to explore the influence of influencers' daily behavioural features on their sales level prediction, iv) marketing behavioral features model: Input all the personal information characteristics of the influencers and the marketing characteristics of the last three months into the classification model to compare the prediction results of the full-time full-featured model, v) pure marketing feature model: In the classification model, only the corresponding amount of marketing features is inputted to compare the prediction results with the marketing behaviour feature model. Prediction results through a decision tree are shown in Table 3.

Table 3: Output of decision tree model

Norm	Full-time full-feature model	Personal attribute information features model	Non-marketing behavioural features model	Marketing behavioral features model	Pure marketing features mode
RMSE	17.759	28.636	27.644	13.525	13.200
Correlation	0.814	0.571	0.540	0.913	0.896
MAE	12.693	22.645	20.6	10.150	10.041
Spearman_rho	0.684	0.657	0.638	0.882	0.898

From the evaluation indexes of the models, we have the following observations:

I. From the comparison of the indicators, it can be found that the pure marketing characteristics model is optimal in all indicators; RSME and MAE are the lowest, while Correlation and Spearman_rho are 0.896 and 0.898, respectively, which indicate a strong correlation. The RMSE of the pure marketing characteristics model can reach 13.2 at the lowest, the RMSE of the marketing behaviour characteristics model is 13.525, which is second only to that of the pure marketing characteristics model, and the RSME of the two is very close to each other, which indicates that the influencer's marketing behaviours are more able to influence its sales level. However, the influencer's sales level is not only influenced by the marketing behaviours.

II. The RSME of the non-marketing behavioural characteristics model and the personal attribute information characteristics model is the highest among all models, which indicates that it is impossible to comprehensively measure the commercial value of an influencer by simply relying on the influencer's influence or simply relying on the influencer's daily behavioural characteristics while ignoring other behavioural manifestations.

III. The RSME of the marketing behavioural characteristics model is lower than that of the full-time full-featured model, which indicates that the marketing behavioural characteristics model is better than the full-time full-featured model, further reflecting that the influencer's daily behavioural characteristics are not the first factor we consider when we consider the influencer's sales ability.

As already mentioned, this paper uses Rapidminer to process the data with the Random Forest algorithm. In order to better validate the model performance, this paper will compare the evaluation of the model under different classifiers by using

RMSE, Correlation, MAE, and Spearman_rho as the model performance metrics, and the output is shown in Table 4.

From the evaluation indexes of the model models, we have the following observations: i) From the comparison of the indicators, we can find that the pure marketing feature model is optimal in all indicators, RSME and MAE are the lowest, while Correlation and Spearman_rho are the highest, the RMSE of the pure marketing feature model can reach 13.459 at the lowest level. The RMSE of the marketing behavioural feature model is 19.640, which is only second to the pure marketing The RMSE of the pure marketing feature model is lower than that of the marketing behavioural feature model, which indicates that the influencer's marketing behavioural features are more likely to affect their sales level. Meanwhile, the level of sales of influencers is mainly influenced by marketing behavior; ii) The RSME of the non-marketing behaviour feature model and the personal attribute information feature model is the highest among all models, so it is the same as the conclusion obtained from the decision tree model, i.e., the influencer's personal information or daily behaviours cannot be used as an indicator to judge the influencer's sales ability alone; iii) The RSME of the full-time, full-feature model and the marketing behaviour feature model are very close to each other, and the marketing behaviour feature model is slightly better than the full-time, full-feature model, which shows that the level of the influencer's sales is not only affected by personal information and marketing behaviour.

Random Forest ranks in the top 10 regarding the importance of feature quantity, as shown in Table 5. The top 10 features include five marketing behaviour features, two daily behaviour features, and three basic personal information features. The number of sales is ranked first and is the most crucial feature quantity, and in the correlation analysis above, we also learned that there is a strong correlation between the number of sales and sales, and here again, the importance of the number of sales on sales is demonstrated. The remaining four marketing characteristics indicate that the interactivity and popularity of the marketing video are also essential. Among the three personal characteristics of the influencer, the label represents the personal attributes of the influencer, the play count represents the popularity of the influencer's video, and the number of likes reflects not only the popularity of the video but also the stickiness of the fans, which indicates that the better the stickiness of the fans is, the higher the quality of the video, and it can also contribute to the level of sales to a certain extent. The evaluation results by means of a *K*-NN algorithm are reported in Table 6.

Table 4: Output of random forest model

Norm	**Full-time full-feature model**	**Personal attribute information features model**	**Non-marketing behavioural features model**	**Marketing behavioral features model**	**Pure marketing features model**
RMSE	19.650	25.741	23.999	19.640	13.459
Correlation	0.835	0.510	0.591	0.855	0.892
MAE	16.185	23.204	20.829	16.527	10.286
Spearman_rho	0.830	0.571	0.573	0.857	0.872

Table 5: Top 10 importance features

Eigenvalue
Sales volume
Hashtag
Views
Average comments(s)
Average likes(s)
Average view(s)
Average comment(d)
Likes
Average share(d)
Average video duration(s)

Table 6: Output of *K*-NN model

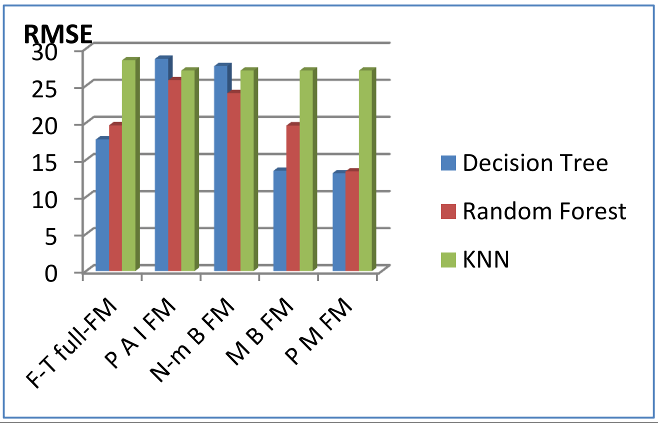
Norm	**Full-time full-feature model**	**Personal attribute information features model**	**Non-marketing behavioural features model**	**Market-ing behav-ioral fea-tures model**	**Pure marketing features model**
RMSE	28.446	27.045	27.045	27.039	27.037
Correlation	0.422	0.516	0.516	0.517	0.517
MAE	22.942	22.473	22.470	22.506	22.506
Spearman_rho	0.523	0.532	0.532	0.532	0.532

From the evaluation metrics of the model outputs, the following are the observations: i) The overall performance of each group of comparison experiments in the K-NN algorithm is close to each other; the RMSE of the pure marketing feature model is still the lowest but higher than the prediction results of the decision tree and the random forest, and the RMSE of the marketing behaviour feature model is only second to that of the pure marketing feature model, which further illustrates that the marketing behaviour is the main factor influencing influencers' sales level; ii) The RMSE of the personal attribute information model and the non-marketing behavioural characteristics model is the same at 27.045, while the RMSE of the full-time full-feature model is the highest at 28.446. Comparing the marketing behavioural characteristics model and the full-time full-feature model further shows that the daily behavioural characteristics are not the primary factor we consider when considering the influencer's sales ability; iii) By comparing the RSME, MAE, Correlation and Spearman_rho of *K*-NN with those of Random Forest and Decision Tree, Random Forest and Decision Tree are more effective in this classification.

Figure 1 plots the experimentation summary related to test RMSE.

6 Conclusions

This paper acquired, modeled, and analyzed influencers' data to assess the factors that influence their sales performance. The key research findings are divided into three parts: a) we identified a list of eligible influencers through data from a third-party platform, then collected their personal information and marketing data over the past three months as our primary research subject; b) we used the influencers' sales data from the past three months as the dependent variable, with the remaining influencer data serving as independent variables. Descriptive statistical analysis was conducted to demonstrate behavioral differences among influencers; c) we constructed marketing behavior characteristics, non-marketing behavior characteristics, and personal information characteristics by analyzing influencers' behaviors. Correlation analysis with sales was performed, and subsequently, we established a sales volume assessment model using decision tree, random forest, and K-NN algorithms, which were then evaluated.



Full-time full-features model	**Personal attribute information features model**	**Non-marketing behavioural features model**	**Marketing behavioral features model**	**Pure marketing features model**

Figure 1: Test RMSE in different scenarios.

The significance of this paper lies in the following contributions: first, we constructed distinct behavioral characteristics by analyzing influencers' behaviors, combined with personal information, and conducted descriptive statistical analysis and correlation tests with sales volume. These features were then used to build a model to predict influencers' sales levels. The predictive effectiveness of the model was assessed using machine learning classification metrics, and the importance of each feature was ranked. Specifically, i) through correlation analysis, we found that the number of sales and the sales volume are strongly correlated; the average number of likes, the average duration of marketing videos, and the average number of views of daily videos are moderately correlated with sales. Fans lev1, fans lev3, Views_lev1, Views_lev2, and Views_lev4 also show moderate correlation with sales, while other behavioral characteristics are weakly correlated. This indicates that the number of sales is the most critical factor influencing an influencer's success; ii) the decision tree, random forest, and K-NN algorithms were used to predict influencer sales based on behavioral data from the past three months. The results showed that the decision tree provided the best classification, followed by random forest and K-NN models. The consistent conclusion across all three algorithms is that marketing behavior is the primary factor in determining an influencer's sales capacity, while daily behavioral characteristics are less significant; iii) the random forest algorithm identified the top ten features in terms of importance, including five marketing behavior features, two daily behavior features, and three basic personal information features. This underscores the greater impact of marketing behavior on influencer sales, with the label and the number of views among the three basic personal information features ranking second and third, respectively, indicating the importance of the influencer's label and content popularity. This study provides influencers with a means to evaluate their current performance. It also offers valuable insights for aspiring influencers, MCN organizations, and merchants seeking to collaborate with influencers. The practical significance of this paper is threefold: a) it provides a theoretical framework for empirical analysis in the field of influencer research; b) it quantitatively analyzes the impact of influencers' behavioral characteristics on e-commerce using machine learning modeling, thereby enriching quantitative analysis methods and empirical research in the influencer e-commerce domain; c) it offers a theoretical basis for MCN organizations and companies in optimizing their operations and selecting the right influencers for partnerships.

References

1. Akthar, F., Hahne, C.: Rapidminer 5 operator reference. *Rapid-I GmbH*, *50*, 65 (2012).
2. Aw, E. C. X., & Agnihotri, R.: Influencer marketing research: review and future research agenda. *Journal of Marketing Theory and Practice*, 1-14 (2023).
3. Barta, S., Belanche, D., Fernández, A., & Flavián, M.: Influencer marketing on TikTok: The effectiveness of humor and followers' hedonic experience. *Journal of Retailing and Consumer Services*, *70*, 103149 (2023).
4. Cho, S.B., Tallón-Ballesteros, A.J.: Visual Tools to Lecture Data Analytics and Engineering. In *International Work-Conference on the Interplay Between Natural and Artificial Computation* (pp. 551-558). Springer, Cham (2017).
5. Dong, G., Liu, H.: Feature Engineering for Machine Learning and Data Analytics. CRC Press (2018).
6. Gillin, Paul: The new influencers: A marketer's guide to the new social media. Linden Publishing (2007).
7. Kubler, Kyle: Influencers and the attention economy: the meaning and management of attention on Instagram. *Journal of Marketing Management* 39.11-12: 965-981 (2023).
8. Lou, Chen, and Shupei Yuan: Influencer marketing: How message value and credibility affect consumer trust of branded content on social media. *Journal of interactive advertising* 19.1: 58-73 (2019).
9. Si, Ruo: *China livestreaming e-commerce industry insights*. Springer Singapore Pte. Limited, 2021.
10. Wang, D., & Guo, B.: Will online shopping lead to more brand loyalty than offline shopping? The role of uncertainty avoidance. *Journal of Marketing Research*, *61*(1), 92-109 (2024).
11. Zarrella, D.: *The social media marketing book*. O'Reilly Media, Inc. (2009).