International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 3, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v3n1.24
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on Product Demand Forecasting and Personalized Recommendation Based on Data Analysis

Yifu Yuan

International Joint Audit Institute, Nanjing Audit University, Nanjing, 210029, China
3189636416@qq.com

ABSTRACT

This paper aims to explore the design and implementation of product demand forecasting and personalized recommendation system based on data analysis. Through the collection and analysis of a large number of user behavior data, an accurate demand prediction model is constructed, and based on this, a personalized recommendation algorithm is implemented. The paper first introduces the research background and purpose, and then elaborates the research method, process and results. The method based on data analysis can effectively predict the product demand and provide users with personalized recommendation services, so as to improve user satisfaction and market competitiveness of enterprises.

KEYWORDS

Data analysis; Product demand forecasting; Personalized recommendation; User behavior analysis; Machine learning

1. INTRODUCTION

In the digital age, user needs are increasingly diversified and personalized. In order to meet the needs of users, enterprises need to accurately predict product demand and provide personalized recommendation services. With the development of Internet technology, platforms such as e-commerce and social media have accumulated a large amount of user data. These data contain rich user behavior information and consumption preferences, which provides huge business value for enterprises. By analyzing this data, companies can predict product demand and provide personalized recommendation services to users, thereby increasing user satisfaction and loyalty. This study aims to use data analysis technology to build accurate product demand prediction model and realize efficient personalized recommendation algorithm to provide strong support for enterprise decision-making [1]. Firstly, by collecting user behavior data and product attribute data, the data is preprocessed and features are extracted. Then, the machine learning algorithm is used to build a product demand prediction model, and the accuracy of the model is verified by experiments. Finally, personalized recommendation algorithm is implemented based on prediction model to provide personalized product recommendation for users [2].

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

2. PRODUCT DEMAND FORECASTING THEORY BASED ON DATA ANALYSIS

2.1. Basis of Data Analysis

Data analysis is a process of mining, sorting and analyzing a large amount of data to reveal the laws, trends and correlations behind the data [3]. In product demand forecasting, data analysis is the basis of accurate forecasting, which involves data preprocessing, feature extraction, pattern recognition and so on. Mastering basic data analysis skills and theoretical knowledge is essential for improving prediction accuracy and effectiveness. Data collection is the starting point of product data analysis, and its quality and completeness directly affect the accuracy and validity of subsequent analysis results. Data comes from a wide range of sources, including market research, user feedback, sales data, social media, and more. During the collection process, it is necessary to ensure the accuracy and reliability of the data, and to carry out the necessary cleaning and collation for subsequent analysis [4]. The analysis of user characteristics is the key step to understand the target user group. Through the analysis of user data, the user's age, gender, region, consumption habits, interests and preferences can be revealed. These characteristics help enterprises to better understand the needs of users and formulate targeted product and market strategies. Data analysis tools are important auxiliary means for data analysis and mining. Choosing the right data analysis tool can greatly improve the efficiency and accuracy of the analysis. Common data analysis tools include Excel, Python, R, etc., which have powerful data processing, visualization and modeling capabilities, which can help enterprises better analyze product data. Product life cycle assessment is a comprehensive analysis of the entire process of a product from market to decline. Through the analysis of sales data, user feedback and other data, it is possible to evaluate the market performance of the product, user satisfaction, and potential risks and opportunities. This helps enterprises to formulate reasonable product strategies and extend product life cycle [5].

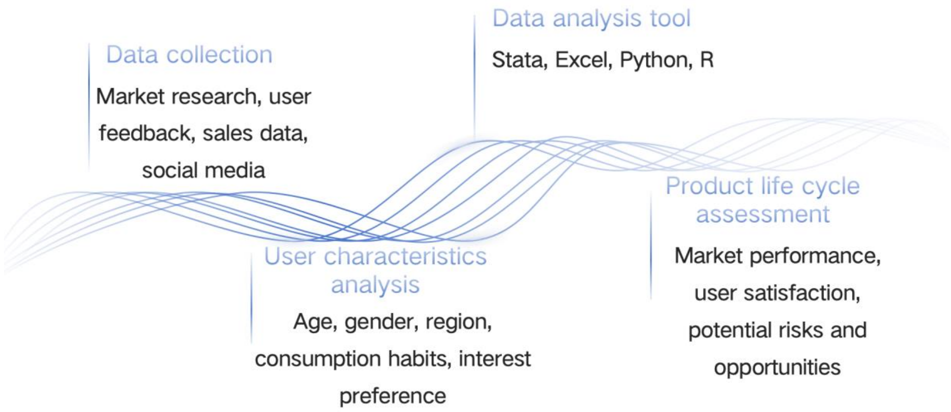

Figure 1. Fundamentals and applications of data analysis

2.2. Market Demand Analysis

Market demand analysis is the core of product demand prediction. Through in-depth analysis of historical sales data and user behavior data, the changing trend and potential law of market demand can be revealed. In addition, it is also necessary to obtain more comprehensive market demand information by means of market research and user interviews to provide strong support for the forecasting model. The study of competitive situation is an important means to understand the performance and development trend of competitors in the market. By analyzing competitors' products, prices and marketing strategies, we can evaluate our own competitive position and market share. In product demand forecasting, the study of competitive situation helps to identify potential market opportunities and threats, and provides important reference for forecasting models.

3. THE SIGNIFICANCE OF DATA ANALYSIS FOR PRODUCT DEMAND FORECASTING

3.1. Reveal Market Trends

Data analysis plays an important role in product demand forecasting, which can help enterprises better grasp market opportunities, improve operational efficiency and market competitiveness. Through in-depth analysis of historical sales data, market research data and user behavior data, market trends such as seasonal fluctuations and changes in consumer preferences can be revealed. This helps enterprises to accurately judge the future market trends and provide strong support for the forecast of product demand.

3.2. Improve Prediction Accuracy

Data analytics can help companies build accurate predictive models using a variety of statistical and machine learning algorithms. These models can take into account multiple influencing factors, such as macroeconomic environment, competitor dynamics, consumer purchasing power, etc., in order to more accurately predict product demand. By accurately predicting product demand, enterprises can better meet market demand and enhance customer satisfaction and loyalty. At the same time, enterprises can also adjust product design and functions according to the forecast results to better meet the expectations and needs of consumers, thereby enhancing market competitiveness

3.3. Optimize Inventory Management

Accurate product demand forecast is helpful for enterprises to make a reasonable inventory plan and avoid the occurrence of inventory overhang or out of stock. This can not only reduce inventory costs, but also improve customer satisfaction and enhance the competitiveness of enterprises. In addition, data analysis can provide a powerful basis for enterprise decision making. By forecasting product demand, enterprises can make production plans, market strategies and resource allocation schemes more wisely, thus improving operational efficiency and market response speed.

4. PRODUCT DEMAND FORECASTING METHOD BASED ON DATA ANALYSIS

4.1. Data Analysis and Classification

Data analysis is a technology that seeks hidden information and knowledge from a large number of practical data stored in the database, excavates potential value information, and applies these value information to business scenarios and generates business value. This technology can calculate or deduce the business results hidden in the data more accurately and shrewdly than the human brain, helping operators to make scientific judgments and develop smart solutions. With the development of information technology, the penetration field of data mining has been very extensive, the use of data mining for analysis and prediction, there are mainly the following commonly used methods:

4.1.1. Classification and regression of data

Classification problem is to divide the database into different classes according to the common characteristics of the data through the classification model: regression is a function of mining the characteristics of the data in time to map the dependency between variables. The predictive model constructs a model by extracting information such as attributes and labels from the data information, so as to predict the unknown data. The difference between classification and regression problem is that if the target predicted value is discrete in data, it should be analyzed as a classification problem. If the data representation of the target predicted value is continuous and the relationship between the variable and the dependent variable can be expressed by regression formula, it should be analyzed as a regression problem.

4.1.2. Cluster analysis

Clustering is considered to be an active method of grouping data into many sets or clusters based on the similarity of features and features of data points. Over the past few years, researchers have proposed and implemented dozens of data clustering techniques to solve data clustering problems, such as K-Means, DBSCAN, and more. In general, cluster analysis techniques can be divided into two broad categories: hierarchical analysis and region-based analysis. The more common applications of clustering problems are segmentation of images, segmentation of customers, analysis of social networks, and so on.

4.1.3. Association Rules

Association rule mining is a rule-based machine learning algorithm that can discover relationships of interest in large databases. Its purpose is to use some metrics to identify the strong rules in the database. In other words, association rule mining is used for knowledge discovery, not prediction, so it is an unsupervised machine learning method. Association rule mining algorithm is not only used in shopping basket analysis, but also widely used in web browsing preference mining, intrusion detection, continuous production and bioinformatics.

4.1.4. Time series analysis

Time series analysis is to predict the future value by observing historical data, including the long-term change trend, seasonal change law, cyclical change law, and forecast the development and change of the future moment. It should be emphasized here that time series analysis is not about the regression of time, it is mainly about the study of its own changes.

Table 1. The components of time series [6]

	Time series	Concept
1	Long-term trend (T)	The general tendency of a phenomenon to change over a longer period of time as a result of some fundamental factor
2	Seasonal variation (S)	A regular cyclical change in a phenomenon over the course of a year as the seasons change
3	Cyclic trend (C)	The regular change in the pattern of waves and fluctuations of a phenomenon over a period of several years
4	Irregular change (I)	Irregular change is a kind of irregular change to follow, including strict random change and irregular sudden change with great impact two types.

4.2. Demand Forecasting Algorithm

4.2.1. Long short-term memory (LSTM)

Long short-term memory (LSTM) is a kind of neural network suitable for sequence prediction with end-to-end modeling capability, can introduce other covariables and automatically extract features. Unlike traditional recurrent neural networks (RNNS), LSTMS employ a special memory unit to store and access previous information, and use a gating mechanism to control the input and output of information. These features make LSTM excellent at dealing with long sequences and short-term memory problems. In recent years, LSTM has been widely used in time series prediction, natural language processing, image recognition and other fields. The LSTM model has been proved by Ogunmolu et al to be effective in processing complex nonlinear change data. The RNN structure is shown in Figure 2, where x is the input at the current time, h is the output at the current time, and A is the network neuron. Compared to RNN, the variant structure of LSTM makes it more suitable for processing long series data and has stronger modeling capabilities [7].

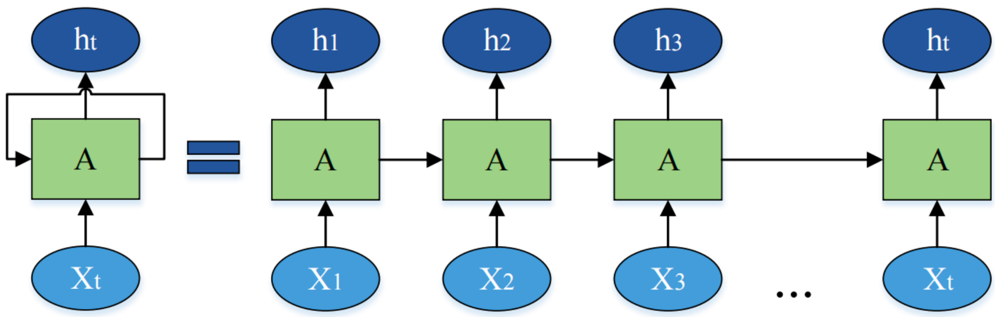

Figure 2. RNN Network Structure Diagram

LSTM is an advanced recurrent neural network, which uses three independent logic control units, namely, forgetting gate, output gate and input gate, to realize the deletion of unimportant information and the control of the input and output of new information, thus greatly improving the reliability and memory function of the system, and effectively solving the problem of gradient disappearance in traditional RNN. In this way, the LSTM network is able to better handle the relationship between sequence information and maintain long-term memory in time.

The main objective of this paper is to forecast the product demand. Therefore, the multi-input single-output mode of LSTM is selected for the network structure. In this mode, the input data contains the feature matrix X of the current sample, as well as the information retained by LSTM neuron h0 at the previous time. The output Y is the predicted value of the demand at the next time. The following diagram shows the network structure:

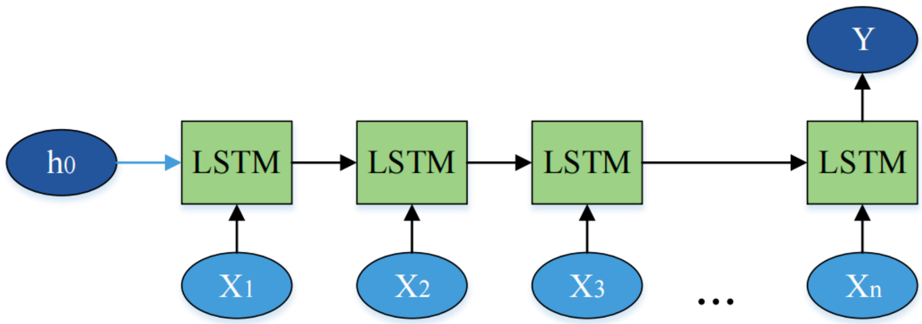

Figure 3. LSTM Network Structure Diagram with Multiple Inputs and Single Output [7]

4.2.2. Random Forest algorithm

Ensemble learning is a technique that combines several weak learners to form a strong learner. By combining weak learners, ensemble learning techniques can be divided into two types: Bagging and Boosting. Bagging is parallel training between weak learners, while Boosting is serial training between weak learners.

Random forest is a kind of bagging algorithm, which has been widely used in various industries. The principle is as follows: several subsets and features are extracted by Bootstrap method, then a multi-base model is trained, and the final result is obtained by arithmetic averaging method. The advantage of random forest is that it can reduce the variance and generalization error of the model, and can effectively deal with high-dimensional and large-data problems. In addition, the interpretability of random forest is also very good, and the importance of each feature can be evaluated, which provides a basis for the interpretation of the model. The core model of random forest is decision tree. Decision tree is a kind of tree structure based on information entropy, which generates child nodes by randomly choosing some features and splitting them each time. If the leaves are of the same class, the decision tree is no longer split, and a complete decision tree is created.

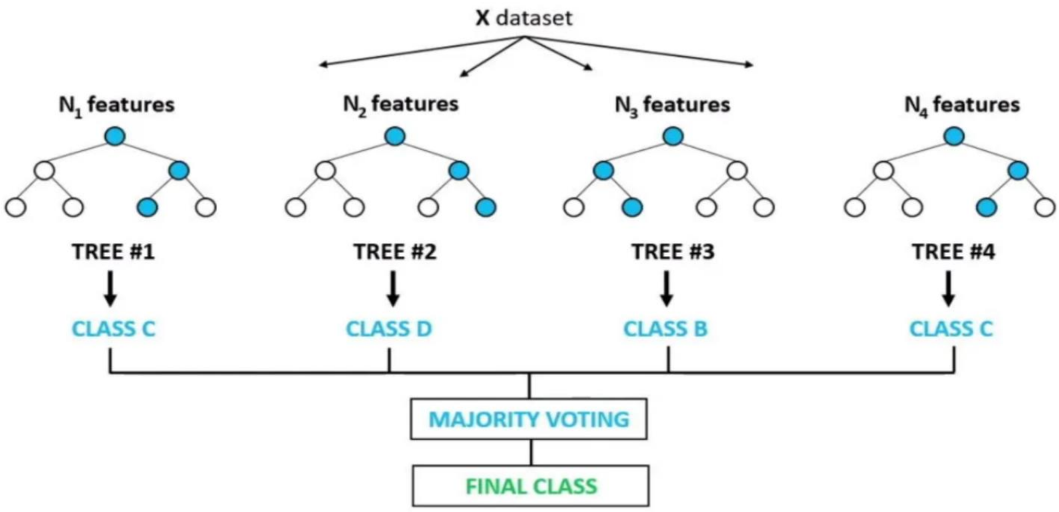

Figure 4. The process of producing Random Forest (RF) results [9]

4.3. Main Process of Data Analysis

4.3.1. Feature selection and processing

The first step in building a predictive model is to collect comprehensive and accurate market data. This data includes, but is not limited to, historical sales data, consumer behavior data, market trend analysis, competitor dynamics, etc. Data collection needs to cover multiple channels to ensure data integrity and authenticity. The collected data needs to be cleaned, collated, and analyzed to eliminate outliers, fill in missing data, and extract features useful for forecasting. By means of data visualization and descriptive statistics, the distribution and correlation of data are further explored to lay a foundation for the subsequent model construction. Select the appropriate prediction model according to the characteristics of the data and the prediction requirements. Common predictive models include time series analysis, regression analysis, machine learning algorithms (such as neural networks, support vector machines, etc.). When selecting a model, it is necessary to consider the complexity, interpretability and prediction accuracy of the model.

4.3.2. Selection and optimization of prediction algorithm

For the selected prediction model, it is necessary to set reasonable parameters according to the actual situation. This includes adjusting the hyperparameters of the model, such as the learning rate, the number of iterations, etc., as well as setting the forecast time range and prediction accuracy according to the business needs. After the model is trained, new data can be input or real-time prediction can be made, and the prediction results can be output. The forecast results are usually presented in the form of charts or reports, so that business decision makers can intuitively understand the forecast trend of product demand.

4.3.3. Model training and evaluation process

In order to verify the validity and accuracy of the prediction model, empirical analysis is needed. Common empirical analysis methods include contrastive analysis, correlation analysis, regression analysis and so on. Evaluate the performance of the predictive model by comparing it with actual sales data or other reliable data sources. Based on the results of empirical analysis, the performance of the prediction model is evaluated. The accuracy, stability and interpretability of the model are evaluated. At the same time, analyze the difference between the forecast results and the actual market demand, and find out the possible reasons and improvement directions. Finally, according to the empirical analysis results, the prediction model is optimized and adjusted. This may include adjusting model parameters, replacing models that are more suitable, introducing more influencing factors, etc.

Through continuous optimization and adjustment, the performance of the forecast model is improved, so that it is closer to the actual market demand.

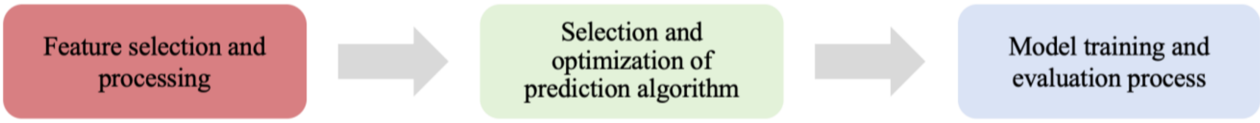

Figure 5. Process of data analysis

5. PERSONALIZED RECOMMENDATION DESIGN AND IMPLEMENTATION

5.1. Personalized Recommended Design

5.1.1. User portrait construction

The design and implementation of personalized recommendation system based on data analysis is a complex and critical process, which involves several key steps, including data collection and processing, user characteristics analysis, recommendation algorithm selection and optimization, and real-time guarantee of the system. The first is data collection and processing. The cornerstone of personalized recommendation system is a large number of user behavior data and item attribute data. This data includes the user's browsing history, purchase history, ratings, reviews, etc., as well as the item's category, label, text description, price and other information. In order to extract valuable information from this data, data cleaning, statistics, and pre-processing are required to eliminate noise and outliers, and to convert the data into a format suitable for analysis. The second is the analysis of user characteristics. The analysis of user characteristics is a key step in the design of personalized recommendation system. Through in-depth analysis of user behavior data, users' interests, preferences, consumption habits and other characteristics can be extracted. These feature information not only help to build the user's interest model, but also can be used for the subsequent recommendation algorithm optimization and model adjustment.

5.1.2. Recommendation strategy formulation

The selection of recommendation algorithm is very important to the performance and effect of personalized recommendation system. At present, the commonly used recommendation algorithms include content-based recommendation algorithm, collaborative filtering algorithm and deep learning algorithm. Each algorithm has its own characteristics and applicable scenarios, so it needs to be selected and optimized according to specific application scenarios and requirements. For example, content-based recommendation algorithms analyze users' historical behavior data and item attribute data to recommend items similar to the content they liked in the past. Collaborative filtering algorithms recommend items that may be of interest to users by analyzing user behavior records and behavior habits among similar users. Deep learning algorithms, by building complex neural network models, learn users' preferences and patterns from large amounts of data to achieve more accurate recommendations. In the process of algorithm optimization, the accuracy and performance of the recommendation algorithm can be improved by adjusting parameters, introducing new features, and using integrated learning.

5.1.3. System architecture and process design

Personalized recommendation systems need to be real-time in order to provide users with the latest recommendation results that meet their interests and needs. In order to achieve this goal, technologies such as streaming processing and online learning can be used to process user behavior data and update recommendation models in real time. In addition, the response speed and stability of the system can be improved by means of caching and load balancing. When designing personalized recommendation system, it is necessary to consider the system architecture, module division, interface design and so on. The system architecture should be scalable and maintainable to handle possible future data growth and functionality expansion. The module division should be clear and clear to facilitate collaboration and communication between modules. The interface design should be simple and easy to use, allowing users to interact with the system. In the process of implementation, we can use microservice architecture, containerization technology and other technical means to improve the flexibility and reliability of the system. At the same time, it is also necessary to pay attention to the security and privacy protection of data to ensure the security and compliance of user data [10].

5.2. Personalized Recommendation Implementation

5.2.1. Implementation details of the recommendation algorithm

In the process of implementation of recommendation algorithm, some details need to be paid attention to, first of all, user interest modeling is one of the core links of recommendation system. By analyzing the behavior data of users, the interest model of users can be built. This model can reflect the user's preferences, needs and behavior habits, and provide the basis for the subsequent feature extraction and recommendation algorithms. Secondly, on the basis of user interest modeling, feature extraction and analysis are needed. This includes extracting useful features from user behavior data and item attribute data, such as user's historical behavior, item attributes, user preferences, etc. Through the analysis of these characteristics, users and items can be more deeply understood, and more valuable information can be provided for the subsequent recommendation algorithm. Finally, similarity calculation and matching are the key steps of personalized recommendation. By calculating the similarity between users, the similarity between items, and the match between users and items, it is possible to find the items that best match users' interests. This can be achieved through various similarity calculation algorithms and matching algorithms, such as collaborative filtering, content recommendation, etc.

5.2.2. System development and deployment

System architecture and implementation is the last step of recommendation system design. When designing the system architecture, it is necessary to consider the scalability, maintainability and real-time performance of the system. In the process of implementation, modern software development techniques such as microservice architecture and containerization technology can be used to improve the flexibility and reliability of the system. After the recommendation system goes online, it needs to evaluate and optimize its effect. The performance and user satisfaction of the recommendation algorithm are evaluated by collecting user feedback and analyzing system logs. Based on the evaluation results, the recommendation algorithm and system architecture are optimized to continuously improve the recommendation quality and user experience. In the process of designing and implementing personalized recommendation systems, privacy protection and compliance requirements must be strictly observed. Through encryption technology, data desensitization and other means to protect the security and privacy of user data. At the same time, relevant laws and regulations and industry standards need to be followed to ensure the compliance and sustainability of the recommendation system.

6. CONCLUSION

This paper discusses the construction process of product demand forecasting model, including market data collection, data processing and analysis, forecasting model selection, model parameter setting, forecasting result output, empirical analysis methods, interpretation of empirical results, and model optimization and adjustment. However, in the process of empirical analysis, we also found some problems and shortcomings, such as the generalization ability of the model needs to be improved, and the prediction ability of some features is weak. To solve these problems, we put forward the corresponding optimization suggestions and improvement directions, such as further optimizing the model structure and introducing more influencing factors. In the future, we will continue to improve and optimize the product demand forecasting model to improve its forecasting accuracy and applicability, and provide more powerful support for the development of enterprises. By building and optimizing forecast models, enterprises can more accurately predict product demand and provide strong support for strategic planning and operation management. In the future, with the continuous progress of technology and the constant changes of the market, the research and application of product demand forecasting model will be further improved and developed.

REFERENCES

[1] Van Nguyen T, Zhou L, Chong A Y L, et al. Predicting customer demand for remanufactured products: A data-mining approach [J]. European Journal of Operational Research, 2020, 281(3): 543-558.

[2] Fan Lei. Research on Demand Forecasting and Personalized Recommendation of Automotive Internet Products based on Big Data Analysis [J]. Brand Research,2023(35):37-39.

[3] Tao Runrong. Research and implementation of early warning and recommendation system for subscriber loss of Telecom Package [D]. Jiangsu: Yangzhou University,2023.

[4] Fu Lijun, Pan Xuwei. Research and application of key technologies of precision marketing of agricultural products e-commerce based on User portrait [J]. Journal of Agricultural Engineering,2024,14(1):50-55.

[5] Abbasimehr, Hossein, Mostafa Shabani, and Mohsen Yousefi. "An optimized model using LSTM network for demand forecasting." Computers & industrial engineering 143 (2020): 106435.

[6] Huang Xue-bo. Research and implementation of enterprise product Information extraction, Analysis and recommendation based on NLP [D]. Shandong: Qingdao University of Technology,2018.

[7] Kadam V, Vhatkar S. Design and Develop Data Analysis and Forecasting of the Sales Using Machine Learning [M]//Intelligent Computing and Networking: Proceedings of IC-ICN 2021. Singapore: Springer Nature Singapore, 2022: 157-171.

[8] Abolghasemi M, Beh E, Tarr G, et al. Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion [J]. Computers & Industrial Engineering, 2020, 142: 106380.

[9] Genuer R, Poggi J M, Genuer R, et al. Random forests[M]. Springer International Publishing, 2020.

[10] Cui Z, Xu X, Fei X U E, et al. Personalized recommendation system based on collaborative filtering for IoT scenarios [J]. IEEE Transactions on Services Computing, 2020, 13(4): 685-695.