

A Review of the Impact of Artificial Intelligence on Consumer Profiling Based on Regression Statistics

Xien Mei^{1,a,*}

¹*Cambridge International School of LITA College, SISU, Shanghai, 200083, China*
a. willow.education.us@gmail.com

**corresponding author*

Abstract: Using regression analysis and statistical methods, this study explores the impact of artificial intelligence (AI) on consumer portraits, focusing on its self-learning ability and adaptability to new environments. With the rapid development of AI technology, its transformative potential and application value in various fields of social economy continue to emerge. The results show that AI, powered by big data, can accurately characterize consumers, significantly improving the efficiency of personalized recommendation, precision marketing, and intelligent customer service. This plays a critical role in improving user satisfaction and market competitiveness of enterprises. However, the research also identifies limitations in AI's performance when the amount and quality of data are insufficient. In complex situations or value trade-offs, AI struggles to completely replace human judgment and ethical thinking ability. Overall, while AI holds great potential in consumer portraits, it is still necessary to further optimize the algorithm model, expand data sources, and strengthen the protection mechanism for data privacy and security. These improvements will lay the foundation for a wider range of cross-domain applications in the future.

Keywords: Artificial intelligence, consumer portrait, big data analysis, personalized recommendation, regression analysis

1. Introduction

The rapid development of artificial intelligence (AI) has profoundly transformed the social and economic landscape, particularly in the business sector. AI not only reshapes traditional production processes but also plays an important role in marketing and consumer behavior analysis, becoming a key enabler for accurate decision-making of enterprises. As the core tool of modern marketing, consumer portrait outlines the characteristics of consumers by analyzing data on user behavior, interests, and preferences to guide enterprises in achieving precision marketing. Traditional consumer portrait, which relies on manual experience and static data, suffers from inefficiency and limited accuracy. The application of AI greatly overcomes these defects.

Academic research on AI's impact on consumer profiling focuses on several key aspects. First, AI's advanced data processing capabilities enable the extraction of valuable insights from massive unstructured datasets, enhancing the scope and depth of consumer portraits. For example, natural language processing can analyze consumer reviews, while machine learning algorithms uncover hidden needs through user behavior analysis. Second, AI introduces real-time and dynamic features

to consumer portraits, allowing continuous updates based on user behavior, ensuring high adaptability to market changes. Furthermore, AI facilitates personalized profiling, though this raises ethical concerns regarding privacy and data security. As technology evolves, discussion around fairness and transparency have become central to academic debates.

Based on this background, this study explores AI's self-learning capabilities and its role in optimizing consumer profiling from the perspective of regression statistics. It compares AI-driven methods with traditional portrait methods, highlighting AI's potential to improve productivity and user experience. This research aims to provide theoretical insights to support the broader application of AI technologies in business contexts.

2. Theoretical Framework

The integration of AI into consumer profiling has triggered multi-disciplinary research, exploring its implications from technological, user experience, and business practice perspectives. From the algorithmic and data processing standpoint, AI demonstrates significant advantages in big data environments. Algorithms represented by deep learning provide greater accuracy and variety for consumer profiling by processing unstructured data including text and images. For example, constitutional neural network (CNN) excel in image recognition, while recurrent neural network (RNN) is good at processing time series data, facilitating the development of dynamic consumer profiles.

From a user experience perspective, AI's core value lies in personalized recommendations and situational awareness. By analyzing behavioral data such as purchase and browsing history, AI generates highly personalized recommendations and improve user satisfaction. Additionally, multi-modal data analysis combines dimensions such as text, images, and audio, enabling AI to perceive user needs and give emotional dimensions to consumer portraits.

At the business application level, AI drives advancements in dynamic pricing strategies and potential demand identification. Real-time analysis of market supply and demand enables dynamic pricing, optimizing enterprise pricing decisions. Meanwhile, clustering algorithms uncover unmet consumer needs, providing data-driven insights for product development and market expansion.

In terms of theoretical methods, linear regression model is often used to quantify the impact of AI on consumer portrait optimization. By constructing regression equations, the contributions of different technical factors to efficiency improvements and real-time adaptability are revealed. In addition, causality analysis methods delve deeper into the internal mechanism of portrait transformation driven by AI and provide rigorous logical support for the theoretical framework [1].

3. Comparison of Relative Algorithms

3.1. Linear Regression

Linear regression is a common regression algorithm that predicts the value of the dependent variable by finding a linear equation that best fits the data. The core idea of linear regression is to assume that there is a linear relationship between the dependent variable and the independent variable, so this linear relationship can be found by fitting a set of training data [2].

In implementation of linear regression usually uses the least squares method, an optimization technique that minimizes the sum of squared errors between the predicted and actual values. By reducing these squared errors, the algorithm identifies the optimal linear equation that represents the relationship between variables.

After finding the best-fitting linear equation, we can then use this equation to predict the dependent variable value of the new data point. In linear regression, the predicted value is obtained by plugging

the input data into the linear equation. Therefore, the linear regression algorithm is simple, intuitive, and easy to understand.

While linear regression can work well in many situations, it has some limitations. For example, it assumes a linear relationship between dependent and independent variables, which may not always hold. In addition, linear regression is very sensitive to outliers in the input data, which can cause the predictive performance of the model to decline [3].

In general, linear regression is a simple, intuitive, and easy-to-implement regression algorithm. It is suitable for scenarios where there is a linear relationship between dependent variables and independent variables and can achieve good results in many cases. However, in practical applications, we also need to choose the right algorithm according to the specific problem and data characteristics and evaluate and adjust accordingly.

3.2. Random Forest Regression

Random forest regression is an ensemble learning algorithm that makes predictions by building multiple decision trees and taking their average or voting results. Random forest regression combines the advantages of randomness and ensemble learning and can improve the generalization performance and robustness of the model [4].

In random forest regression, multiple decision trees are first generated, and each tree is trained on a randomly selected subset of training samples and a randomly selected subset of attributes. This way, each tree captures some different feature of the data, reducing the risk of overfitting and underfitting.

Then, for a new input sample, it makes predictions through each decision tree and takes the average of all predictions or votes as the final prediction. This approach improves the stability and accuracy of the model because each tree provides a different view of the data, reducing the error of a single decision tree.

In addition, random forest regression is also very interpretable, because each decision tree is independent and can be interpreted separately. This makes random forest regression have great advantages in explaining the model prediction results [5].

In general, random forest regression is a powerful and flexible regression algorithm that can handle high-dimensional data, handle nonlinear relations, automatically select the best features, and has good robustness and generalization performance. In practice, random forest regression is widely used in various regression problems, such as housing price forecasting, stock price forecasting, etc.

3.3. Operation Method of Data Analysis in Regression Statistics

Clarifying the problem: The first step in data analysis is to define the research objectives and core questions within a structured framework. For example, when predicting purchasing behavior through consumer profiles, it is essential to specify whether the objective is to forecast the likelihood of purchasing a specific product or to predict overall consumption trends. This requires identifying the dependent variables (target variables) and independent variables (influencing factors). Only by fully understanding the business context and actual needs of the problem can you ensure that the analysis is targeted.

Data collection: Valid data is the foundation of any analysis. Data sources can include internal user behavior records (such as shopping history, browsing data), external market research reports, and even public information on social media. When collecting data, it is also important to comply with privacy protection regulations (such as GDPR). For example, a retail enterprise may extract user purchase data through a CRM system and combine user interest labels from third-party data platforms to enrich consumer profiles [6].

Data cleaning: Since data in the real world is often noisy and defective, the data cleaning step is one of the key steps to ensure the reliability of the analysis results. For example, median or mean values can be used to fill in missing data, and outliers can be identified and eliminated by boxplot. Not only that, for duplicate records or non-uniform format data (such as postal codes in different formats in the address bar), it is also necessary to normalize processing so that subsequent analysis can proceed smoothly.

Feature selection: The purpose of feature selection is to select the most influential factor on the target variable from many variables. This can be achieved through statistical tests (such as T-tests, correlation coefficient analysis) and machine learning methods (such as recursive feature elimination). For example, when analyzing consumer profiles, variables such as age, gender, and income may affect consumption behavior, while irrelevant variables (such as randomly generated user numbers) need to be eliminated to avoid increasing the calculation burden.

Model selection: Select the most suitable regression model according to the data characteristics. For example, multiple linear regression can be used for tasks where the target variable is continuous. If the goal is a classification problem (such as buying willingness), logistic regression or support vector machines are more appropriate. In complex scenarios, hybrid models or deep learning models can also be tried to improve the flexibility and accuracy of prediction [7].

Model training: Model training is the process of letting algorithms learn from data, and it is also a key step in data processing. By dividing data sets (including training sets, verification sets, test sets, etc.), optimization algorithms (such as gradient descent method, etc.) are used to find the optimal solution of model parameters. For example, in consumer portrait modeling, the form of feature weights and bias values is iteratively adjusted to ensure that the model effectively fits the behavior pattern in the training set.

Model evaluation: Evaluating model performance ensures reliability and accuracy. Common metrics include R^2 , mean square error (MSE), and classification metrics such as accuracy, precision, and recall. Validation data or cross-data validation techniques help assess the model's stability and robustness. If the model performs better on the training data than on the test data set, overfitting may occur, requiring adjustments to model complexity or the use of regularization technology techniques [8].

Prediction and interpretation: Once trained, the model can predict outcomes for new input data. For example, a consumer profile can be used to predict the likelihood of a user purchasing a certain product [9]. Regression models also provide insights into the relationships between variables. By analyzing the parameters of the regression model, businesses can find out which variables have the greatest impact on the prediction results, offering valuable insights for decision-making.

Interpretation and application of results: The final step is to interpret the results and apply them to real-world problems. For instance, if the analysis indicates a specific consumer segment prefers a specific product, a targeted marketing strategy can be developed. Similarly, if a characteristic is found to be associated with low loyalty, incentives can be developed to increase user stickiness. In addition, the unsolved potential problems in the analysis can also provide directions for future research [10].

4. The Impact of Artificial Intelligence on Consumer Profiling

AI has revolutionized consumer profiling by leveraging advanced data analysis techniques to enhance precision, forecasting, marketing, and customer experience. The following subsections discuss the key areas where AI has had a significant impact.

4.1. Precise Positioning

AI enables businesses to accurately identify and match consumer needs and preferences through deep learning of consumer behavior data, such as purchase history and browsing patterns, using techniques like deep learning. This allows for the delivery of personalized products and services, which in turn enhances consumer satisfaction and loyalty. For example, Amazon's recommendation system employs collaborative filtering algorithms to analyze user behavior and suggest items of potential interest. This targeted approach significantly improves click-through rates, improves the shopping experience, and reduces marketing costs for merchants.

4.2. Forecasting Consumption Trends

AI's ability to process and analyze big data allows businesses to predict consumer consumption trends and future demand with high accuracy. By dynamically identifying shifts in the consumer market, businesses can maintain a competitive edge. For example, using time series analysis, AI can predict the demand for goods during specific holidays, helping businesses adjust inventory strategies to reduce overstocking or out-of-stock situations. This trend prediction can also provide scientific basis for new products to be launched and reduce the risk of market testing.

4.3. Optimizing Marketing Strategies

AI can help enterprises more accurately understand the needs and preferences of consumers, so as to develop more accurate marketing strategies. The optimization of AI-enabled marketing strategies is reflected in several aspects. For example, Baidu's advertising platform uses deep learning algorithms to achieve precise delivery, helping advertisers choose the best delivery time, audience, and content format. This kind of intelligent delivery mechanism not only improves the advertising conversion rate but also enables consumers to obtain a more personalized shopping experience and promotes the improvement of brand loyalty [11].

4.4. Improving Customer Experience

Artificial intelligence technology improves customer experience through a variety of applications (such as intelligent customer service, intelligent shopping guides, etc.). At the same time, artificial intelligence can also provide consumers with richer ways of interaction through voice recognition, image recognition, and other technologies, and enhance consumers' sense of participation and stickiness. For example, intelligent customer service can answer user questions around the clock and solve common problems several times faster than traditional customer service. In addition, AR fitting technology allows users to virtually try on clothes or cosmetics, greatly enhancing the immersion of online shopping. The improvement of this experience can not only meet the personalized needs of consumers but also promote the repurchase rate.

In short, the impact of artificial intelligence on consumer portraits is mainly reflected in precise positioning, predicting consumption trends, optimizing marketing strategies, and improving customer experience, which helps to improve the market competitiveness of enterprises and enhance the shopping experience of consumers.

5. Conclusion

Artificial intelligence has a significant impact on consumer portraits. It accurately depicts consumer characteristics through big data technology, thus realizing functions such as personalized recommendation, intelligent customer service, precision marketing, and consumer trend prediction. These capabilities significantly improve consumers' purchase willingness and satisfaction and also

enhance the market competitiveness of enterprises. However, there are still some shortcomings in the current research: First, data privacy and security issues are becoming increasingly prominent, and how to use consumer data while protecting their privacy needs to be further studied. Secondly, the accuracy and adaptability of AI models in processing diverse data still need to be improved. Finally, the existing research focuses more on the technology application itself, and the in-depth discussion on consumer psychology and behavior mechanism is relatively weak.

Future research can be carried out from the following aspects: First, expand data sources and incorporate non-traditional data (such as voice, image, social relations, etc.) into the consumer portrait system to fully reflect the characteristics of consumers. The second is to optimize the algorithm model to improve the accuracy and efficiency of artificial intelligence in data analysis. The third is to explore the cross-field application of artificial intelligence technology in consumer portraits, such as medical care, education, etc., to tap more potential value. The fourth is to strengthen the research on consumer psychology, behavior rules and the impact of technology application on society, so as to provide more scientific theoretical support for the application of artificial intelligence in consumer portraits.

References

- [1] Zhou Chunmei. *The big model craze has swept the AI market and wages have risen*. *Securities Times*. 2023.12.06 (A01): 1-2.
- [2] Hao S Y. *Consumer portrait analysis in the context of big data marketing [J]*. *National Circulation Economy*, 2019(20): 8-9. DOI:10.3969/j.issn.1009-5292.2019.20.003.
- [3] Du Jiaju, Chen Zhiwei. *Path analysis using SPSS linear regression [J]*. *Chinese Journal of Biology*, 2010, 45(2): 4-6. DOI:10.3969/j.issn.0006-3193.2010.02.002.
- [4] Li Yong, Tan Xiaoling, Chen Xiaoting, et al. *User portrait analysis of agricultural products based on e-commerce evaluation data: A case study of Anhua Dark tea [J]*. *The Rural Economy and Science and Technology*, 2019, 30 (19): 4. DOI: CNKI: SUN: NCJI. 0.2019-19-045.
- [5] liu yuanhai. *Error theory and data processing [D]*. *Dalian University of Technology*, 2009. DOI: 10.7666 / d.y. 1417708.
- [6] Yang Xiuzhi, Jiang Yuhui, Wang Xingdong, et al. *Precision Detection and Error Compensation Analysis of CNC Machine Tool Based on Linear Regression Theory [J]*. *Manufacturing Technology & Machine Tool*, 2022 (000-001): 2-6.
- [7] KuangJianHua. *Big data under the background of the informationization development of tobacco marketing research [J]*. *Journal of New Communications in China*, 2016 (4): 1. DOI: 10.3969 / j.i SSN. 1673-4866.2016.04.027.
- [8] ShuTong. *Supplier selection of supply chain coordination with the sales forecast [D]*. *Hunan University*, 2009. DOI: 10.7666 / d.y. 1260964.
- [9] Cao Zhengfeng. *Random forest algorithm optimization research [D]*. [2024-11-22], *Capital University of Economics and Business*. DOI: CNKI: CDMD: 1.1014.220587.
- [10] Cui Shangshu, Yang Lian, Li Ting. *Research on air quality prediction algorithm based on Principal component analysis and multiple linear regression [J]*. *Science and Technology Innovation*, 2022(6): 4.
- [11] Cheng Xueqi, Jin Xiaolong, Wang Yuanzhuo, et al. *Big data system and analysis technology review [J]*. *Journal of software*, 2014, 25 (9): 20. The DOI: 10.13328 / j.carol carroll nki jos. 004674.