

Analysis on E-Commerce Data Based on Machine Learning

Wei Zhang^{1*}

¹ Chewy, Inc, 343 Congress St, Boston, USA

Email Address

zweivt@hotmail.com (Wei Zhang)

*Correspondence: zweivt@hotmail.com

Received: 22 December 2023; **Accepted:** 27 February 2024; **Published:** 26 March 2024

Abstract:

As one of the most important commercial activities in the digital era, e-commerce has become an indispensable part of People's Daily life with the continuous development and popularization of Internet technology. With the continuous expansion and development of the e-commerce industry, the business activities and customer interaction data faced by enterprises show an explosive growth. How to find the business value from these data has become a major problem faced by the current e-commerce enterprises. Machine learning, as a method capable of capturing data rules, has been widely used in the field of e-commerce. In order to make better use of and analyze the business data of enterprises, it is necessary to use machine learning algorithms to dig out the hidden rules and patterns from the data. E-commerce data analysis based on machine learning can predict user behavior, realize personalized marketing, improve user loyalty, and more importantly, it can provide an important reference basis for enterprise development. This paper will discuss the application of machine learning in e-commerce data analysis, aiming to analyze the opportunities and challenges brought by machine learning to the development and innovation of e-commerce enterprises, and put forward corresponding strategies.

Keywords:

Machine Learning, E-Commerce, Data Analysis

1. Introduction

With the rapid development of Internet technology, the e-commerce industry is booming and gradually becomes the mainstream business model. E-commerce enterprises provide goods and services through online platforms to bringing more convenient and diversified shopping experiences to consumers. At the same time, with the continuous accumulation of a large number of user behavior data, e-commerce enterprises are also facing the challenge of how to use these data to bring business value. In this context, machine learning, as a method to automatically learn and extract knowledge from data, has become a popular research direction in e-commerce data analysis. Machine learning algorithms can analyze a large amount of e-commerce data to identify users' behavior patterns, preferences and needs, and then predict users' purchase intentions, achieve personalized recommendations and improve user satisfaction. At the same time, machine learning can also automatically identify

suspicious fraud, and take corresponding measures to prevent and combat it. Therefore, it is of great significance for e-commerce enterprises to deeply analyze the methods and technologies, explore its advantages and limitations, and put forward the corresponding improvement and innovation.

2. Machine Learning, E-Commerce Data Analysis Advantages

2.1. Predicts user behavior

E-commerce enterprises have accumulated a large amount of user data, including users' purchase records, browsing history, search queries, social media behavior, etc. Machine learning can use this data to reveal the patterns and patterns behind user behavior. Machine learning algorithms can dig out the complex relationships between user behavior. Through data mining and pattern recognition technology, users' purchase habits, interest preferences, product relevance and other information can be found. In this way, e-commerce enterprises can predict the future behavior of users based on these relationships, such as purchase intention, attrition rate, etc. By predicting user behavior, e-commerce companies can provide personalized product recommendations. According to the historical behavior and interests of the user, machine learning can match the most suitable products for the user, thus improving the probability and satisfaction of the user purchase [1]. Personalized recommendation can enhance users' stickiness to e-commerce platforms and promote user growth and repeat purchases. Predict user behavior can also help e-commerce enterprises to segment customers and develop effective marketing strategies. By dividing users into different groups, customized marketing plans can be formulated according to the behavior patterns and needs of each group. Machine learning can identify potential high-value users and loss-risk users, thus optimizing marketing input and resource allocation. By predicting user behavior in real time, e-commerce enterprises can provide real-time personalized services. For example, when a user views a product, machine learning can analyze the user's behavior and preferences, and recommend related products or offer coupons. This real-time personalized service can enhance users' purchase intention and satisfaction.

2.2. Fraud detection

Another advantage of machine learning in e-commerce data analysis is fraud detection. E-commerce platforms often face various kinds of fraud, such as false orders, credit card fraud, account theft and so on. Machine learning can analyze transaction data and user behavior patterns, automatically identify possible fraud, and take corresponding measures to prevent and combat it. Machine learning can analyze large-scale transaction data and user behavior data. By mining and comparing the characteristics of normal and fraudulent transactions, machine learning can learn the patterns and rules of fraudulent behavior. In this way, in unknown transactions, machine learning can determine whether there is a risk of fraud based on existing experience. Machine learning algorithms can identify possible fraud. By learning the difference between normal transactions and fraudulent transactions, machine learning can build models to determine whether new transactions are at risk of fraud. Common machine learning algorithms include decision tree, logistic regression, random forest, support vector machine, etc. Machine learning can perform fraud detection in real-time. When a new transaction occurs, machine learning can quickly analyze the transaction data, determine whether there is a risk of fraud, and take corresponding

actions, such as blocking transactions, prompting users, alarm, etc. This real-time fraud detection can effectively prevent and combat fraud. In addition to real-time monitoring, machine learning can also predict possible future fraud behavior [2]. By analyzing a user's purchase history, browsing behavior, and device information, machine learning can determine whether a user is at risk of fraud and take preventive measures in advance.

2.3. Inventory control

Machine learning can help e-commerce companies optimize their inventory management and replenishment strategies by analyzing historical sales data and market trends to predict product demand and sales volume. Avoid the inventory backlog and stock shortage phenomenon, improve the capital utilization rate and customer satisfaction. Machine learning can optimize product pricing strategies by analyzing factors such as market competition, product demand and user behavior. Prices can be adjusted on real-time data, and users can be attracted through personalized price offers and promotions, increasing sales and profits. Machine learning can provide more accurate and personalized search results and product recommendations by analyzing user search query and click behavior, and improve the efficiency and experience of user search and purchase. Related products and services can be recommended according to users' interests and preferences to increase users' purchase intention and conversion rate.

3. Machine Learning and E-Commerce Data Analysis Dilemma

The accuracy and effects of machine learning are influenced by the data quality. If there are problems with missing data, errors or outliers, it will affect the training and prediction results of machine learning models. In e-commerce, data quality problems may come from incorrect user input, incorrect system records, or incomplete data integration, etc. Therefore, data quality assurance and data cleaning are the key premises. E-commerce companies need to process a large amount of users' personal data, such as purchase records, browsing history, payment information, etc., which involves users' privacy and security issues. It is crucial to process and protect user data reasonably, but it also limits the availability of data, which may bring some challenges to the training and application of machine learning models. In the e-commerce data analysis, there is often a category imbalance. For example, in fraud detection problems, the number of normal transactions is far greater than that of fraudulent transactions, so the trained model may be less accurate in identifying fraudulent transactions. Solving the problem of training sample imbalance requires some special methods, such as resampling, undersampling, or oversampling. Machine learning models are often treated as black boxes, and their decision-making process is difficult to interpret and understand. In e-commerce, interpretability and transparency are essential for trust building and user satisfaction. Therefore, when applying machine learning models, we need to weigh the relationship between the model performance and interpretability and take appropriate measures to increase the interpretability of the model. The generalization ability of machine learning models refers to the ability to adapt on new data. In e-commerce, market trends, user behavior and so on may change, therefore, the model needs to have a good generalization ability, to be able to adapt to the new scenarios and data distribution.

4. Optimize the Machine Learning E-Commerce Data Strategy

4.1. Feature engineering

Feature engineering is a crucial step in machine learning. Model performance and predictive capabilities can be improved by mining and constructing suitable features. In e-commerce data, effective features, such as user age, gender, purchase frequency, commodity category, etc., can be extracted by combining domain knowledge and user behavior characteristics. According to the complexity of the problem and the characteristics of the data, select the appropriate machine learning model [3]. Common models include decision trees, random forests, logistic regression, deep learning, and so on. In the selection process of models, factors such as model performance, complexity and interpretability are considered. Moreover, parameter tuning and cross-validation were performed to improve the generalization ability and prediction accuracy.

4.2. Data balancing and sampling

When dealing with the problem of category imbalance, the number of samples in one class of data is much more than that of the other category, which will have a certain impact on the performance and effect of the machine learning model, because the model is biased towards the majority class and cannot learn the characteristics of the minority class well. Therefore, some technical means are needed to balance the samples and improve the ability of the model to identify a few categories. Oversampling refers to increasing the number of samples for minority classes, so that the number of samples for majority and minority classes is equal or approximate. Oversampling can be done by replication samples, interpolation, etc. The advantage of oversampling is that the existing data is not used with new data; the disadvantage is that it may cause overfitting and reduce the generalization ability of the model. Undersampling refers to reduce the number of samples in the majority class, so as to be equal or close to the two classes. The advantage of undersampling is simplicity and no additional calculation and storage space; the disadvantage is that a lot of useful information may be lost, resulting to underfitting. Synthetic samples refer to increase the number of minority classes by synthesenew minority samples.

4.3. Continuous improvement and monitoring

Machine learning models and data analysis are an iterative process. Continuously monitor the model performance, timely adjust and iterate the model to adapt to new data and scene changes. At the same time, keep the attention to business needs and user feedback, timely adjust strategies and algorithms, to continuously optimize the effect of e-commerce data analysis. With the emergence of new products, new business processes, new users, etc., the model needs to be adjusted and improved in time to ensure that the model can cope with new data and scenario changes. Continuous monitoring of model performance is important because the original model may lose utility as the data updates and changes. Therefore, new data need to be collected and retrain the model to maintain the accuracy and generalization ability of the model. For the indicators of monitoring model performance, it can be selected according to the specific business requirements. On the basis of continuously monitoring the model performance, the model needs to be updated and iterated. This may involve adjusting the model parameters, changing the model structure, adjusting the feature engineering, etc. If necessary, you may also need to retrain the model or

add new data. Through continuous improvement and iterations, the model can be better adapted to new data and scene changes. In e-commerce data analysis, it is necessary to pay close attention to business needs and user feedback. For changes in business requirements, we need to follow up and add new features and data [4] to the model. At the same time, for user feedback, we need to timely solve problems and improve algorithms to improve user satisfaction. In short, continuous improvement and monitoring are an indispensable part of machine learning and e-commerce data analysis. By continuously updating and iterating models, paying close attention to business requirements and user feedback, and adopting new algorithms and technologies, we can help us build high-performance, high-reliability machine learning models, thus improving the effect and value of e-commerce data analysis.

4.4. Privacy and security protection

In machine learning e-commerce data analysis, privacy and security protection is very important, because it involves the processing of users' personal information and sensitive data. Applicable regulations and privacy policies must be strictly followed. This may include regional data protection regulations, industry norms, or privacy policies established within the business. Ensuring legal and compliant data processing is the basis for protecting user privacy. Data encryption is an important means to protect users' data security. Appropriate encryption algorithms and technologies are used to encrypt the sensitive data to ensure that the data is not easy to be stolen or tampered with during the transmission and storage process. Encryption can be applied to data transfer channels (such as the HTTPS protocol) and to data storage media (such as database encryption). For some sensitive information, such as the user's ID card number, mobile phone number, etc., you can use data desensitization technology. Data desensitization can be protected by removing sensitive information or replacing it, generalization, etc. In data processing and sharing, the use of desensitized data can reduce the risk of data leakage. In the process of data processing, a strict access permission control mechanism needs to be established. Only authorized personnel can access and process user data. This can be achieved through authentication, rights management, access audit, and other measures.

5. Conclusions

With the continuous development of machine learning technology and the continuous growth of the e-commerce industry, the e-commerce data analysis based on machine learning has achieved remarkable results in practice. Through the analysis of large-scale transaction data and user behavior patterns, machine learning algorithms can help e-commerce enterprises predict user behavior, provide personalized recommendation, realize fraud detection, etc., providing important support for the development and innovation of enterprises. Machine learning-based e-commerce data analysis is a frontier and challenging research area. By deeply tapping and utilizing the potential of e-commerce data, it can provide enterprises with more accurate and personalized services, improve user satisfaction and loyalty, and promote the vigorous development of e-commerce industry.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Zhao, J.M.; Zhang, L.P.; Liu, D.; Ren, Y.F. Machine learning-based power data regression analysis and prediction technology research. *Journal of Test Technology*, 2022, 36, 6, 525-529.
- [2] Ma, Q. Research on intelligent analysis technology of visitor review data based on machine learning methods. *Electronic Design Engineering*, 2021, 29(12), 5-6.
- [3] Ni, J.M.; Zhang, Y.; Liu, P, et al. Real-time cloud data correlation rule extraction and operation and maintenance analysis method based on machine learning. *Electronic Design Engineering*, 2022, 30(4), 5-6.
- [4] Tang, Z.F.; Zhou, Y.J. Prediction analysis of key performance indicators of a distributed database based on machine learning algorithms. *Communication World*, 2022, 002, 029-030.



© 2024 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)