

Research Article

Research on the Influencing Factors of Film Consumption and Box Office Forecast in the Digital Era: Based on the Perspective of Machine Learning and Model Integration

Qi He and Bin Hu 

School of Management, Shanghai University of Engineering Science, Shanghai 201620, China

Correspondence should be addressed to Bin Hu; 201771274@yangtzeu.edu.cn

Received 24 August 2021; Revised 12 September 2021; Accepted 15 September 2021; Published 14 October 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Qi He and Bin Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The film industry is one of the core industries of the digital creative industry, which has great positive externalities to the digital creative economy. Movie box office revenue is an important indicator to measure the realization of the market value of movie consumption, and it is also the basic guarantee for the sustainable development of the movie industry. This paper relies on the professional database of the Maoyan movie market to use Python software to collect a total of 830 domestic movie-related consumption characteristic data from 2017 to 2019. In this study, the stacking method in the machine learning ensemble algorithm combines the fivefold crossfolding training method based on distributed random forest, extremely randomized trees, and generalized linear models. The model is good at handling different data types. It has higher fitting and model accuracy in feature mining and model construction, so as to effectively grasp the relevant feature factors affecting movie consumption and accurately predict the future movie box office. Based on the innovative design method of model fusion, the extracted feature vector is used to build a more accurate movie box office prediction model through stacking with a fivefold crossfolding training method. It is aimed at opening the black box that affects the realization of the value of the film content consumption market in the digital age and putting forward corresponding countermeasures and suggestions.

1. Introduction

With the continuous development of digital technology, the digital transformation characterized by artificial intelligence [1–3] and big data applications has promoted the continuous evolution of the connotations, boundaries, and forms of creative economy and industrial development. The role of enhancing national competitiveness, promoting the development of industrial integration, and inducing new models and new business forms is increasingly deepening, and the impact on social development is becoming more and more profound. Vigorously promoting digital consumption has become an important driving engine for China to build a new development pattern that focuses on the domestic cycle and domestic and international dual cycles.

The film industry fully embodies the integration of humanities and art and technological innovation, the integration of traditional media and digital media, and the inte-

gration of producers and consumers. In the planning and classification of cultural and creative industries and digital content industries in different countries and regions, it has always been in the core category, which has great positive externalities to the digital economy. Movie products are a typical representative of the development of creative cultural products and digital content. Movie box office revenue is an important indicator to measure the realization of the consumer market value of the movie industry. As of 2019, the Chinese film industry has leaped to the second place in the world in terms of market size and has made important contributions to the economic benefits and social impact of the domestic digital content industry, although the development of the new crown epidemic in 2020 has affected the offline film industry to a certain extent. But at the same time, the reshaping of the film industry by digital transformation has penetrated the entire industry chain, profoundly changing the format and ecosystem of the film industry. The

operation logic of deep integration of technology and creativity is deeply rooted in the hearts of the people. With the gradual development of big data and artificial intelligence, digital technology has penetrated into the entire industry chain of production, distribution, and sales of the film industry, including the algorithm strategy to open the technical support for audiovisual streaming media on the online distribution of movies, as well as the opening of the artificial intelligence system's intervention in film production management such as box office prediction and audience positioning [4]. Since 2020, many film and television groups, including Hollywood giant Warner Bros., have built their own artificial intelligence project management systems, trying to gradually use artificial intelligence-related technologies to evaluate the value of content and main creation, so as to assist the decision-making reference of film distribution strategies [5].

However, the consumption of film products in the digital economy era is affected by multiple factors, and its box office forecasts are more challenging. Although previous studies have conducted a series of empirical analyses using statistical analysis methods and related indicators, the simple use of statistical analysis models is not enough to deconstruct the complex characteristics and structural relationships of film consumption under the new pattern. At present, there is still no method that comprehensively considers the comprehensive characteristics of film consumption in the context of digital transformation to conduct in-depth systematic research, which is insufficient for accurately grasping the characteristics of factors affecting digital content creative consumption and interpreting and predicting the future value of the box office. Therefore, based on the original research, this article systematically analyzes the multidimensional factors affecting film consumption in the digital age, relying on the professional database of the Maoyan film market, and comprehensively using the research methods of big data and machine learning to extract and construct the characteristics of relevant consumption influencing factors. Through model fusion training an innovative and enhanced predictive model, it attempts to build a research framework for the factors affecting film consumption in the context of digital transformation and opens the black box that affects movie box office. The main tasks of this study are as follows:

- (1) Data collection and preprocessing. The data sources of this research mainly come from the famous professional movie website Maoyan professional database, Sina Weibo, IMDB professional database, and WeChat official account platform. The data provided by these platforms were manually screened for unit and text errors, as well as data cleaning of error data, redundant data, and missing data during the data transmission process. A total of 830 movies were indexed
- (2) The stacking method in the machine learning ensemble algorithm combines the fivefold crossfolding training method based on distributed random

forest, extremely randomized trees, and generalized linear models. The model is good at handling different data types. This method has higher fitting and model accuracy in feature mining and model construction, so as to effectively grasp the relevant feature factors affecting movie consumption and accurately predict the future movie box office

2. Related Research

2.1. Research on Influencing Factors of Movie Box Office. The research on the factors affecting movie box office has a long history. It can be traced back to the 1940s. Early research mainly focused on research techniques [6]. For the first time, Gallup [7] and Handel [8] have systematically sorted out the influencing factors of movie box office such as actors, marketing, story, and evaluation to predict box office revenue. Subsequent scholars carried out in-depth research under this research framework. Generally speaking, the box office success of a movie is mainly based on three dimensions: the characteristics of the movie (such as director, star, screenwriter, and genre), the strength of the marketing strategy (mainly through advertising budget, number of screens, trailers, etc.), and reviews (from critics and movie audiences, etc.). Study the influencing factors of movie box office from both the supply side and the demand side of movie products. Researchers explored many potential influencing factors, including film origin, film cost, schedule, director's influence, award-winning influence, professional rating, word of mouth, genre, celebrity influence, film content, reviews, cultural familiarity, and consumer factors. Among them, the three major factors of celebrity influence, comment, and word of mouth have received extensive attention [9]. In addition, in view of the huge economic impact of film sequel products in the film industry, scholars have begun to study the impact of this factor [10]. With the explosive growth and development of digital technology, movie consumers can express their opinions or attitudes towards products across space and time. Therefore, in recent years, electronic word of mouth (eWOM) in the form of online reviews has increased exponentially [11]. Many researchers study the impact of eWOM indicators on box office performance. With the development of big data technology, more scholars use social media and digital marketing activities as influencing factors to predict box office [12]. On the whole, traditional box office forecasting studies use factors such as budget, actors, directors, producers, story locations, screenwriters, screening time, music, screening locations, target audiences, and sequels as variables. Research based on the background of digital transformation extends the influencing factors to include social media topics, search engines, marketing activities, and other variables with connotative characteristics of digital consumption.

2.2. Box Office Prediction Model Research. Early box office prediction methods were based on audience surveys. Since Litman et al. (1989) put forward a model that affects movie box office income factors and movie rental income through regression analysis [13], the research on movie box office

prediction model methods has continued to advance. Scott Sochay (1994) made improvements based on the above model [14]. Representative scholars such as de Vany and Walls (2004) used the OLS model, and Deuchert et al. (2005) proposed a two-stage model [15]. Other researchers have carried out extensive linear regression research on this basis. Ramesh et al. (2006) first proposed a box office prediction model using neural network methods, which opened up the research on innovative methods of box office prediction models in the digital age [16]. Based on big data and machine learning technology, the accuracy of the box office prediction model has been further improved. Choudhery et al. (2017) constructed a polynomial regression model for box office prediction by extracting chat data to analyze user emotions and other three methods [17]. Although the accuracy of the neural network model is improved compared with the previous two prediction models, the results are still not satisfactory.

In summary, there is a solid research foundation on the influencing factors of movie consumption and box office forecasting models, and the framework of the influencing factor evaluation system including the main creative team, movie characteristics, marketing promotion, and word-of-mouth comments has been basically formed. And in terms of research methods, the research methods of statistical measurement models such as market survey questionnaire interviews and linear regression have gradually expanded to neural networks, machine learning, and data fusion in the context of big data. However, in previous studies, different research methods have only considered the linear effects of some factors on box office forecasts, and empirical research on box office forecasting models using machine learning and model fusion on the basis of fully considering the complex and comprehensive digital age influence factors is relatively lacking. This has laid a certain theoretical foundation for this research from influencing factors to the improvement of research methods.

3. Design of Characteristic Index System for Influencing Factors of Movie Box Office in the Era of Digital Economy

Based on the mature experience of film product attribute feature selection at home and abroad, combined with consumers' personalized characteristics and aesthetic preferences, this study focuses on the impact of digital environment elements on consumption logic under the background of digital transformation; explores the three-dimensional feature factors of consumers, film products, and digital environment, which have a great impact on film consumption in the digital era; and constructs an index system. In order to ensure the comprehensiveness of the evaluation of the characteristics of the influencing factors of film product consumption in the digital age, first of all, according to the personal influencing factors of movie consumption generally mentioned in the existing literature, the indicators of gender, age, education level, active area, and preference type are selected to reflect the basic information of the per-

sonal characteristics of movie consumption, aesthetics and preferences, and the influence of herd atmosphere. Second, fully consider the determinants of the film's main creative team and the characteristics of the film product. The cultural awareness of the core creative subject directors, screenwriters, and main actors, such as word-of-mouth popularity, box office appeal, number of movies, release schedule, 3D, and IMAX factors, is added to the film product feature evaluation indicators. In this way, the original value, artistic value, experience and emotional symbolic value, and cultural recognition related to the characteristics of the film product are measured. Third, it focuses on the most important changes in film consumption under the influence of the digital economy era, such as online social support, social marketing activities, and digital opinion leaders. Include the marketing activities in the environmental characteristics of the digital age, the popularity of public opinion, the number of publicity placements under the influence of Internet word of mouth, the platform, the amount of broadcast, the public opinion evaluation and popularity of online media, the score of Internet word of mouth, and the schedule and other factors. Based on the above reasons and the availability of data, the evaluation index system for the influencing factors of film consumption characteristics in the digital era set in this study, that is, the follow-up characteristic data collection system setting, is shown in Table 1.

4. Machine Learning Fusion Forecasting Model Construction and Demonstration

4.1. Data Collection and Processing. The data sources for this research mainly come from the professional database of Maoyan, a well-known professional film website in China, Sina Weibo, IMDB professional database, and WeChat official account platform. The relevant professional database mainly provides timely, accurate, and professional film creation and box office data analysis for practitioners in the domestic and foreign film industry. Among them, the Maoyan database has fully opened up the online movie information database, which is more suitable for studying the influencing factors of domestic movie consumption. Sina Weibo and WeChat are mainly used as the source of digital environment feature collection. In order to fully reflect the impact of environmental changes in the digital economy era on film consumption characteristics, considering the comprehensiveness and continuity of the data, the sample collection interval is the index information related to the consumption characteristics of all domestic films from 2017 to 2019. Preliminary data collection uses Python to complete the data capture and analysis. First, collect information about the personal characteristics of each movie consumer displayed on the website, and secondly, collect the culture, experience, and cognition information of the movie, such as the main creators and the company's topical discussion on social media, historical box office, number of representative works and movie awards, IP information, genres, and sequels. In addition, collect information on external environmental elements such as related distribution and promotion and the representative work of a film distribution

TABLE 1: Setting of the characteristic index system of influencing factors of movie box office in the digital age.

First-level indicator	Secondary index	Three-level indicators	Explanation of related indicators
Personal consumption characteristics	Basic information Aesthetics and preferences Herd mentality	Gender	Consumer gender distribution information
		Age	Consumer age distribution information
		Education level	Consumer education level distribution information
		Active area	Consumer active area distribution information
		Preferred movie type	Consumers' favorite movie genres in the past
		Film awards or nominations	Refers to all valid awards won by the film at major film festivals
		Movie type	The categories, types, or forms of films formed due to different themes or techniques, including 13 categories such as action, science fiction, and comedy
		Visual effect	Whether it belongs to 3D, IMAX, or giant screen
		Whether adapted from IP	Whether the movie is adapted from classic classics, best-selling novels, animation works, game works, etc.
		Whether the sequel	Does the movie belong to the sequel of a certain series of movies
Features of movie products	Core cultural value Experience and emotional value Cultural awareness	Director's box office appeal	Director's cumulative historical box office
		Starring box office appeal	The top ten leading actors in their respective cumulative historical box office
		Screenwriter box office appeal	The top three screenwriters have their cumulative historical box office
		Director topic discussion volume	Number of director's online topic discussions
		Leading topic discussion volume	Top ten leading actors in their respective online topic discussions
		The volume of screenwriting topics	The top three screenwriters discuss their respective online topics
		Number of masterpieces produced by the company	Cumulative number of masterpieces from major production companies
		Number of masterpieces produced by the company	Cumulative quantity of all masterpieces of major production companies
		Number of representative works issued by the company	Cumulative number of masterpieces of major issuing companies
		Trailer run times	Trailer network runs
Digital environment characteristics	Marketing activities Popularity of public opinion Internet word of mouth	Total number of trailers played	Cumulative play volume of trailer
		Distribution platform	The distribution of trailer delivery platforms
		Cumulative number of popular Weibo	Number of related hot Weibo discussions
		Cumulative Weibo interactions	Number of Weibo interactions
		Weibo topic discussion volume	Number of topics discussed on Weibo
		Cumulative number of official account articles	Number of related public account articles
		Cumulative article reading	Cumulative reading volume of related public account articles
		Cat eye score	Word-of-mouth score of Maoyan movie website
		IMDB score	IMDB word-of-mouth score
		Screening time/schedule	Film first round screening schedule

company to identify the company's ability, as well as the number of promotional materials, quantity, platform, topic level of professional mass social media, public opinion popularity indicators, and the impact of the planned cycle of

movie schedules. Subsequently, manual screening of units and text errors, as well as data cleaning of erroneous data, redundant data, and missing data due to the data transmission process, was carried out, totaling all the index

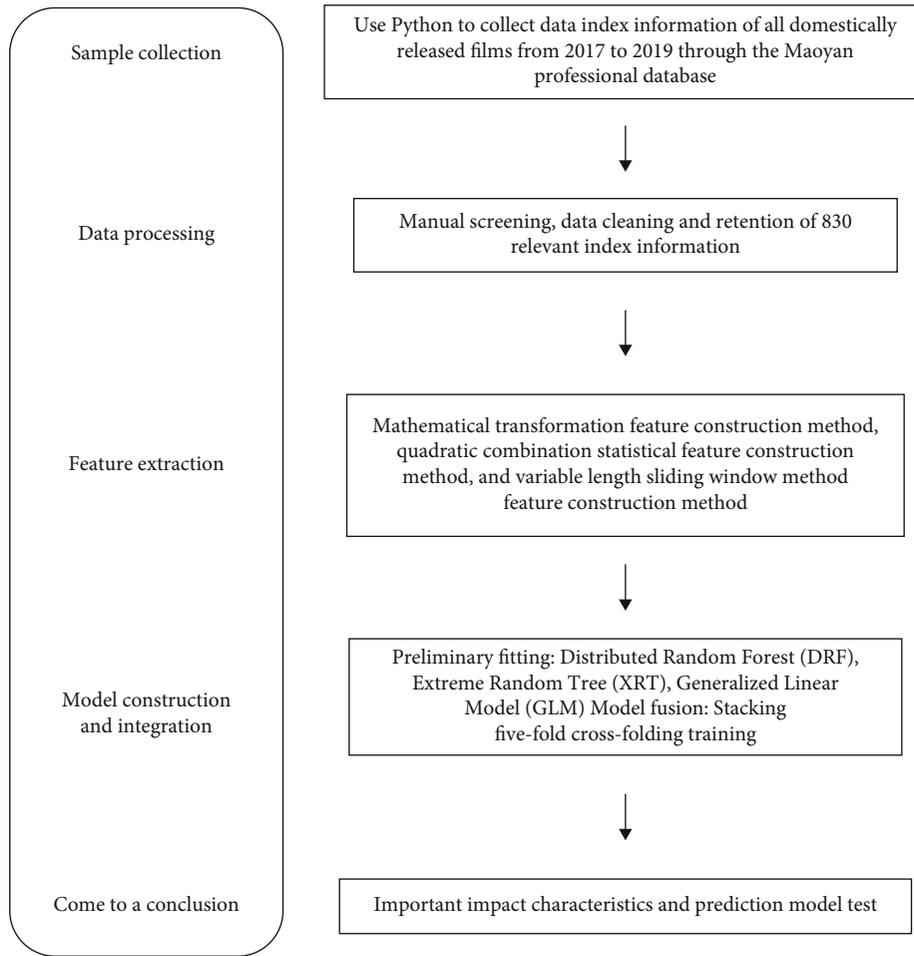


FIGURE 1: The research flow chart of machine learning and model fusion based on the analysis of the influencing factors of film consumption in the digital age.

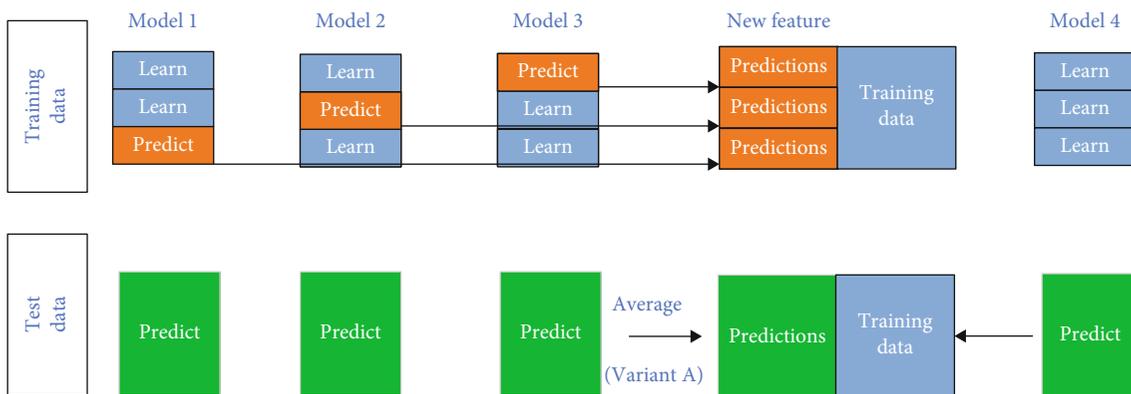


FIGURE 2: Stacking process model structure diagram.

information of 830 movies. In the future, new feature construction will be carried out according to research needs and specific scenarios. In view of the different data types having their own characteristics, different processing methods will be used to fit the research model.

4.2. *Research Method Selection.* The use of machine learning methods for movie box office prediction has made some research results in recent years, but most of the research only converts box office prediction from a regression problem to a classification problem. However, the use of classification

TABLE 2: Fitting effect analysis of distributed random tree forest model.

DRF: distributed random forest	
Model	DRF_1_AutoML_20200707_105158
Frame	automl_training_Key_Frame__movie_r2.hex
Description	Metrics reported on out-of-bag training samples
Model_category	Regression
MSE	0.039718
RMSE	0.199295
r2	0.941232
Mean_residual_deviance	0.039718
mae	0.134591
rmsle	0.023422

methods to predict the box office will lose a lot of characteristic information, which may cause certain restrictions on the use of prediction results. Feature engineering methods can extract core features, which have a vital impact on the accuracy of prediction models [18]. Through the innovative integration of machine learning feature engineering and regression models that process multiple data, it is more conducive to accurately assess the influencing factors and box office expectations of movie consumption in the digital age. Therefore, this study first uses Python computer program design language-related data directed crawler (requests + bs4 + re) library to complete the analysis of digital movie product consumers' personal characteristics, product characteristics, and digital environment network interaction behavior characteristics. Through manual screening, data cleaning, and preprocessing, combined with feature engineering research methods in the field of machine learning, Scikit-learn is used for feature extraction and feature construction. Then, according to the diversity of data types related to movie influencing factors. The innovation uses the stacking method in the machine learning integration algorithm to fuse models based on the fivefold cross-folding training method for distributed random tree forests (distributed random forest), extremely randomized trees (extremely randomized trees), and generalized linear model which are good at handling different data types. It has a higher fit and model accuracy in feature mining and model construction, so as to more effectively grasp the relevant feature factors that affect movie consumption and more accurately predict the future movie box office.

4.3. Research Idea Design. This research is based on the exploratory structure and deep insight of data characteristics and innovatively adopts the model fusion perspective for machine learning applications. The design of the research ideas is shown in Figure 1. First of all, comprehensive preliminary research and literature research combining the design of influencing factor index system and accurate data collection are the basic guarantees for the construction of feature engineering models. Good data preprocessing can explore the direction and accuracy of model training. Second, carry out data cleaning and screening to retain valid information. After that, input the effective data reflecting

the different influencing factors of the movie into different feature learning models, extract the features of the corresponding movie, and try to construct new features. Due to the very large scales of different types of variables in this study, exploratory statistical analysis found that data such as cumulative box office, first week box office, and star cumulative box office are all exponentially distributed, so logarithmic transformation of these features can construct new features. Finally, the stacking model fusion method is selected to construct the box office prediction model, and the three basic models are learned and fused by designing the feature vector fivefold crossfolding training method to construct a more accurate prediction model. In this way, it can more accurately identify the feature vector that conforms to the film consumption in the digital age and reveal the source of film box office revenue.

The processing of high-dimensional and complex data is a difficult point in machine learning. In traditional classification algorithms, it is difficult to deal with problems in practical applications, presenting problems such as low accuracy and overfitting. Stacking model is essentially a hierarchical structure, which is good at dealing with model fusion problems, and it is also particularly suitable for model training and learning that deal with multidimensional complex factors. Through the fitting and learning of different types of models, fusion builds an innovative fusion model that is more in line with the characteristics of the data. It is very suitable for the complex and multivariate influencing characteristic variable types of this research and the actual needs of accurate box office prediction. Figure 2 shows the basic process structure of this method.

4.4. Model Construction and Empirical Analysis. The feature engineering construction method of machine learning is used to analyze, collect and construct features, and determine which consumption features are the most important, which plays a role in the performance of the predictive model. It helps to avoid errors in human factor judgments and some inertial problems of traditional statistical measurement models and helps to obtain a more explanatory characteristic variable system. According to the data characteristics of the influencing factor index system, the following three types of classic models are used for fitting, respectively,

TABLE 3: Fitting effect analysis of extreme random tree model.

ERT: extremely randomized trees	
Model	ERT_1_AutoML_20200707_105158
Frame	automl_training_Key_Frame__movie_r2.hex
Description	Metrics reported on out-of-bag training samples
Model_category	Regression
MSE	0.037427
RMSE	0.193462
r2	0.944621
Mean_residual_deviance	0.037427
mae	0.132382
rmsle	0.022773

TABLE 4: Analysis of fitting effect of generalized linear model.

GLM: generalized linear model	
Model	GLM_1_AutoML_20200707_105158
Frame	automl_training_Key_Frame__movie_r2.hex
Description	.
Model_category	Regression
MSE	0.051241
RMSE	0.226365
r2	0.924182
Mean_residual_deviance	0.051241
mae	0.16952
rmsle	0.027431

TABLE 5: Analysis of fitting effect of fivefold crossfolding training fusion with three models.

Stacked ensemble	
Model	StackedEnsemble_AllModels_AutoML_20200707_105158
Frame	automl_training_Key_Frame__movie_r2.hex
Description	.
Model_category	Regression
MSE	0.005488
RMSE	0.074081
r2	0.99188
Mean_residual_deviance	0.005488
mae	0.050961
rmsle	0.008718

and the stacking model fusion method is used to perform fivefold crossfolding training on different models to construct a new fusion model. This makes the fusion model stronger in fusion and generalization and forms a model structure that is more suitable for the identification of influencing factors of film consumption and box office prediction in the digital age.

4.4.1. Distributed Random Tree Forecast Model Experiment. Bernard et al. proposed that stochastic forest is one of the

most classical data processing models of integrated learning algorithm. It provides users with reasonable and effective classification label information by using the integrated thought, thus providing reliable and effective data information recommendation [19]. Fernández-Delgado et al. found that the random forest algorithm has the best classification performance by comparing the classification performance of 179 classification algorithms [20]. Lizhi et al. found that the distributed random forest algorithm in Spark is more suitable for feature learning of two-dimensional variables

TABLE 6: Analysis of fusion combination factor for fivefold crossfolding training with three models.

Names	Coefficients	Standardized_ coefficients
Intercept	-0.1308	7.6077
XRT_1_AutoML_20200707_195606	0.9967	0.7847
DRF_1_AutoML_20200707_195606	0.0043	0.0033
GLM_1_AutoML_20200707_195606	0.0166	0.0129

[21]. The data collection conforms to the data structure and characteristics of movie consumption factors in the digital age. The empirical study also shows a good fit effect. Table 2 shows that the goodness of fit with this model reaches about 94.12%, and the prediction error RMSE of the model reaches 19.9%.

4.4.2. Extreme Random Tree Prediction Model Experiments. The extreme random tree algorithm proposed by Geurts et al. [22] is very similar to the random forest algorithm, but the extreme random tree features are randomly selected. Selecting the best partitioning feature with the specified threshold as the optimal partitioning attribute not only guarantees the utilization of training samples but also reduces the final prediction bias, so it is superior to the results obtained by random forests to some extent. Therefore, it is also used as a prediction model method to carry out experiments. The results obtained in this paper also conform to a high level of goodness of fit, basically reaching about 94.46%, and the RMSE prediction error reaches 19.3%, as shown in Table 3.

4.4.3. Experiments of Generalized Linear Prediction Model. The generalized linear model is an extension of the general linear model. It establishes the relationship between the expected value of the response variable and the predicted variable of the linear combination through a join function. It is characterized by not forcibly changing the natural measures of the data. The data may have a nonlinear and non-constant variance structure, or it may be the most popular machine learning algorithm at present. This study also uses this algorithm to fit according to the structure characteristics of the data indicators. The results of the analysis are relatively consistent with the data characteristics, reaching a good fit of 92.41%, but the RMSE prediction error is as high as 22.63% (see Table 4).

4.4.4. Triple Model Fusion Fivefold Crosstraining Experiment. Fitting the data of movie consumption characteristics with the above three models, we can find that, first, the selection of the initial index system is more effective, making these basic characteristics representing movie consumption more regular. At the same time, the three algorithms have more than 90% fitting accuracy and strong explanatory power, but there is still room for further improvement of prediction

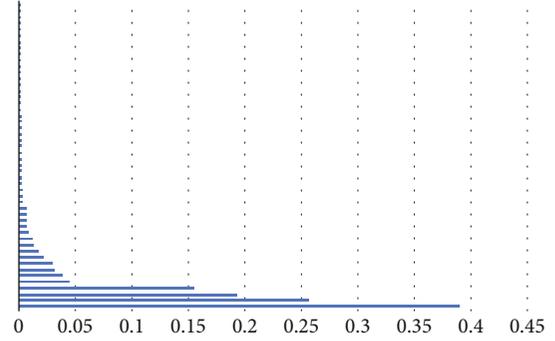


FIGURE 3: Feature extraction of influencing factors of movie consumption characteristics in the digital age.

accuracy. In order to further explore the consumption characteristics, the stacking model fusion method was used to train the three models with fivefold crossfolding, and a more accurate model was obtained. The goodness of fit reached 99.18%, while the RMSE was 7.4%, and the RMSLE was significantly lower than the classification prediction error of the first three classical models which was only 0.8%. This fully demonstrates that the features extracted by learning from this model are very consistent with the characteristics and actual results of the database. The model has more generalization ability and basically matches the feature structure of the current factors affecting movie consumption perfectly (see Table 5).

The results of this fusion model are integrated by the three model algorithms mentioned above, and their combination coefficients are shown in Table 6.

4.5. Results and Discussion. Based on the above model learning results, it can be found that through innovative model fusion training, the goodness of fit is higher and the prediction bias is lower than that of a single prediction model. In the digital economy era, the extraction of influencing factors of movie consumption is more accurate and can provide more effective box office prediction model scheme. Based on the analysis results, this study further discusses and analyzes which extracting features can better reflect the explanatory and influence of movie consumption in the digital age. Feature extraction and learning are carried out through different models. Important indicators in the influence feature variables of digital movie consumption are shown in Figure 3.

Therefore, it can be found that the most influential feature is the cumulative historical box office of star writers in the main body of the core content creator, which fully reflects the importance of the current market on the core content. First, the writer is the core creator of the current creative source of digital content products, and also the source of IP core stories. Past box office represents the creative ability of the writer, the cultural and artistic value of the work and the docking ability of the market, highlighting the significance of content as king.

Second, digital marketing has become an important influencing factor of movie consumption. Essential changes have taken place in the form of movie marketing promotion

in the digital age. The broadcast of marketing materials put on the Internet has become an important feature affecting movie consumption. Movie consumption in the digital age has a wider audience. In the market of digital content products, the voice and influence of social media play an important role. Precise delivery and distribution mechanism based on Internet platform can help achieve digital marketing effect. Thirdly, the cumulative historical box office of star creators shows that the past artistic performance and recognition of star creators are very important, and star is still the core value creator of content products. Fourth, the type of movie still has an important impact. Although this factor has been proven by many studies to be closely related to movie consumption, special types such as love, action, and science fiction still become an important factor to trigger the resonance of movie consumption and stimulate market vitality. Fifth, hot spots of public opinion have become important variables affecting movie consumption, including different types of self-media comments and word-of-mouth communication and discussion, such as Weixin Public Number and Weibo topic discussion.

5. Conclusions

Based on the empirical results of these factors affecting movie consumption, the following suggestions are given on how to improve the consumption related to digital content in the future:

First of all, attach great importance and increase capital investment to creative subjects of high-quality cultural content and treat the flow effect with caution. Along with the continuous innovation of digital content form, it brings consumers a more pleasant consumption experience but also changes people's traditional consumption habits and consumption concepts. Secondly, further standardize the network environment and strengthen the network ecological governance. The most important influencing factors of consumption of digital content products are the guidance and evaluation of network public opinion. The network environment should further be standardized; major movie and television websites should do a good job in related management, focus on "zombie" number and accounts with malicious scoring records, and rectify the black industry chain that the network breeds. Third, encourage the construction of a diversified digital content value evaluation system. For cultural creative consumers, big data on the Internet only means the display and prediction of large probability events, which can only be used as reference. Digital content products are essentially cultural creative products. Its cultural value and aesthetic experience cannot be pale and shallow only represented by a series of data. Finally, encourage content providers such as digital content creative subject, production producer, and dissemination subject to adhere to the original intention of content creation. Make good use of digital diffusion channels and create a win-win situation between content providers and consumers by using "big data." However, the prediction model used in this article is sensitive to noise, and the prediction accuracy needs to be

further improved. These two shortcomings are also the direction for future work.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

All the authors do not have any possible conflicts of interest.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 71704102).

References

- [1] M. Zhao, A. Jha, Q. Liu et al., "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Medical Image Analysis*, vol. 71, article 102048, 2021.
- [2] Z. Chu, M. Hu, and X. Chen, "Robotic grasp detection using a novel two-stage approach," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 19–29, 2021.
- [3] M. Zhao, Q. Liu, A. Jha et al., "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, <https://arxiv.org/abs/2106.11480>.
- [4] W. Wang and D. Bin, "Machine green light system' and' algorithmic matrix movie' - the impact of artificial intelligence on the film production industry," *Contemporary Movies*, vol. 12, pp. 30–36, 2020.
- [5] T. Siegel and W. Bros, *Signs Deal for AI-Driven Film Management System*, no. 1, 2020Hollywood Reporter, 2020.
- [6] R. Handel et al., *How Hollywood Understands Audiences*, Chinese Press, 2014.
- [7] B. Ayoub, "George Gallup in Hollywood," *Pacific Historical Review*, vol. 77, no. 4, pp. 693–695, 2008.
- [8] L. A. Handel, *Hollywood Looks at its Audience. A Report of Film Audience Research*, The University of Illinois Press, Urbana, IL, 1950.
- [9] F. Peng, L. Kang, S. Anwar, and X. Li, "Star power and box office revenues: evidence from China," *Journal of Cultural Economics*, vol. 43, no. 2, pp. 247–278, 2019.
- [10] B. Belvaux and R. Mencarelli, "Prevision model and empirical test of box office results for sequels," *Journal of Business Research*, vol. 130, no. 1, pp. 38–48, 2021.
- [11] H. Ma, J. M. Kim, and E. Lee, "Analyzing dynamic review manipulation and its impact on movie box office revenue," *Electronic Commerce Research and Applications*, vol. 35, article 100840, 2019.
- [12] Z. Wang, J. Zhang, S. Ji, C. Meng, T. Li, and Y. Zheng, "Predicting and ranking box office revenue of movies based on big data," *Information Fusion*, vol. 60, pp. 25–40, 2020.
- [13] B. R. Litman and L. S. Kohl, "Predicting financial success of motion pictures: The '80s experience," *Journal of Media Economics*, no. 2, pp. 35–50, 1989.
- [14] S. Sochay, "Predicting the performance of motion pictures," *Journal of Media Economics*, vol. 7, no. 4, pp. 1–20.

- [15] A. de Vany and W. D. Walls, "Uncertainty in the movie industry: does star power reduce the terror of the box office?," *Journal of Cultural Economics*, vol. 23, no. 4, pp. 285–318, 1999.
- [16] S. Ramesh and D. Delen, "Predicting box-office success of motion pictures with neural networks," *Expert Systems with Applications*, vol. 30, no. 2, pp. 243–254, 2006.
- [17] D. Choudhery and C. K. Leung, "Social Media Mining: Prediction of Box Office Revenue," in *Proceedings of the 21st International Database Engineering & Applications Symposium on - IDEAS 2017*, 2017.
- [18] Y. Ru, B. Li, J. Chai, and J. Liu, "A movie box office prediction model based on in-depth learning," *Journal of China Media University: Natural Science Edition*, vol. 26, no. 1, pp. 30–35, 2019.
- [19] S. Bernard, S. Adam, and L. Heutte, "Dynamic random forests," *Pattern Recognition Letters*, vol. 33, no. 12, pp. 1580–1586, 2012.
- [20] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, "Do we need hundreds of classifiers to solve real world classification problems?," *Journal of Machine Learning Research*, vol. 15, pp. 3133–3181, 2014.
- [21] M. Lizhi, D. Jiyao, and L. Chong, "Breast cancer risk prediction analysis based on Spark and random forest," *Computer Technology and Development*, vol. 29, no. 8, pp. 142–146, 2019.
- [22] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006.