International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 6861-6869

https://doi.org/10.70517/ijhsa463588

Consumer demand identification method based on K-nearest neighbor algorithm and multimodal data fusion

Tianyi Yu[1,*]
[1] Zhejiang Business College, Hangzhou, Zhejiang, 310000, China
Corresponding authors: (e-mail: lynn118211@163.com).

Abstract Consumer demand identification is the core problem of modern business intelligence, and traditional methods are difficult to effectively integrate heterogeneous data from multiple sources. With the explosive growth of data scale on e-commerce platforms, how to accurately recognize consumer demand from multimodal data such as massive reviews and product information has become a key challenge. Existing demand identification methods are deficient in feature extraction and classification accuracy, and lack the ability to deeply integrate multimodal data, so there is an urgent need for more accurate and efficient technical means to realize the accurate identification of consumer demand. Purpose: Aiming at the problems of low accuracy and limited data fusion capability of traditional consumer demand identification methods, we propose a consumer demand identification method based on K nearest neighbor algorithm and multimodal data fusion. Methods: Web crawler technology is used to obtain 11,896 Jingdong Mall review data, and 10,861 valid texts are obtained after preprocessing; Jieba partitioning and TF-IDF algorithm are used to extract key features, and Particle Swarm Optimized K Nearest Neighbors (PSO-KNN) classification model is constructed to integrate multimodal data, such as semantic analysis and commodity information, for demand identification. Results: the proportion of positive consumer evaluations higher than 0.5 is 86.7%, the price of goods is mainly concentrated in the range of 10-20 yuan, and the goods with PH value distribution in the range of 3.5-4.5 are the most popular; compared with the traditional KNN and SVM algorithms, the average absolute percentage errors of PSO-KNN model are reduced by 1.87% and 3.18%, respectively. Conclusion: The proposed method effectively improves the accuracy of consumer demand identification and provides scientific support for enterprise precision marketing and product optimization decision-making.

Index Terms K-nearest neighbor algorithm, multimodal data fusion, consumer demand identification, particle swarm optimization, TF-IDF, sentiment analysis

I. Introduction

With the rapid development of science and technology, in the current wave of market economy, the increasing abundance of new things embellish people's modern life, the competitive market forces enterprises to constantly seek innovation and uniqueness [1], [2]. The source of this phenomenon lies in the fact that product innovation is the basis for the survival of manufacturing enterprises, and the production and manufacture of innovative products that meet the market demand is the magic weapon for manufacturing enterprises to maintain competitiveness in the competitive market [3]-[5]. With the emergence of e-commerce platforms, consumers can easily access a variety of products and services through online channels, and at the same time, they are also plagued by the problems of massive choices and information overload [6], [7]. Therefore, identifying consumer demand for products and services is important for understanding consumer behavior, optimizing e-commerce platforms, and developing effective corporate marketing strategies.

Since consumers tend to realize that they have certain needs or problems in their daily lives, and thus enterprises carry out demand identification is an important part of promoting consumers to execute purchase decisions [8]. For e-commerce platforms, demand identification is more diverse and personalized, and consumers may perceive the demand through social media, search engines, online advertisements, and other channels [9]-[11]. Consumers at this stage may be influenced by the external environment, such as recommendations from friends and advertisements, or they may be driven by internal factors, such as personal preferences and demand levels [12], [13]. Understanding consumers' behaviors and motivations in the demand identification stage not only helps enterprises to discover users' needs, but also helps enterprises to design more accurate marketing strategies and provide personalized products and services, thus promoting consumers to carry out purchasing behaviors and achieve a win-win situation [14]-[16].

In the context of the rapid development of Internet economy, consumer demand identification has become an important foundation for enterprises to formulate marketing strategies, optimize product design and improve service quality. The booming development of e-commerce platforms has generated a huge amount of user behavior data, which contains a wealth of demand information, and how to accurately extract and identify consumer demand from these complex and diverse data has become a common focus of attention in both academia and industry. Traditional demand identification methods mainly rely on a single data source, such as questionnaire surveys or sales data analysis, which have obvious limitations in handling large-scale and multi-dimensional data, and are difficult to comprehensively reflect the real demand characteristics of consumers. In recent years, the breakthrough of machine learning technology in the field of text analysis and pattern recognition has provided new ideas to solve this problem, especially the K-nearest neighbor algorithm has been widely used in classification tasks due to its simple and efficient features, but there is still room for improvement of the traditional K-nearest neighbor algorithm in terms of parameter selection and distance calculation. Meanwhile, the development of multimodal data fusion technology provides the possibility of understanding consumer needs more comprehensively, and a more complete consumer portrait can be constructed by integrating different types of data, such as text comments, product attributes, and user behavior. The current research needs to go further in terms of feature extraction methods, algorithm optimization strategies, and multi-source data fusion, especially in terms of how to effectively deal with the semantic information of Chinese text and improve the accuracy of classification algorithms need more exploration. In this paper, firstly, we use web crawler technology to obtain large-scale user comment data from Jingdong Mall, and use Chinese word splitting and TF-IDF algorithm for text preprocessing and feature extraction; secondly, we improve the traditional K-nearest neighbor algorithm through particle swarm optimization algorithm to improve the classification accuracy and convergence speed; lastly, we integrate the multimodal data such as comment text, product information, and user ratings to construct the demand identification model, and verify its effectiveness through experiments. Through this study, we expect to provide scientific method support for enterprises to accurately identify consumer demand and optimize product strategy, and at the same time provide new ideas for the application of multimodal data fusion in the field of business intelligence.

II. Consumer demand multimodal feature data mining

II. A. Semantic analysis

II. A. 1) Web crawling techniques

Web crawlers, also known as web spiders, are essentially information programs or scripts that automatically crawl information on the World Wide Web according to a pre-designed program [17]. The Internet produces a huge amount of information every day, and how to effectively extract and utilize this information in the face of such a huge amount of data has become a great challenge. In this information contains countless shapes and colors of data, in which most of the retrieved results are useless and without a certain structure for us. In order to solve the above problems, crawling techniques have emerged that enable targeted crawling.

II. A. 2) Chinese word separation technology

Chinese word division means to divide a complete Chinese sentence into individual words to form a word sequence. Since there is no space between Chinese words to act as a natural delimiter, a whole coherent sentence has only a few punctuation marks for differentiation, and Chinese word combinations and syntax are more complex than English, so the English lexical technology can not be well adapted to Chinese. At present, there are three main Chinese word separation methods: dictionary and thesaurus matching based on the word separation method, knowledge-based understanding of the word separation method, based on statistical word separation method.

The development of Chinese participle has already had a relatively mature methodology and operational tools, for different scenarios and application requirements, also applies to different participle methods. So there are many Chinese word separation toolkits on the market that can be used directly by ordinary users, the more common are Jieba, ICTCLAS, IK Analyzer, SCWS, HTTPCWS, CC-CEDICT and Butcher's word separation package.

II. A. 3) TF-IDF algorithm

TF-IDF is an algorithm, this algorithm is based on the theory of statistical weighting. One of the key parameters is TF frequency, which represents the number of occurrences of a keyword, the larger the value of TF frequency means the higher the importance of the word; the other is IDF frequency, which mainly represents the differentiation of the keyword word, which is different from the TF frequency, the IDF frequency considers that the lower the number of occurrences of a keyword word in the other documents, the better the differentiation, and that's the reason why it's called the reverse frequency [18]. For example, a high number of occurrences of a word in one document but a low frequency of occurrences of that word in other documents of this text collection corresponds to a highly weighted

TF-IDF value. Thus if a large TF-IDF value is guaranteed it is possible to screen out common words and keep only those that are important for text differentiation.

The main idea is that if a word has a high frequency of occurrence in a passage of text (high TF value) and a low number of occurrences in other texts (high IDF value), it is considered to have an excellent category differentiation ability and is suitable to be used as a subject keyword.

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{i,j}} \tag{1}$$

In Eq. (1): n_{ij} denotes the number of times the word w_i appears in a forum post, and the denominator is the sum of the number of times all words appear in the forum post. That is, the more times w_i appears in the text of all posts, the larger the TF value.

$$IDF_i = log\frac{N}{n+1} \tag{2}$$

In Eq. (2), N denotes the total number of forum posts involved in the data, while n denotes the number of posts containing w_i, and the denominator is increased by 1 in order to prevent the word from not being in the text base, which would result in the denominator to be 0. i.e., the fewer times that the word w_i occurs in other posts, the larger the IDF value.

$$TF\_IDF_i = TF_{i,j} \times IDF_i \tag{3}$$

The multiplication of the two equations leads to equation (3), which is the TF-IDF formula for word w_i, to represent the weight of the word.

II. B. Consumer demand capture

II. B. 1) Data acquisition

The huge review data of online shopping platform is getting more and more attention, and the research and mining of consumer evaluation has become one of the research hotspots in some specialties. This paper uses “Octopus Collector” to realize data crawling. Jingdong Mall platform is selected as the data source, the keyword “consumer demand” is searched, and the list of products is sorted according to the sales volume, and all the review data of different consumers' demand for the purchase of goods are captured. The user review data captured this time totaled 11,896 after removing the default positive feedback, blank text and duplicate text, and the exported data contains the user name, product information, evaluation star rating, evaluation time, and evaluation information.

II. B. 2) Data pre-processing

1) Data cleaning. Data cleaning refers to a series of processing steps to remove errors, noise, and redundant information from data to improve the quality and credibility of the data. Remove comment text of less than 8 characters as well as meaningless comments. The Pandas library in Python is used to realize the removal of comment text within 5 characters, system default comments, duplicate comments, and meaningless comments such as “positive”, “good”, “like”, etc.; the total number of comment texts obtained is 10861 in the end. “The total number of comments is 10861.

2) Jieba Segmentation. Chinese lexical technology can accurately separate large text content into several words, which is a key step for subsequent Chinese information processing. In this paper, we use the Jieba lexical library in Python.

3) De-duplication of words. Chinese deactivated words refer to some meaningless virtual words, prepositions and auxiliaries that have no practical significance to the research work. In this paper, the deactivation lexicon includes the common deactivation word lists such as the Chinese invalid word list and Baidu keyword list, and the deactivation words are expanded with appropriate amount of customized deactivation words according to the research features.

II. B. 3) TF-IDF word frequency statistics

Words with high TF-IDF values indicate higher representation as keywords to the text base. Python programming was applied to calculate the TF of all independent words with the formula:

$$TF_{ij} = \frac{n_{i,j}}{\sum_k n_{ik}} \tag{4}$$

where $n_{i,j}$ is the number of times an independent word i occurs in a comment j, and $\sum_k n_{ik}$ is the total number of words in the comment j.

To prevent an independent word from being overweighted in the corpus, the IDF statistics are:

$$IDF_i = \log \frac{|D|}{|\{j : t_i \in d_j\}| + 1} \tag{5}$$

$|D|$ denotes the total number of documents in the corpus, $|\{j : t_i \in d_j\}|$ denotes the number of documents containing the word t_i, and the addition of one is to prevent computational errors caused by a zero denominator.

The formula of TF-IDF algorithm is as follows:

$$TF - IDF = TF_{ij} \times IDF_i \tag{6}$$

The top 20 keywords can be calculated as shown in Table 1.

Table 1: Keyword frequency statistics

	Key words	TF-IDF	Word Count
1	Mass	0.461	1619
2	Spirit	0.424	1418
3	Substance	0.358	1062
4	Safety	0.338	954
5	Physiology	0.306	778
6	Respect	0.301	752
7	Value	0.300	746
8	Function	0.281	647
9	Environmental protection	0.263	548
10	Innovate	0.247	463
11	Aesthetic feeling	0.245	452
12	Affections	0.240	422
13	Socializing	0.227	352
14	Price	0.222	324
15	After sale	0.220	315
16	Service	0.216	291
17	Culture	0.207	246
18	Desires	0.207	245
19	Attitude	0.207	244
20	Speed	0.206	240

II. C. Characterization of consumer demand

II. C. 1) Commodity information analysis

(1) Commodity information analysis: In order to extract the characteristics of sanitary napkin commodity information on the Jingdong shopping platform and explore consumer purchase preferences, the information of the top 100 commodities sold on the platform was captured and analyzed, and the results were plotted as a more intuitive diagram as shown in Figure 1. Commodity brand positioning: After the research, it was found that there are more than 100 brands such as Nursing Care, High Knot Silk, Free Point, etc. sold online. The top 90 selling products are from 36 brands, with a large price area between brands, ranging from 100 yuan to tens of dollars.

Commodity price range: the price is mainly concentrated in 10~20 yuan. The overall price difference between the top 90 sales of commodity styles is not large, only one reached at 38 dollars. Visible, high-priced goods sales are low, showing that consumers are sensitive to price, acceptance of the average concern about price reasonableness.

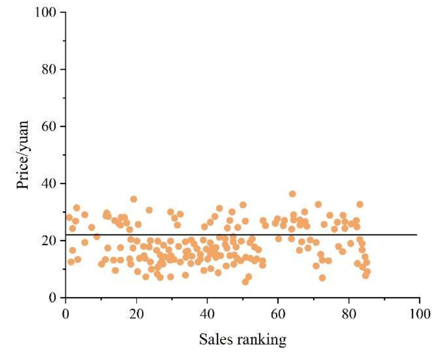

Figure 1: Commodity price range

The distribution of commodity PH values and selection preferences are shown in Figure 2. From the point of view of commodity PH value, there is a large range of choices in the extracted information of 90 commodities. Among them, the main distribution in the range of 3.5~4.5 is the most, and a few choices in the range of 5~7, which shows that consumers pay more attention to the comfort of the goods as well as the health of the body.

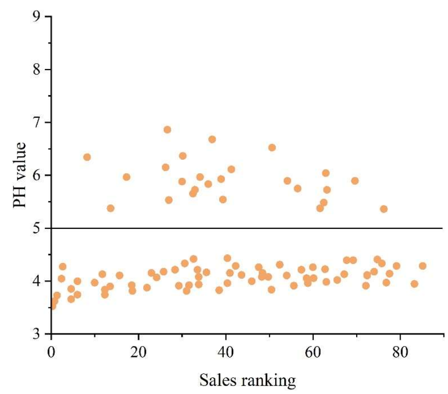

Figure 2: Consumer preferences

II. C. 2) On-line evaluation analysis

After understanding consumers' shopping preferences through product information, sentiment analysis is used, with higher sentiment scores representing higher satisfaction. In the category sales volume of the first 90 products, extract each product under the first 200 comments for sentiment analysis, to get the sentiment score shown in Figure 3. The sentiment score in the figure is centrally distributed in the range of 0.7-1.0, with comments higher than 0.5 accounting for about 86.7%. Consumers are still satisfied with the hot models.

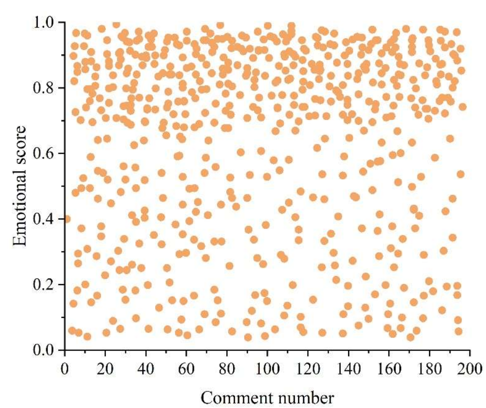

Figure 3: Consumer emotional score

III. Consumer demand identification model

III. A. Requirement identification based on particle swarm optimization kNN algorithm

III. A. 1) K Nearest Neighbor Algorithm

The k-nearest neighbor algorithm (knn) is a non-parametric estimation method in the field of pattern recognition, and the algorithm is simple, fast and efficient. The idea is: if a sample data x to be recognized, in the feature space k nearest neighbor training samples represent points of the majority of which belong to one of the categories, then x also belongs to this category. Generally k 's should not be large or too small, the time complexity of the knn algorithm is directly related to the size of the sample [19].

Suppose there are c sets of samples of categories $\omega_1, \omega_2, \ldots, \omega_c$, with $N_i (i = 1, 2, \cdots, c)$ samples labeled with the category in each class. Assuming that there are z attributes of the sample, the attribute index of the sample data can constitute a z-dimensional feature space, and all sample points have a unique point corresponding to it in this z-dimensional feature space, and any sample x to be identified can be put into this z dimensional feature space, and the k nearest neighbors of the sample x can be found by constructing a distance formula (generally using Euclidean distance). Suppose again that there are N_1 samples from the category ω_1, and so on, and N_c samples from the category ω_c. If $k_1, k_2, \cdots, k_c$ are the number of samples belonging to the category of $\omega_1, \omega_2, \cdots, \omega_c$ in the k nearest neighbors, respectively, then the discriminant function can be defined:

$$g_i(x) = k_i, j = 1, 2, \cdots, c \tag{7}$$

The category x belongs to ω_j if $g_i(x) = \max k_i$. For unknown samples, one can compare the distance between x and N samples of a known category, and just find x and the closest sample to it.

III. A. 2) Particle Swarm Optimization Algorithm

Particle Swarm Optimization (PSO) algorithm is based on the fact that group flight and aggregation behavior of birds during foraging can share information and find food quickly [20]. Assuming that a group of birds are randomly searching for food and there is only one piece of food in the area where the birds are searching for food, and none of the birds know the exact location of the food, the easiest and most effective way to find the food is to search the area around the bird that is closest to the food. The PSO algorithm is inspired by this model and is used to solve the optimization problem.

Mathematical expression of the algorithm: in a d-dimensional search space, there is a population of n particles x=$(X_1, X_2, \cdots, X_n)$ where the i th particle is denoted to be a d-dimensional vector $X_i = (X_{i1}, X_{i2}, \cdots, X_{id})^T$ denotes the position of the i th particle in the search space, which may also be a solution. The fitness value of each particle's position X_i can be calculated based on the objective function, and the velocity of the i th particle is $v_i = (v_{i1}, v_{i2} \cdots, v_{id})^T$ with an individual extremum value of $p_i = (p_{i1}, p_{i2}, \cdots, p_{id})^T$, and the global extreme of the population is $p_g = (p_{g1}, p_{g2}, \cdots p_{gd})^T$.

During each iteration, the velocity and position formulas in the particle swarm algorithm are as follows:

$$V_{id}(K+1) = \omega * V_{id}(K) + C_1 \times \phi_1 \times (P_{id} - X_{id}(K)) + C_2 \times \phi_2 \times (P_{gd} - X_{id}(K)) \tag{8}$$

$$X_{id}(K+1) = X_{id}(K) + V_{id}(K+1) \tag{9}$$

where K is the number of iterations of the particle swarm, V_{id} is the particle's moving speed, X_{id} is the d th dimensional position of the i th particle, ω is the inertia weight coefficient, c_1, c_2 is the learning factor or acceleration constant, ϕ_1, ϕ_2 are any two random numbers in the interval (0, 1), P_{id} is the individual extremum, and P_{gd} is the global extremum in particle swarm search.

III. A. 3) Particle swarm optimization based KNN algorithm

The procedure of KNN algorithm based on PSO optimization is as follows:

Step1: Initialization. Generate particles based on the data samples, the generated particles are the numbers of k training samples, in the training samples, P_{id} is the initial position of each particle, and P_{gd} is the number of the first k samples of all particles that have the smallest distance from the test sample in the training samples.

Step2: Update particles. According to the formula of updating the velocity and position of the particle swarm in the iterative process to get the new position of the particle.

Step3: Calculate the distance. Query the distance table according to the training sample represented by each particle, calculate the distance from that sample to the test sample if that sample has not been compared yet, and then update the distance table of the sample.

Step4: Update P_{id}, P_{gd}. Calculate the P_{id} of the particles, then calculate the distance from each training sample to the test sample in this particle iteration, sort the results in ascending order according to the calculated distances, where the top k samples with the smallest distances are assigned with the sample number P_{id}; Calculate the distances from the training samples to the test samples in all the particles P_{id}, and then sort the results in in ascending order, the first k samples with the smallest distance are numbered as P_{gd}; update P_{id}, P_{gd} according to the distance obtained in the above method.

Step5: Termination judgment. If the sum of the distances from each training sample to the test sample in this iteration of gbest is the same as the sum of the distances from each training sample to the sample to be categorized in the previous iteration, the algorithm can be terminated and the gbest of the knn combination can be output, or else continue to go back to STEP2 for searching.

III. B. Consumer demand identification validation

III. B. 1) PSO-KNN model validation

This subsection uses the above dataset for model validation of the constructed PSO-KNN consumer demand prediction model, using the cross-validation method, when the value of K is 23, the model accuracy is higher, and thus in this paper, K takes the value of 23. 1,100 purchases of data in June 2024 are taken as experimental samples, and the training and test sets are divided according to the ratio of 7:3, and programmed prediction is carried out by Python, and the comparison of the prediction the results are shown in Fig. 4. From the figure, it can be seen that the training set as well as the test set data and the real value data are roughly in the same direction between them, presenting a good fit and verifying the reliability of the model.

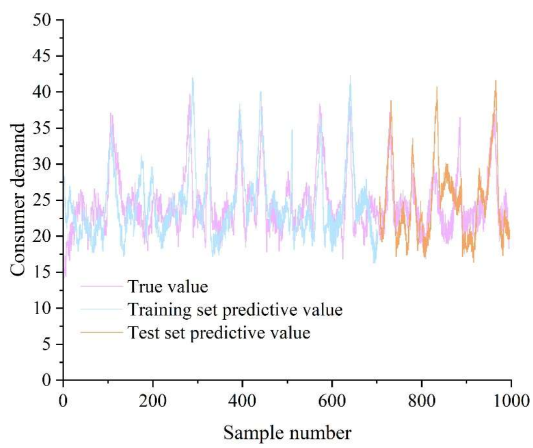

Figure 4: Comparison chart for consumer demand prediction results

Further on June 18th consumer purchase data for consumer consumer demand prediction, prediction comparison results as well as the relative error as shown in Figure 5 and Figure 6. It can be seen from the figure that the predicted consumer demand in each interval compared to the true value of the error fluctuation range is small, the error value of each consumer within 1-7. Because the purchase time of each consumer is shorter, the base is smaller, so the relative error fluctuation range is larger, the value is within 30%, which proves the effectiveness of the model's practical application.

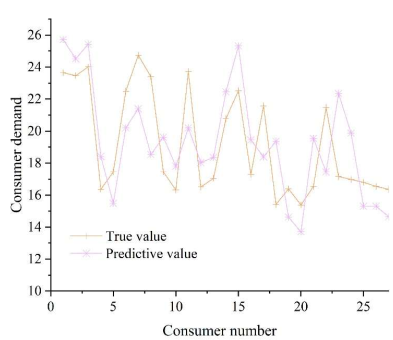

Figure 5: Comparison of consumption demand prediction results

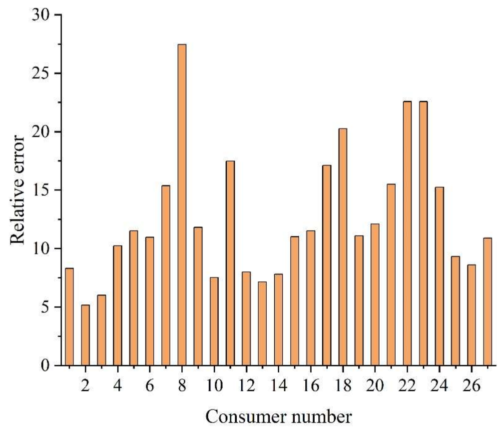

Figure 6: The relative error diagram of consumer demand prediction

III. B. 2) Comparative analysis with existing model accuracy

In order to further verify the accuracy of the model proposed in this paper, the PSO-KNN prediction model established in this paper is compared with the unimproved KNN algorithm and SVM algorithm for the average absolute percentage error. Selecting the data of the 2nd week of June, the three models are used to predict the consumer demand for the goods purchased by consumers every day of the week, and the results of the comparison of the average absolute percentage error of the models are shown in Table 2. As can be seen from the table, the average absolute percentage error of the combined model established in this paper for each day of the week is lower than that of the traditional SVM model and KNN model, which proves the high accuracy of the model.

Table 2: Comparison of model mape values

Time	SVM	KNN	PSO-KNN
June 10	19.67%	16.78%	14.91%
June 11	18.43%	17.42%	15.38%
June 12	18.39%	16.29%	15.57%
June 13	17.83%	17.41%	16.73%
June 14	18.79%	17.84%	15.49%
June 15	18.14%	17.95%	16.31%
June 16	17.54%	16.42%	15.84%

A comparison of the average absolute percentage errors of the three models in consumer demand forecasting over a one-day period is shown in Figure 7. As can be seen from the figure, the average absolute error value of the combination model represented by the blue line is basically lower than that of the SVM model represented by the yellow color and the KNN model represented by the pink color in one day time. It effectively verifies that the PSO-KNN prediction model established in this paper is more accurate and has better validity.

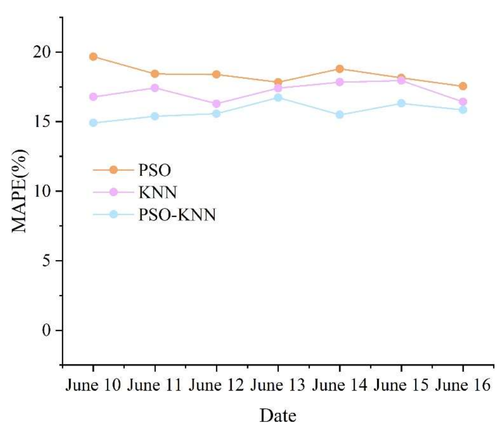

Figure 7: Comparison diagram of consumer demand prediction model

IV. Conclusion

The study has successfully constructed a consumer demand recognition model based on particle swarm optimization K nearest neighbor algorithm through the deep mining and analysis of user comment data of Jingdong Mall, which realizes the effective fusion of multimodal data. The experimental results show that the key demand features such as quality, spirituality and materiality are effectively extracted from the preprocessed 10,861 effective comment data, in which the TF-IDF value of quality demand reaches the highest value of 0.461. Consumers' overall satisfaction with the hot-selling goods is high, and the sentiment scores are centrally distributed in the range of 0.7-1.0, which reflects a good degree of market acceptance. In terms of model performance, the PSO-KNN algorithm shows superiority in the consumer demand prediction task, with a significant decrease in the average absolute percentage error compared with the traditional SVM model and the standard KNN algorithm, verifying the effectiveness of the particle swarm optimization strategy. Commodity feature analysis reveals the consumer preference law, which is mainly reflected in the favor of medium price products and the importance of product safety and comfort. The method provides a more comprehensive and accurate solution for consumer demand identification through multimodal data fusion technology, which comprehensively considers multidimensional information such as text semantics, commodity attributes and user behavior. The research result not only enriches the method system of demand identification theoretically, but also provides powerful data support for enterprises to make decisions on product design, market positioning and precise marketing, which has important practical value and promotion prospects.

References

[1] Yang, Y. (2024). Study on strategies for enhancing enterprise management decision-making ability facing market demand. International Journal of Product Development, 28(4), 301-322.

[2] Herman, H. (2018). The influence of market orientation and product innovation on the competitive advantage and its implication toward Small and Medium Enterprises (UKM) performance. International Journal of Science and Engineering Invention, 4(8), 8-21.

[3] Priem, R. L., Wenzel, M., & Koch, J. (2018). Demand-side strategy and business models: Putting value creation for consumers center stage. Long range planning, 51(1), 22-31.

[4] Kaul, A., & Luo, J. (2018). An economic case for CSR: T he comparative efficiency of for‐profit firms in meeting consumer demand for social goods. Strategic Management Journal, 39(6), 1650-1677.

[5] Arslan, I. K. (2020). The importance of creating customer loyalty in achieving sustainable competitive advantage. Eurasian Journal of Business and Management, 8(1), 11-20.

[6] Raji, M. A., Olodo, H. B., Oke, T. T., Addy, W. A., Ofodile, O. C., & Oyewole, A. T. (2024). E-commerce and consumer behavior: A review of AI-powered personalization and market trends. GSC Advanced Research and Reviews, 18(3), 066-077.

[7] Yrjölä, M., Rintamäki, T., Saarijärvi, H., & Joensuu, J. (2017). Consumer-to-consumer e-commerce: outcomes and implications. The International Review of Retail, Distribution and Consumer Research, 27(3), 300-315.

[8] Rosário, A., & Raimundo, R. (2021). Consumer marketing strategy and E-commerce in the last decade: a literature review. Journal of theoretical and applied electronic commerce research, 16(7), 3003-3024.

[9] Wu, J., Wang, Y., Shafiee, S., & Zhang, D. (2021). Discovery of associated consumer demands: Construction of a co-demanded product network with community detection. Expert Systems with Applications, 178, 115038.

[10] Arunachalam, M. P. (2024). Sentiment Analysis of Social Media Data for Product and Brand Evaluation: A Data Mining Approach Unveiling Consumer Preferences, Trends, and Insights. International Journal of Engineering and Management Research, 14(3), 46-52.

[11] Shen, Z. (2023). Mining sustainable fashion e-commerce: Social media texts and consumer behaviors. Electronic Commerce Research, 23(2), 949-971.

[12] Ansari, S., Ansari, G., Ghori, M. U., & Kazi, A. G. (2019). Impact of brand awareness and social media content marketing on consumer purchase decision. Journal of Public Value and Administrative Insight, 2(2), 5-10.

[13] Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. (2017). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. International Journal of Production Research, 55(17), 5142-5156.

[14] Khrais, L. T. (2020). Role of artificial intelligence in shaping consumer demand in E-commerce. Future Internet, 12(12), 226.

[15] Li, L., Chi, T., Hao, T., & Yu, T. (2018). Customer demand analysis of the electronic commerce supply chain using Big Data. Annals of Operations Research, 268, 113-128.

[16] Trappey, A. J., Trappey, C. V., Fan, C. Y., & Lee, I. J. (2018). Consumer driven product technology function deployment using social media and patent mining. Advanced Engineering Informatics, 36, 120-129.

[17] YuushiroUeda & KenjiroYazawa. (2024). Dragline silk fibers from golden orb‐web spider Trichonephila clavata ensure structural and mechanical robustness against ultraviolet radiation. Polymers for Advanced Technologies,35(4),

[18] Emre Delibaş. (2025). Efficient TF-IDF method for alignment-free DNA sequence similarity analysis. . Journal of molecular graphics & modelling,137,109011.

[19] Muhammad Rajabinasab,Farhad Pakdaman,Arthur Zimek & Moncef Gabbouj. (2025). Randomized PCA forest for approximate k-nearest neighbor search. Expert Systems With Applications,281,126254-126254.

[20] Xinyue Li,Yu Zhang & Wang Hu. (2025). Ensemble learning training strategy based on multi-objective particle swarm optimization and chasing method. Expert Systems With Applications,281,127777-127777.