ISSN 2616-5775 Vol. 7, Issue 7: 98-104, DOI: 10.25236/AJCIS.2024.070713

Research on the Application of Machine Learning Algorithm in the Prediction of College Students' Consumption Behavior

Xiaoyan Liu[1,2], Lilibeth Cuison[1,*]

[1]*Graduate School, Angeles University Foundation, Angeles City, 2009, Philippines*
[2]*Department of Virtual Reality and Gaming, Jiangxi Tellhow Animation College, Nanchang, Jiangxi Province, 300200, China*
[*]*Corresponding author*

Abstract: *In response to the challenges faced by machine learning algorithms in predicting college students' consumption behavior, corresponding optimization strategies are proposed. These strategies aim to enhance data quality and stability through data augmentation and preprocessing. By employing model integration and optimization techniques, the predictive and generalization abilities of the models are strengthened, while efforts are made to increase their interpretability and practicality. This paper aims to provide an effective optimization scheme for predicting college students' consumption behavior using machine learning algorithms.*

Keywords: *prediction of college students' consumption behavior; Machine learning algorithm; Model selection; Model interpretation*

1. Introduction

With the development of Internet and big data technology, machine learning algorithm has been widely used in the field of college students' consumption behavior prediction, especially in college students' social networks, economic conditions, personal preferences and other consumption behaviors are affected by many factors. Traditional statistical methods have limitations in dealing with complex consumer behavior prediction problems, while machine learning algorithms can improve the accuracy and precision of prediction by learning a large number of data and establishing complex models. Machine learning faces many challenges in the prediction of college students' consumption behavior, such as data quality and stability, the complexity of model selection and effect evaluation, and the need for model interpretation and interpretability. This paper will analyze and discuss these challenges, and put forward corresponding solutions and methods. This paper will discuss the importance and influence of data quality and stability on machine learning prediction model. This paper discusses the strategies to improve the interpretability and interpretability of the model, so that decision makers can better understand and trust the prediction results of the model. Through these analyses and discussions, this paper aims to provide practical suggestions and guidance for researchers and practitioners on how to optimize the machine learning application of college students' consumption behavior prediction.

2. Characteristics of machine learning algorithm in college students' consumption behavior prediction

2.1 Mining personalized features

As one of the important applications of machine learning in the field of college students' consumption, the mining of personalized features is of great significance for improving the accuracy and practicability of the prediction model. Because of the diversity of age, region and cultural background, the consumption behavior of college students is complex and changeable, and it is often difficult for traditional statistical analysis methods to capture these personalized consumption habits and trends. By analyzing and learning big data, machine learning algorithm can effectively mine hidden personalized features from massive data, so as to predict and interpret college students' consumption behavior more accurately. The mining of personalized features involves the comprehensive application of multiple technologies and methods. Data collection and processing is the basis of personalized

ISSN 2616-5775 Vol. 7, Issue 7: 98-104, DOI: 10.25236/AJCIS.2024.070713

feature mining. By collecting large-scale consumption data including multi-dimensional information such as consumption amount, consumption frequency and consumption time, we can build a comprehensive and in-depth understanding of college students' consumption behavior. Cluster analysis, association rule mining and deep learning model of machine learning algorithm can identify unique consumption preferences and behavior patterns for different consumer groups and individuals [1].

On the other hand, the mining of personalized features also needs to fully consider the dynamic and real-time nature of data. With the change of college students' consumption behavior and the influence of social environment, the consumption characteristics may change significantly, and the traditional static model can't cope with this dynamic change. Machine learning model needs to have good adaptive ability, and can analyze and adjust the consumer behavior prediction model in real time to cope with the ever-changing consumer market [2]. The mining of personalized features is not limited to data analysis technology, but also needs to be deeply understood and explained by combining the knowledge and experience of domain experts. The features mined by machine learning algorithms often need to be explained by professional domain knowledge, which not only helps to improve the interpretability and understandability of the model, but also provides more accurate and targeted suggestions and strategies for consumer behavior prediction.

2.2 Analysis of time series and periodicity

In the prediction of college students' consumption behavior, machine learning algorithm can effectively capture and analyze the temporal and periodic characteristics of consumption behavior. Time series analysis mainly focuses on the pattern of consumption behavior changing with time, while periodic analysis focuses on the repetitive law of consumption behavior changing with time. By analyzing the time series and periodicity of college students' consumption behavior, we can better understand the changing law of consumption behavior and predict the future consumption trend more accurately. By analyzing historical consumption data, machine learning algorithm can find the time series pattern of consumption behavior [3]. This model can help consumers understand the changing trend of consumption behavior in different time periods, and the influence of special periods such as shopping season and exam week on consumption behavior. Through the recognition of these time series patterns, enterprises and businesses can adjust their commodity stocking and marketing strategies in advance to meet the consumption needs of college students.

The machine learning algorithm can also discover the periodicity of consumption behavior. The periodicity of consumption behavior may be caused by the start and end of school semester, holiday shopping season, periodic events in college students' lives and other factors. Through the analysis of historical consumption data, machine learning algorithm can identify these cyclical laws and predict future consumption trends. By analyzing the historical sales data, enterprises can predict the sales changes in each shopping season, so as to make the adjustment of promotion activities and commodity stocking in advance. When analyzing time series and periodicity, machine learning algorithm can adopt time series prediction model and periodicity prediction model. Time series prediction models, such as ARIMA model and LSTM model, can capture the pattern of consumption behavior changing with time and predict the future consumption trend. Periodic prediction models, such as Fourier analysis and seasonal decomposition, can capture the periodic laws in consumption behavior and predict the future consumption trend. Through the application of these models, machine learning algorithm can provide more accurate consumer behavior prediction and provide targeted suggestions and strategies for enterprises and individuals. It is helpful for enterprises and businesses to better meet the consumption needs of college students and improve sales and customer satisfaction.

2.3 Comprehensive analysis of multi-factor influence

In the prediction of college students' consumption behavior, machine learning algorithm can comprehensively analyze the influence of various factors on consumption behavior, thus providing more accurate prediction results. These factors include but are not limited to economic factors, social factors and psychological factors.Through the comprehensive analysis of these factors, machine learning algorithm can reveal the interaction and influence mechanism between different factors and provide targeted suggestions and strategies for enterprises and individuals. Economic factors have an important influence on college students' consumption behavior. Family income, personal scholarships, loans and other economic factors will affect the consumption ability and propensity of college students. Machine learning algorithm can comprehensively analyze these economic factors and predict the consumption behavior of college students under different economic conditions. Through the analysis of

ISSN 2616-5775 Vol. 7, Issue 7: 98-104, DOI: 10.25236/AJCIS.2024.070713

historical consumption data, machine learning algorithm can find the correlation between family income and college students' consumption expenditure,thus helping enterprises and businesses to formulate corresponding marketing strategies[4].

Social factors also have an important impact on college students' consumption behavior.Social factors such as the interaction between classmates, public opinion and fashion trends will all affect college students' consumption decisions. Machine learning algorithm can capture the influence of these social factors on consumption behavior and bring it into the prediction model.By analyzing the discussions and comments on social media, the machine learning algorithm can understand the evaluation and preferences of college students on a product, thus providing targeted product promotion strategies for enterprises. Psychological factor is another important factor that affects college students' consumption behavior. Psychological factors such as personality characteristics, mood swings and values will all affect college students' consumption behavior. Machine learning algorithm can comprehensively analyze these psychological factors and predict the consumption behavior of college students in different psychological states[5].By analyzing college students' emotional fluctuations and consumption records, machine learning algorithm can find the influence of emotions on consumption behavior, thus helping enterprises and businesses to formulate corresponding promotion strategies.The machine learning algorithm flow is shown in Figure 1.

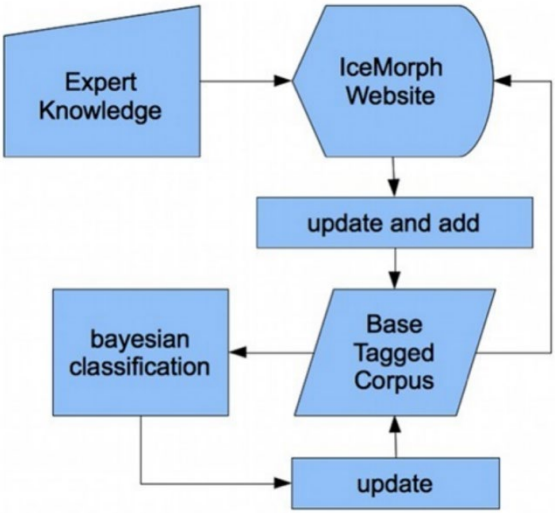

Figure 1: Machine learning algorithm flow

3. Problems of machine learning algorithm in college students' consumption behavior prediction

3.1 Data quality and stability

Data quality has an important influence on the prediction results of machine learning algorithm. In the prediction of college students' consumption behavior, the accuracy, completeness and representativeness of the data have an important influence on the accuracy of the prediction results. There are often data quality problems in practical applications. There may be missing values, abnormal values or noise in the data, which may lead to the wrong prediction results of machine learning algorithms. Before machine learning analysis, it is necessary to check and clean the data to improve the quality of the data. The stability of data is also a problem that machine learning algorithm faces in the prediction of college students' consumption behavior. Consumption behavior is a dynamic process, which is influenced by many factors, such as economic environment, social events and personal mood. Changes in these factors may lead to fluctuations and instability in the data of consumer behavior.

Machine learning algorithm needs stable data to train and predict, and needs to analyze the stability of data, and adopt corresponding processing methods such as data smoothing and time series decomposition to improve the stability of data. The acquisition and collection of data is also a challenge. The data of college students' consumption behavior usually comes from shopping websites, bank accounts, social media and other channels. These data may have different formats and structures, and need to be integrated and preprocessed. The acquisition and collection of data may be restricted by privacy protection and data access rights, so it is necessary to find suitable data sources and establish a reliable data collection and update mechanism to ensure the quality and stability of data.

ISSN 2616-5775 Vol. 7, Issue 7: 98-104, DOI: 10.25236/AJCIS.2024.070713

3.2 Model selection and effect evaluation

Model selection is a key step in the application of machine learning algorithm. For different consumer behavior prediction problems, we can choose different machine learning models such as linear regression, decision tree, support vector machine and neural network. Model selection is not a simple matter. Different models have different characteristics and applicable scenarios, and need to be selected according to specific problems and data characteristics. Model selection often depends on prior knowledge and experience, and it may become a difficult problem for users who lack relevant background knowledge. It is necessary to develop some automatic or semi-automatic model selection methods to help users choose the appropriate model.

Effect evaluation is an important means to evaluate the performance of machine learning model. The effect evaluation in the prediction of college students' consumption behavior can help us understand the prediction ability of the model, so as to adjust and optimize the model. Effect evaluation is also a complicated process. Select appropriate evaluation indicators such as accuracy, recall and F1 value. Different evaluation indicators may lead to different evaluation results, so it is necessary to select appropriate evaluation indicators according to specific problems and needs. The evaluation process needs to consider the generalization ability of the model, that is, the prediction ability of the model on unknown data. In order to evaluate the generalization ability of the model, cross-validation or set-aside method is usually needed. These methods may increase the calculation and time cost of evaluation.

3.3 Model interpretability and interpretability

In the prediction of college students' consumption behavior, the interpretability of the model is an important issue in the application of machine learning algorithm. Although machine learning algorithm has high accuracy and stability in predicting consumption behavior, the interpretability and interpretability of the model has become a challenge in its practical application. The interpretability of the model means that the model can provide the explanation and understanding of the prediction results. Users want to know how the model obtains the prediction results from the input features, so as to increase the trust and acceptance of the model. Many machine learning algorithms, especially deep learning algorithms, are often regarded as "black box" models, and it is difficult to provide an explanation for the prediction results. This may lead users to doubt the prediction results of the model, thus affecting the practical application of the model. Improving the interpretability of the model is an important problem that machine learning algorithm needs to solve in the prediction of college students' consumption behavior.

The interpretability of the model means that the model can provide interpretability for the prediction results. In the prediction of college students' consumption behavior, interpretability is of great significance for the practical application of the model and decision support. Users want to understand how the weights of the model are allocated from the input features, so as to have a deeper understanding of the prediction results of the model. Many machine learning algorithms are often difficult to provide an explanation of weight distribution. It may lead users to be confused about the prediction results of the model, thus affecting the practical application of the model. Improving the interpretability of the model is an important problem that machine learning algorithm needs to solve in the prediction of college students' consumption behavior.

4. Optimization of machine learning algorithm in college students' consumption behavior prediction

4.1 Data enhancement and preprocessing

In the prediction of college students' consumption behavior, data enhancement and preprocessing are important countermeasures to optimize the machine learning algorithm. By enhancing and preprocessing the data, the quality and stability of the data can be improved, thus improving the prediction accuracy and reliability of the model. Data enhancement refers to a series of transformation and processing of the original data to increase the diversity and expressive ability of the data. In the prediction of college students' consumption behavior, data enhancement can help the model to better learn the inherent laws and characteristics of data. The original data can be randomly disturbed, scaled and rotated to generate more diverse training samples. This can increase the generalization ability of

ISSN 2616-5775 Vol. 7, Issue 7: 98-104, DOI: 10.25236/AJCIS.2024.070713

the model and improve the prediction performance of the model on unknown data. Data enhancement can also simulate the data distribution in the real world by introducing noise and outliers, thus improving the robustness of the model.

Data preprocessing refers to cleaning and preparing data before model training. In the prediction of college students' consumption behavior, data preprocessing is of great significance to improve the quality and stability of data. It is necessary to interpolate, delete or estimate the missing value of data. The data needs to be filtered, smoothed or replaced with outliers. It is also necessary to normalize and standardize the data to eliminate the dimensional influence between different features and improve the training efficiency and prediction performance of the model. In addition to data enhancement and preprocessing, other optimization countermeasures can be adopted to improve the application effect of machine learning algorithm in college students' consumption behavior prediction. Integrated learning methods such as random forest and gradient hoist can be used to improve the overall prediction accuracy by combining the prediction results of multiple models. Transfer learning method can be used to initialize the weights by using the pre-trained model, thus speeding up the training speed of the model and improving the prediction performance.The data increase and method effect are shown in Table 1.

Table 1: Data Enhancement Methods and Effects

Optimization countermeasure category	Specific method	Purpose and effect
Data enhancement	random disturbance	Increase the diversity of data, help the model learn the inherent laws and characteristics of data, and improve the generalization ability.
	zoom	Change the data scale and enhance the recognition ability of the model to different scale features.
	rotate	Increase the adaptability of the model to the change of direction
	Introduce noise	Simulate the real world data distribution and improve the robustness of the model.
	Introducing outliers	The ability of training model to identify abnormal situations

4.2 Integration and optimization of models

Model integration and optimization are the key countermeasures to improve the performance of machine learning algorithm in the prediction of college students' consumption behavior. By integrating the prediction results of multiple models, the prediction accuracy and stability of the models can be effectively improved. Model integration refers to combining the prediction results of multiple models to obtain more reliable prediction results. The model integration method can effectively improve the generalization ability and robustness of the model in the prediction of college students' consumption behavior. The Stacking method can be used to input the prediction results of multiple models into a new model as new features to produce the final prediction results. Weighted average or voting method can be used to weight or vote according to the prediction results of different models to obtain the final prediction results.

In addition to model integration, model optimization is also an important countermeasure to improve the performance of machine learning algorithms. By adjusting the parameters and hyperparameters of the model, the prediction ability and generalization ability of the model can be improved. Grid Search or Random Search can be used to traverse the combination of parameters and superparameters of the model and select the optimal parameters and superparameters. Bayesian optimization method can be used to express the optimization problem of parameters and superparameters in a probabilistic way by constructing Bayesian optimization model, thus improving the optimization efficiency and accuracy of the model. In order to further improve the effect of model integration and optimization, some advanced optimization countermeasures can be adopted. Transfer learning method can be used to initialize the weights by using the pre-trained model, thus speeding up the training speed of the model and improving the prediction performance. Adaptive learning method can be used to dynamically adjust the learning rate and optimization algorithm according to the training progress and performance of the model to improve the training effect and prediction performance of the model.The specific optimization effects of the model are shown in Table 2.

ISSN 2616-5775 Vol. 7, Issue 7: 98-104, DOI: 10.25236/AJCIS.2024.070713

Table 2: Model optimization effect

Optimization Strategy	Description	Purpose and Effect
Model Integration	Optimizing the combination of multiple models.	Enhances prediction accuracy and generalization capability.
Parameter Tuning	Adjusting model parameters and hyperparameters.	Improves the model's predictive power and its ability to generalize to new data.
Grid Search	Systematically searching through a grid of parameter combinations.	Finds the optimal set of parameters by exhaustively searching the parameter space.
Random Search	Randomly sampling hyperparameter combinations.	Efficiently explores the hyperparameter space to find a good combination.

4.3 Enhancement of interpretability

Explanatory enhancement and application in college students' consumption behavior prediction is an important countermeasure to optimize machine learning algorithm. By improving the interpretability of the model, users' trust and acceptance of the model can be enhanced, thus improving the practical application effect of the model. Explanatory enhancement refers to improving the interpretability of machine learning model, so that users can understand and explain how the model obtains the prediction results from the input features. The interpretability of the model is very important for users' decision support and trust establishment. Explanatory machine learning algorithms such as decision tree and linear regression can be used to provide interpretation and understanding of the prediction results. Model interpretation tools and techniques such as feature importance analysis and weight visualization can be used to help users better understand the prediction results and weight distribution of the model.

In addition to explanatory enhancement, explanatory application is also an important countermeasure to improve the performance of machine learning algorithms. By combining interpretation with practical application, the practicability and user acceptance of the model can be improved. Interpretation can be applied to the decision-making process of the model to help users understand the prediction basis and decision-making logic of the model. Users can trust the prediction results of the model more and make corresponding decisions according to the suggestions of the model. The application of interpretation can also promote the continuous optimization and improvement of the model. By collecting users' feedback and requirements, the model can be further adjusted and optimized to meet users' practical application needs. In order to further improve the effect of explanatory enhancement and application, some advanced countermeasures can be adopted. Integrated learning method can be used to combine the interpretiveness of multiple models to obtain a more comprehensive explanatory perspective. Transfer learning method can be used to transfer the application experience in the pre-trained model to the target model to improve the explanatory ability of the model. User research can also be carried out to deeply understand users' needs and preferences, so as to design a more explanatory machine learning model.

5. Conclusion

This paper systematically discusses three key problems that machine learning algorithm faces in college students' consumption behavior prediction: data quality and stability, model selection and effect evaluation, model interpretation and interpretability. Through in-depth analysis of the challenges and solutions of each problem, this paper puts forward corresponding theoretical support and practical suggestions. Future research can further deepen the understanding of these problems and explore more accurate and effective solutions to promote the application and development of machine learning in consumer behavior prediction.

References

[1] Wang Yiwu, Jiang Shouhua, Zeng Manjiang. Analysis of student behavior model based on big data [J]. Changjiang Information Communication, 2024(003):37-38.
[2] Huang Meiting. Research on data clustering statistics of college students' consumption behavior based on data mining [J]. Information and Computer, 2023, 35(4):4-6.
[3] Dou Zuqi, Lu Bing, Liu Jinshuo. On the involution psychology of college students' consumption

ISSN 2616-5775 Vol. 7, Issue 7: 98-104, DOI: 10.25236/AJCIS.2024.070713

from the perspective of social comparison theory-Research and analysis on the involution behavior of contemporary college students [J]. Psychological Progress, 2024, 14(5):9-10.
[4] Cai Zhiyan. A summary of college students' consumption view and behavior [J]. Shopping Mall Modernization, 2022(8):3-4.
[5] Lin Dan, Chen Xiaoyue, Chen Tianyi. Investigation on the consumption behavior of college students in Suzhou [J]. Shopping Mall Modernization, 2023(7):17-19.