International Journal for Housing Science and Its Applications
Publish August 4, 2025. Volume 46, Issue 3 Pages 2074-2086

https://doi.org/10.70517/ijhsa463164

A neural network-based prediction model for consumer purchase decisions and its application in marketing

Qinglan Xu[1,*], Zhucheng Xiong[2], Junyi Fang[3] and Xinyu Zhang[4]
[1] Dundee International Institute of Central South University, Changsha, Hunan, 410000, China
[2] Dundee International Institute, Central South University, 410083 Changsha, China
Corresponding authors: (e-mail: 7805220120@csu.edu.cn).

Abstract Neural networks have received increasing attention recently, which provide a relatively effective and simple method for dealing with highly complex problems. In this paper, a neural network-based prediction model for consumer purchase decision is constructed. The quantile regression function can reveal the characteristics of the entire conditional distribution of the response variable, and then using the neural network structure, the nonlinear structure in the factors affecting consumer purchasing can be simulated. The article selects the data related to the sales of notebook computers from January to June 2024 as the research object for empirical research, and the results show that, after comparing and analyzing with the linear quantile regression model prediction method, it is clear that the model of this paper investigates and predicts with a higher degree of accuracy, and with a better goodness of fit. The five quartiles of 0.1, 0.3, 0.5, 0.7 and 0.9 were selected for prediction, word of mouth and quality service, which can promote consumer purchase decision. Higher selling prices date lower consumer purchase decisions. Notebook higher memory does not significantly promote consumer purchases and should be in line with the normal needs of consumers.

Index Terms neural network, quantile regression, prediction model, distributional features

I. Introduction

In the daily operation of enterprises, both online and offline have generated a large amount of user consumption behavior data. These data bring new development opportunities for enterprises but also make enterprises face great challenges, how to identify high-quality users and channels, optimize the marketing cost has become a pain point for enterprises in various fields [1]. For enterprises, traditional marketing channels have been unable to effectively meet the personalized and diversified needs of users, and the development of the Internet and data mining technology has broadened sales channels for companies [2], [3]. Therefore, precision marketing with the analysis of consumer purchasing behavior as the core, relying on Internet technology and intelligent algorithms, is increasingly becoming the focus of attention of major enterprises [4]. In 2020, according to the research data of Puhua Industry Research Institute, more than 60% of the enterprises applied big data algorithms to enterprise marketing [5]. The role of consumer purchase behavior data to help enterprises is becoming more and more prominent, and enterprises have gradually begun to pay attention to adding big data to all aspects of marketing [6]. Therefore, how to fully mine and analyze user consumption behavior data, and make accurate predictions on consumer purchase decisions based on the analysis results has become an important issue that enterprises in various industries need to solve.

In order to improve the accurate prediction of user consumption behavior, scholars at home and abroad have done some research on user consumption behavior prediction methods. Valecha, H et al [7] studied the relationship between consumer product purchasing behavior and changes in parameters such as environmental, organizational, individual, and interpersonal factors, and used unique feature engineering to design a real-time evolutionary random forest classifier to predict consumer purchasing behavior decisions. Wang, Y et al [8] conducted purchase prediction across heterogeneous social networks and proposed a game-theoretic approach based on a stable matching model that combines account profiles and historical behaviors to identify personally linked accounts between social networking sites and online shopping, and the results showed that the approach provided more accurate predictions of consumers' purchasing behaviors in relation to goods. Dou, X [9] studied unbalanced real shopping data on e-commerce platforms and used the cat-boost model with symmetric decision tree as the base learner to analyze and predict whether a consumer will buy a certain product or not, and proved that the model has a higher accuracy rate.

With the development of machine learning methods, many scholars began to apply machine learning methods to the problem of consumer purchase decision prediction. Gallicchio, C [10] proposed that for RNN (Recurrent Neural Networks) the higher the level of the network the easier the model forgets the learned historical data, the application of hierarchical recurrent neural networks can improve the model's ability to memorize and better train the model results. Rojas, J. S et al [11] compared traditional and incremental machine learning algorithms for modeling user consumption behavior of OTT services with the goal of personalization and finally concluded that the best traditional machine learning prediction algorithm is Support Vector Machine (SVM). Ketipov, R et al [12] used machine learning models (e.g., decision trees and random forests) to accurately predict consumer behavior based on personality traits based on the link between personality dimensions and shopping behavior, which can be used to accurately predict user preferences based on their personality traits. Martínez, A et al [13] used a gradient tree augmentation algorithm to predict whether a consumer will make a purchase in the near future using data containing more than 10,000 customers and a total of 200,000 purchases, and the method achieved an AUC of 95%. Wang, W et al [14] proposed an XGBoost (eXtreme Gradient Boosting) based machine learning model for prediction of user buying behavior decisions which outperforms other machine learning models in terms of F1 scores, ROC values, and accuracy. Guo, L et al [15] researched and developed a method called SeqLearn to predict consumer payment behavior by analyzing consumer behavioral sequence data and verified the feasibility and effectiveness of the SeqLearn algorithm in experiments.

Because of the complexity of the consumer purchase decision prediction problem, a single model often produces overfitting phenomenon, so there are also a number of scholars who use a combination of models to mine and predictive analysis of purchase decision behavior data [16]. Xu, J et al [17] proposed an SE-stacking fusion model based on information fusion and integrated learning to predict consumer purchase behavior decisions in e-commerce, which outperforms traditional single machine learning models and recommender systems. Xiao, S and Tong, W [18] used a combination of text processing and clustering methods to construct consumer-product matrices and predicted consumers' future purchasing behavior decisions for products in a target category through a logistic regression classifier, and achieved 98% prediction accuracy in the Jingdong competitor dataset. XingFen, W et al [19] improved the traditional XGBoost by combining the Logistic regression algorithm with XGBoost algorithm to predict the purchasing behavior decisions of consumers on an e-commerce website, and the results showed that the fusion method outperforms the single model in all evaluations. Tian, Y et al [20] proposed a consumer behavior decision prediction model based on improved deep neural network model (rDNN), and the results showed that the AUC value of the method reached 0.8422, which solved the problem of low accuracy and unsatisfactory effect of traditional consumer behavior prediction model to some extent. In summary, the research on the problem of consumer purchase behavior decision prediction is still in the stage of continuous development, and scholars at home and abroad have carried out in-depth research on consumer purchase decision prediction by way of model construction, from the initial statistical method to the development of a single machine learning method, and then to the combined model method. However, in specific practical problems, the prediction performance of current methods is not yet very satisfactory.

This paper mainly selects the relevant data of laptop computer sales on the e-commerce platform from January to June 2024 as the target index. The data are cleaned and organized to screen out the relevant factors affecting the online sales volume of laptop computers. The acquired data are analyzed with descriptive statistics on the distribution and the relationship with sales, so as to initially understand the distribution characteristics of the various influences on film and television and the connection between them. In order to ensure that the quantile regression can more accurately ensure the accuracy of the results, this paper introduces the neural network model into the quantile regression model, according to the relevant influencing factors, at the five quantile points of 0.1, 0.3, 0.5, 0.7, 0.90, to construct the neural network based on the consumer's purchase decision prediction model. The data are brought into the decision prediction model to reveal the nonlinear relationship between the variables. And the simulation modeling effect is analyzed by comparing and analyzing with the linear quantile regression method.

II. Quantile regression model

In real life, in order to study the effect of the explanatory variables on the response variable, a linear regression model is usually employed and fitted using least squares estimation or least absolute deviation estimation. Least squares is a commonly used method of modeling mean regression, which describes the effect of the explanatory variables on the conditional mean distribution of the response variable. Least squares is an unbiased, efficient and computationally simple method, and the best linear unbiased estimates can be obtained using this method only when the error term satisfies the assumptions of homoskedasticity and zero mean. When the error terms obey normal distribution, the least squares estimation and the maximum likelihood estimation are consistent. However, in real problems, it is difficult to obtain the error terms under various assumptions [21].

II. A.Quartile regression parameter estimation methods

Suppose that the random variables $Y, \{y_1, y_2, \dots, y_n\}$, $\{y_1, y_2, \dots, y_n\}$ are random samples of the random variable Y, and that the distribution of the Y function is:

$$F(y) = P(Y \le y) \tag{1}$$

For any $0<\tau<1$, define the τ quantile of the random variable Y as:

$$Q_\tau(y) = F^{-1}(\tau) = \inf(y : F(y \ge \tau)) \tag{2}$$

Where the proportion of the region less than $Q_\tau(y)$ is τ, and $\inf(y : F(y \ge \tau))$ is the smallest element in the set of y for which the condition $F(y \ge \tau)$ holds.

The method of solving for the unknown parameter in least squares regression is to minimize the sum of squares of the residuals, unlike the least squares regression method of solving for the unknown parameter in quantile regression is to minimize the sum of the absolute values of the weighted errors. According to Koenker and Bassett, the quantile regression loss function can be expressed as:

$$\rho_\tau(u) = (\tau - I_{(u)})u \tag{3}$$

where $I_{(u)}$ is a schematic function, $I_{(u)} = 1$ when $u \le 0$, and $I_{(u)} = 0$ when $u > 0$, then the loss function $\rho_\tau(u)$ can be written as:

$$\rho_\tau(u) = \begin{cases} (\tau - 1)u & u \le 0 \\ \tau u & u > 0 \end{cases} \tag{4}$$

Clearly, the loss function $\rho_\tau(u)$ is composed of multiple linear segmented functions that take values greater than or equal to 0, and at the same time is discontinuous at the position $u=0$, and therefore not trivial at that point.

Rewrite equation (4) in the following form:

$$\rho_\tau(u) = (\tau - 1)uI_{(u \le 0)} + \tau u I_{(u>0)} \tag{5}$$

When $u = Y - \hat{Y}$ is taken, it is obtained by substituting into the above equation:

$$\rho_\tau(Y - \hat{Y}) = (\tau - 1)(Y - \hat{Y})I_{(Y-\hat{Y} \le 0)} + \tau(Y - \hat{Y})I_{(Y-\hat{Y}>0)} \tag{6}$$

In least squares regression, it is minimizing the sum of squares of the residuals, and in quantile regression, it is minimizing the sum of the absolute values of the weighted errors, i.e., finding the estimate that minimizes $E[\rho_\tau(Y-\hat{Y})]$:

$$\min E[\rho_\tau(Y - \hat{Y})] = \min\left[(\tau - 1)\int_{-\infty}^{\hat{y}} (y - \hat{y})dF(y) + \tau\int_{\hat{y}}^{+\infty} (y - \hat{y})dF(y)\right] \tag{7}$$

Taking the first order derivative of $\hat{y}$ in equation (7) and making it zero gives:

$$(1-\tau)\int_{-\infty}^{\hat{y}} dF(y) - \tau\int_{\hat{y}}^{+\infty} dF(y) = F(\hat{y}) - \tau = 0. \tag{8}$$

The set $F(y)$, as a distribution function of y, is monotonically non-decreasing and right-continuous, so the set $\{\hat{y} : F(\hat{y}) = \tau\}$ is a non-empty set, and then the estimation parameter that makes Eq. (7) hold exists. When the solution is unique, that is, there is a unique corresponding point on the distribution function, $\hat{y} = F^{-1}(\tau) = Q_\tau(y)$.

Let the random vector Y, given a matrix of explanatory variables X, the regression equation for $Y|X$ at the τ quantile can be expressed as:

$$Q_\tau(Y \mid X) = X^T\beta_\tau \tag{9}$$

where β_τ denotes the unknown parameters.

Assuming $\{(y_i, x_i), i = 1, 2, \dots, n\}$ is a random sample, where x_i is a p-dimensional column vector, establish the quantile regression equation:

$$Q_{\tau}\left(y_i \mid x_i\right) = x_i^T \beta_{\tau}, i = 1, 2, \ldots, n \tag{10}$$

The unknown parameter β_{τ} is estimated:

$$\hat{\beta}_{\tau} = \arg\min \sum_{i=1}^{n} \rho_{\tau}\left(y_i - x_i^T \beta_{\tau}\right) \tag{11}$$

After obtaining the estimate β_{τ} of the unknown parameter $\hat{\beta}_{\tau}$, i.e., the quantile regression equation is:

$$\hat{Q}_{\tau}\left(y_i \mid x_i\right) = x_i^T \hat{\beta}_{\tau}, i = 1, 2, \ldots, n \tag{12}$$

According to the formula for the estimated value of the unknown parameter β_{τ}, it can be found that the result of the coefficient β_{τ} of quantile regression depends on the choice of the quantile point τ. The value of the quantile point τ ranges from 0 to 1, so for different quantile points τ, there will be a corresponding fitting function, which in turn will give different parameter estimates. This also means that a cluster of fitted curves can be obtained for the conditional quantile of the response variable, and these fitted curves can reveal the location and shape characteristics of the distribution. Thus, quantile regression models have advantages in analyzing the distributional characteristics of the data.

II. B.Lasso regression method

The Lasso regression method is based on least squares estimation, so it needs to satisfy the assumption of normal distribution and is sensitive to outliers. So in this context, with the continuous development and application of quantile regression and regularization methods, combined with the practical needs of the problem in the quantile regression model to add L_1 penalty term, proposed the Lasso quantile regression model and gives the effective algorithm of the entire solution path, overcoming the shortcomings of the least squares regression method.Lasso quantile regression is similar to the Lasso regression, the combination of Lasso variable selection and quantile regression function, on the one hand, can realize the variable selection and outliers. Similarly, the combination of Lasso variable selection and quantile regression function can realize variable selection on the one hand and parameter estimation on the other hand [22].

Assuming that there are p explanatory variables $X_1, X_2, \ldots, X_p$ and the response variable is Y, if $\left(x_{i1}, x_{i2}, \ldots x_{ip}, y_i\right)$ is $\left(X_1, X_2, \ldots, X_p, Y\right)$ for the sample data where $i = 1, 2, \ldots, n$ and $X = \left(X_1, X_2, \ldots, X_p\right)$, $Y = \left(y_1, y_2, \ldots, y_n\right)^T$, and Lasso quantile regression can be expressed as:

$$\hat{\beta}_{\tau-Lasso} = \arg\min\left(\sum_{i=1}^{n} \rho_{\tau}\left(y_i - x_i^T \beta_{\tau}\right)\right), \; s.t \sum_{j=1}^{p}\left|\beta_{\tau,j}\right| \le t \tag{13}$$

The Lasso penalty function form can be expressed as:

$$\hat{\beta}_{\tau-Lasso} = \arg\min\left(\sum_{i=1}^{n} \rho_{\tau}\left(y_i - x_i^T \beta_{\tau}\right) + \lambda \sum_{j=1}^{p}\left|\beta_{\tau,j}\right|\right) \tag{14}$$

where $\lambda \sum_{j=1}^{p}\left|\beta_{\tau,j}\right|$ is the penalization term, and λ is the penalization coefficient $(\lambda > 0)$. the strength of the L_1 penalty term. When the value of the penalty coefficient λ is larger, it indicates that the degree of compression is more severe. As λ increases, the L_1 penalty term will cause some of the coefficients to be compressed to 0, thus realizing the effect of variable selection. Therefore, for practical problems, it is very important to choose the appropriate penalty coefficients λ. AIC, BIC criterion is usually used in least squares estimation, and the mean quantile loss function is obtained by cross-validation under quantile regression making the mean quantile loss function to be the smallest, and the expression of the mean quantile loss function is as follows:

$$\frac{1}{n} \sum_{i=1}^{n} \rho_{\tau}\left(y_i - x_i' \hat{\beta}_{\tau,\lambda}\right) \tag{15}$$

where n is the sample size and $\hat{\beta}_{\tau,\lambda}$ is the estimated coefficients under the quantile τ penalty coefficient λ.

III. Consumer purchase decision model based on neural network quantile regression

The neural network quantile regression model contains the neural network model and the quantile regression model at the same time with two major advantages: on the one hand, through the simulation of the nonlinear structure in the system, so that the neural network in the simulation of the system to obtain a more accurate simulation results at the same time, and do not need to set a specific form of the function; at the same time, using the advantages of the quantile method, by choosing different locations of the quantile point to obtain the corresponding conditional quantile method, the Thus, the state distribution of the response variable is more comprehensively and finely portrayed, so as to characterize its state distribution in an all-round way [23]. The neural network quantile modeling process is shown in Figure 1.

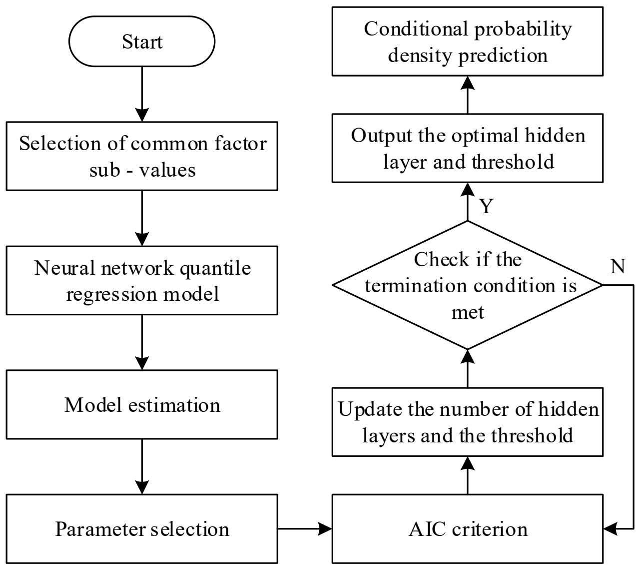

Figure 1: Neural network fractal model process

III. A. Model representation

For the generation of the Quartile Regression by Neural Networks (QRNN) model, it was formed by Taylor's combination of non-parametric regression model properties together in a neural network framework. The model framework underlying the QRNN is set up with a vector of p explanatory variables $x_1, x_2, \ldots, x_p$ as the set of explanatory variables, and the response variable is set to be y_t, and the model estimation process is embodied in the explanatory effect of $x_1, x_2, \ldots, x_p$ on y_t. Taking the underlying three-layer perceptron neural network analysis, the input layer is the explanatory variables $x_1, x_2, \ldots, x_p$, the output layer is y_t, and the node parameter of the implicit layer is set to be J number of nodes, the form of the QRNN model is organized as follows:

$$Q_t(\tau) = f^{(0)}\left(\sum_{j=1}^{J} \omega_j^{(0)}(\tau) g_j(\tau) + b^{(0)}(\tau)\right) \tag{16}$$

Interpretation of symbols in the model:

τ: the quantile of the response variable y_t.

$f^{(0)}(\cdot)$: the transformation function corresponding to the response variable y_t in the output layer.

$Q_t(\tau)$: conditional quantile of the response variable y_t at different τ.

$\omega_j^{(0)}(\tau)$: measured weights of the layer where the response variable y_t is located.

$b^{(0)}$: similar to the intercept of a function, measures the offset in the response variable y_t estimate data.

$g_j(\tau)$ is an implicit layer node in the network model. Satisfaction:

$$g_j(\tau) = f^{(h)}\left(\sum_{i=1}^{p} \omega_j^{(h)}(\tau) x_i + b_j^{(h)}(\tau)\right) \tag{17}$$

where $w_{ij}^{(h)}$ is the implied weight; $b_j^{(h)}(\tau)$ is the implied layer offset, and $f^{(h)}(\cdot)$ is the implied layer transformation function.

In the setting, the number of hidden layer nodes J must not be equal to 1, otherwise the neural network quantile regression model loses its nonlinear characteristics and transforms into a simple linear quantile regression model.

III. B. Model estimation

Combined with the implicit layer node function, this is brought into the total model function. The nonlinear relationship that shrinks between the input and output layers can be expressed by the function as:

$$Q_t(\tau) = f\left(x_1, x_2, \cdots, x_p; \theta(\tau)\right) \tag{18}$$

A representation of the latter's parameter vector in the above function, which consists of connection weights and offsets:

$$\begin{aligned} &\theta(\tau) \\ &= \left[\omega_{11}^{(h)}(\tau), \cdots, \omega_{pJ}^{(h)}(\tau), b_l^{(h)}(\tau), \cdots, b_J^{(h)}(\tau), \omega_l^{(0)}(\tau), \cdots, \omega_J^{(0)}(\tau), b^{(0)}(\tau)\right]^T \end{aligned} \tag{19}$$

Parameter estimation of the neural network quantile regression model to construct the minimization objective function:

$$\frac{1}{T}\sum_{t=1}^{T}\left[\tau - I\left(y_t < Q_t(\tau)\right)\right]\left[y_t - Q_t(\tau)\right] \tag{20}$$

For the estimation of Eq. (20) in the estimation, in order to prevent the QRNN model from overfitting during the fitting process which leads to the distortion of the model results, it is necessary to introduce a penalty function for overfitting in the objective function. The structure-building optimization function is as follows:

$$\begin{aligned} \hat{\theta}\left(\tau; \lambda, J\right) = \arg\min_{\theta} \frac{1}{T}\sum_{t=1}^{T}\left[\tau - I\left(y_t < Q_t(\tau)\right)\right]\left[y_t - Q_t(\tau)\right] \\ + \lambda \frac{l}{pJ}\left\|\omega^{(h)}(\tau)\right\|_2 \end{aligned} \tag{21}$$

where $\omega^{(h)}(\tau) \equiv \left(\omega_{11}^{(h)}(\tau), \omega_{12}^{(h)}(\tau), \cdots, \omega_{pJ}^{(h)}(\tau)\right)^T$ is the implicit layer weight vector, $\|\cdot\|_2$ is the paradigm; λ is the penalty parameter.

III. C. Guidelines for parameter selection

In the estimation of the neural network quantile regression model, the optimal values of the penalty function λ value and the number of nodes in the hidden layer J are to be found by referring to the functions described in the previous section to find the optimal point, and the AIC information criterion is introduced to make a judgment based on the need for optimal selection.

The AIC criterion is a model selection criterion that takes into account the goodness of fit of the model and the complexity of the model. Specifically, the AIC criterion is calculated based on maximum likelihood estimation and the number of model parameters [24]. The smaller the AIC value of the model, the better the model's goodness-of-fit and the lower the model complexity. When using the AIC criterion to determine the optimal penalty function value and the number of hidden layer nodes in a neural network quantile regression model, one is essentially looking for an optimal point of trade-off between the model's goodness of fit and model complexity. A larger penalty function value or a smaller number of hidden layer nodes will result in a lower complexity of the model but may reduce the model's goodness-of-fit; conversely, a smaller penalty function value or a larger number of hidden layer nodes may improve the model's goodness-of-fit but the model's complexity will also increase, and the optimal model can have the smallest AIC value. Define the AIC criterion as follows:

$$\begin{aligned} &AIC(\lambda, J) \\ &= \min\left(\ln\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\tau - I\left(y_t < Q_t(\tau; \theta(\tau, \lambda, J))\right)\right]\right.\right. \\ &\left.\left.\left[y_t - Q_t(\tau; \theta(\tau, \lambda, J))\right]\right\} + \frac{p(J+1)}{T}\right) \end{aligned} \tag{22}$$

In Eq. p is the total number of individuals corresponding to the vector of explanatory variables $x_1, x_2, \ldots, x_p$.

III. D. Model predictions

(1) Conditional quantile prediction

Based on the results of the AIC information criterion, the optimal values of the penalty parameter λ and the number of nodes in the hidden layer J in the neural network quantile regression model are screened to construct the estimation of the parameter $\theta(\tau;\lambda,J)$, which will be estimated as $\hat{\theta}(\tau;\lambda,J)$, and substituting $\hat{\theta}(\tau;\lambda,J)$ into Eq. (22), the conditional quantile prediction of the response variable Y_t, $\hat{Q}_t(\tau)$, is obtained [25].

(2) Probability density prediction

On the basis of obtaining the conditional quantile function $\hat{Q}_{Y_T}(\tau | X_t)$, the conditional density in the neural network model into the influence of the change curve, the quantile point τ of the value range of $(0,1)$, the use of asymptotic value and differential extraction in the value range of τ when the continuous value, you can get the condition of the distribution curve F, F to meet the $F\left(F^{-1}(\tau)\right)=\tau$, on both sides of the equation, respectively, the calculus of differentiation:

$$\begin{aligned} \frac{dF\left(F^{-1}(\tau)\right)}{d\tau} &= 1 \Rightarrow f\left(F^{-1}(\tau)\right)\frac{dF^{-1}(\tau)}{d\tau}=1 \\ &\Rightarrow f\left(F^{-1}(\tau)\right)=\frac{d\tau}{dF^{-1}(\tau)} \\ &\Rightarrow f(\hat{Q}(\tau))=\frac{d\tau}{d\hat{Q}(\tau)} \end{aligned} \tag{23}$$

where $F^{-1}(\tau) \triangleq \hat{Q}(\tau)$ is the τ quantile.

Conditioning the explanatory variables and discretizing the quantiles yields a simulation of the probability density of the output layer $\hat{f}\left(\hat{Q}_{Y_t}(\tau)\right)$. Conditioning on X_t and discretizing the quantile τ yields a conditional probability density simulation of the output layer $\hat{f}\left(\hat{Q}_{Y_t}(\tau | X_t)\right)$.

IV. Empirical Analysis of Consumer Purchase Decision Models

IV. A. Data sources

This paper takes the influence factors of laptop online sales as an example, in order to make the research have rich empirical significance, the acquisition of data should rely on the e-commerce platform which is widely used by consumers and is representative. Regarding the configuration attributes of electronic products and the basic information of stores, Jingdong Mall has a comprehensive system, and the sales of various products account for a large proportion of the entire market, so it is chosen to crawl the relevant data on Jingdong Mall to provide sufficient data support for the research of this paper.

The dataset is crawled on August 31, 2024, and the crawling range is the relevant data of laptop sales from January to June 2024, and the original data obtained initially is 3838 items. The obtained data is first cleaned and organized to make it usable, because the quality of the data will directly affect the effectiveness of the model.

Regarding the processing of missing values, the data with serious missing features that cannot be filled are directly deleted, and the data with less missing values are filled by using methods such as plurality and mean. Regarding the treatment of outliers, they are directly deleted or treated as missing values, and the data are organized after determining the integrity of the data. After cleaning and organizing, the final valid data obtained is 3310. The factors affecting the online sales of laptop computers to be selected in this paper are price ($x1$), brand ($x2$), screen size ($x3$), product net weight ($x4$), thickness ($x5$), core number ($x6$), processor ($x7$), theoretical endurance time ($x8$), warranty life ($x9$), commodity rating ($x10$), logistics rating ($x11$), and after-sale rating ($x12$), Commodity rating ($x13$), logistics rating ($x14$), after-sale rating ($x15$), battery capacity ($x16$), graphics card type ($x17$), number of battery cells ($x18$), screen resolution ($x19$), screen refresh rate ($x20$), computer type ($x21$) and memory capacity ($x22$).

IV. B. Descriptive statistical analysis

The actual data on laptop sales span a wide range of values, so to avoid heteroskedasticity, the sales are converted to logarithmic form and the logarithmic sales kernel density is plotted as shown in Figure 2. The figure shows that the logarithmically converted sales show a left-skewed distribution, i.e., the distribution of logarithmic sales is non-normal and concentrated between 2 and 5. In other words, when using sales volume as the dependent variable to build a model, the traditional linear regression model may not have better applicability, and instead, an analysis method that is more in line with the characteristics of the data should be chosen.

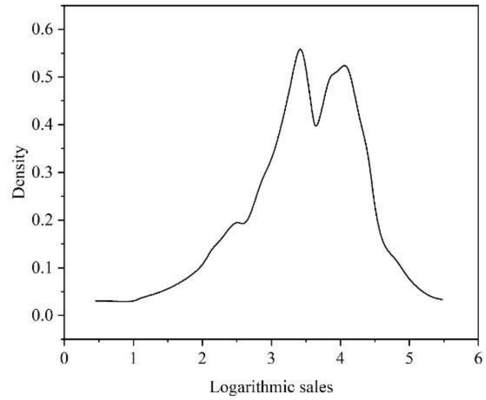

Figure 2: The nuclear density diagram of sales

The configuration level of a laptop computer reflects its excellent performance, which affects the consumer's purchasing decision and thus has an impact on the sales of the item. The laptop attribute variables selected in this paper include categorical and continuous variables, and memory capacity, number of cores, processor and graphics card type are all categorical variables, so the relationship between sales and each factor can be analyzed by drawing box-and-line diagrams, and the results are shown in Figure 3, which shows that there are different degrees of differences in memory capacity, number of cores, processor and graphics card type. The box-and-line plot for memory capacity shows that the mean value of 8GB-16GB memory capacity is relatively large and corresponds to a wide distribution of sales, and the difference in laptop sales between 4GB and 6GB compared to other subgroups is obvious.

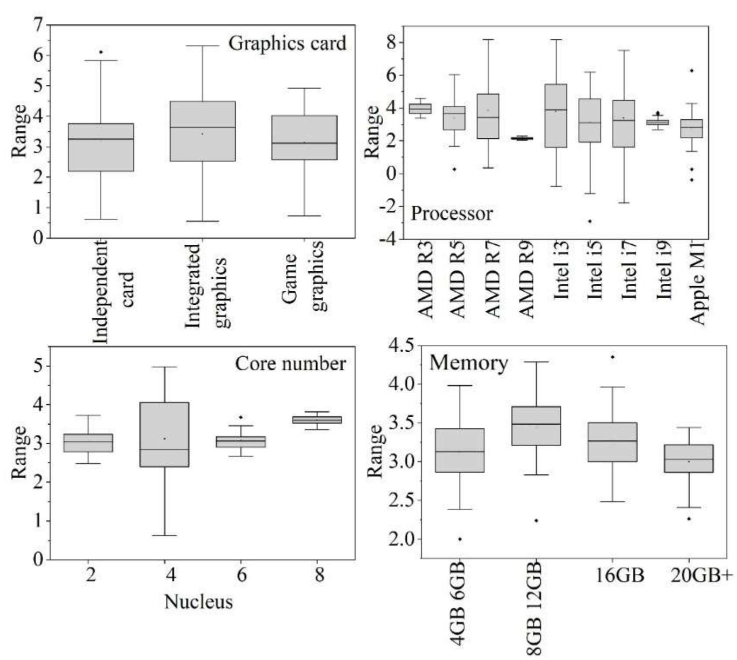

Figure 3: A box diagram of the relationship between factors and sales

In this paper, screen size and battery capacity are transformed into categorical variables to analyze the impact of these two variables on sales. According to the distribution of the variables, screen size is divided into six categories, namely, below 13.0 inches, 13.0 inches-13.9 inches, 14.0 inches-14.9 inches, 15.0 inches-15.9 inches, 16.0 inches-16.9 inches, and 17.0 inches or more. The battery capacity is divided into three categories, namely, below 30Wh, 30Wh-60Wh, and above 60Wh. The box line plots of the two variables versus sales are plotted separately, and the results are shown in Figure 4, from which it can be seen that either too large or too small a value of the screen size has a significant impact on sales, while the intermediate level has no significant impact on sales. A smaller battery capacity corresponds to fewer sales, indicating that a larger battery capacity can purchase promote sales.

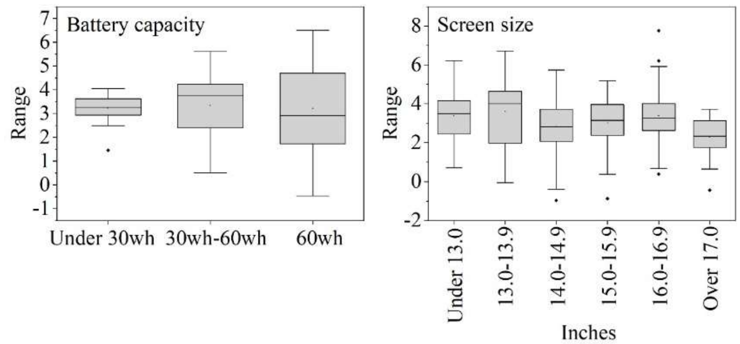

Figure 4: Screen size and battery capacity and sales relationship

The data obtained in this paper covers more than 20 laptop brands, and it is not easy to present all of them, so the brands with larger sample sizes are chosen to plot the kernel density of their prices. The results are shown in Figure 5, the prices of all laptops are roughly between 2000-10000 yuan, which is consistent with the price distribution about laptops in the market. The price distribution graphs of different brands show that the prices of each brand of laptops are mostly skewed. Among them, the prices of Dell, Acer, ASUS, Honor, MI, and HP are mainly distributed around 3,000-6,000 yuan, which indicates that these brands mainly focus on low-end and mid-range computers. And the prices of Apple, Lenovo, and Huawei are mainly distributed around 5,000-10,000 yuan, indicating that these brands sell more high-end laptops.

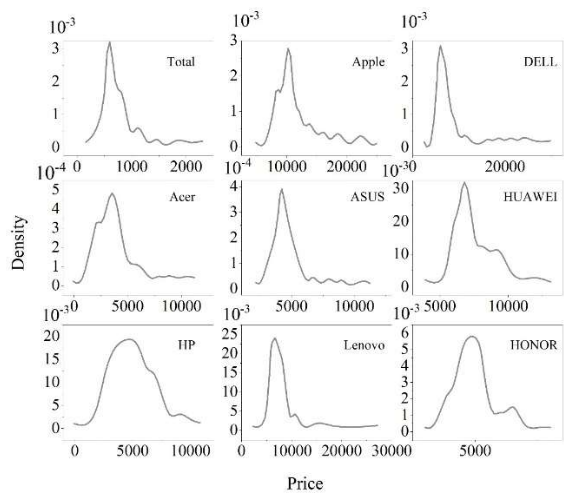

Figure 5: Different brand price distribution

In order to further explore the relationship between price and sales of different brands, the average sales of different brands are compared with the average price, and the comparison results are shown in Figure 6. Where the horizontal coordinate indicates the brand, the vertical coordinate on the left indicates the average sales volume of each brand, and the corresponding image is a bar chart, while the vertical coordinate on the right indicates the average price of each brand, and the corresponding image is a line graph. As can be seen from the graph, with 6000 yuan While Apple, Lenovo, Huawei's products are relatively high-end, their prices are higher and so are their sales.

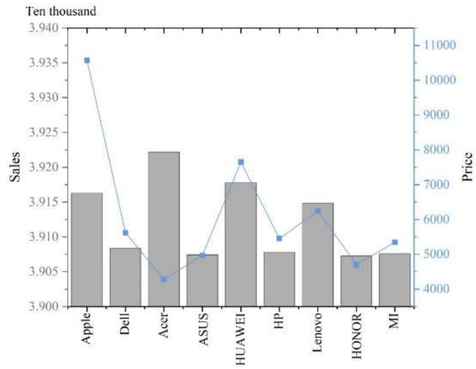

Figure 6: Average and price average diagram of different brand sales

IV. C. lasso regression screening variables

In this paper, we use the method of cross-validation and select the ten-fold cross-validation with K=10 to determine the parameters of the model. The parameter results are shown in Figure 7, where each curve represents the trajectory of the coefficients of each independent variable, the vertical coordinate is the value of the coefficients, the upper horizontal coordinate is the number of steps, and the lower horizontal coordinate characterizes the degree of sparsity, which can be seen that there is a process of compressing the coefficients of the variables to zero.

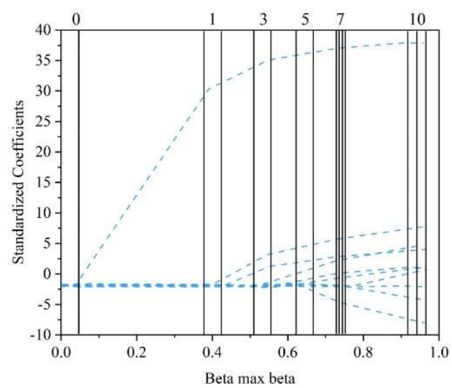

Figure 7: The path diagram of the minimum Angle regression method

IV. D. Consumer Decision Modeling Study

IV. D. 1) Model selection and setup

Secondly, the explanatory variables are analyzed by descriptive statistics, and the results are shown in Table 1, which show that the explanatory variables have different degrees of left or right skewness, and the standard deviation of the price and the number of collections is large, which indicates that the dispersion of the two variables is large. Based on the above characteristics, if the traditional mean regression is still used, the model will not be easy to estimate and the explanatory ability of the model will be worse, as mentioned above, the neural network quantile regression does not exist the assumption of normal distribution and is not sensitive to the perception of outliers in the model, and the constructed model is more robust, so the consumer decision-making model constructed in this paper should choose the quantile regression method. Therefore, the quantile regression method should be chosen for the consumer decision-making model constructed in this paper. In addition to the strong applicability of quantile regression to the model of this paper, it is also very favorable to the explanation of the consumer purchase decision in this paper. The previous chapter on the theory of consumer purchasing decision established the model of consumer purchasing decision in this paper, in which the explanatory variable is sales volume, sales volume represents the consumer's herd purchasing behavior, which is the consumer's purchasing decision, the higher the sales volume represents the easier for the consumer to make a purchasing decision, and the advantage of quantile regression is that it describes the full picture of the conditional distribution of the explanatory variable, and therefore can describe the different quantile levels of sales volume, and the quantile regression is also very advantageous for explaining the consumer purchasing decision. The advantage of quantile regression is that it describes the full picture of the conditional distribution of the explanatory variables, so that it can describe the different quantile levels of sales, which is very helpful in explaining the consumer's purchase decision.

Table 1: Explain the variable description statistics

Variable name	Mean	Median	Standard deviation	Degree of bias	Kurtosis
Running memory	3.67	4	1.492	0.354	0.911
Screen size	5.51	5.4	0.5683	-1.312	4.025
Core number	3.53	4.1	1.085	-0.952	-0.267
Listing duration	2.45	2	1.112	0.936	0.041
Price	2071.58	1786	1282.5	1.752	6.894
Collection number	2125.84	503	4836.5	5.490	43.135
Evaluation rate	0.3	0.15	0.193	3.024	11.394
Image comment rate	0.22	0.16	0.192	1.52	2.456
Evaluation rate	0.98	1	0.052	-11.482	212.2
Comprehensive score	5.12	5.13	0.08	-1.38	7.714

IV. D. 2) Estimation of model coefficients

The coefficients of the model were estimated using the quanterg package in R language, and five quartiles of 0.1, 0.3, 0.5, 0.7 and 0.9 were selected for the selection of the quartiles, and the mean regression was applied using the same data in the actual results, and the coefficients estimation results obtained were compared with the estimation results of the quartiles, and the results of the comparison are shown in Table 2.

The neural network quantile regression, although only five quantile points, 0.1, 0.3, 0.5, 0.7, and 0.9, were selected, clearly shows that the impact of each explanatory variable on sales is different at different quantile points, which suggests that there is a heterogeneous effect. Comparing mean regression with neural network quantile regression also reveals that there are differences between the two methods in terms of coefficient significance as well as the direction of influence.

Consumers cannot fully access information about the quality of goods and other information, so word-of-mouth will help consumers reduce the perception of shopping risk and enhance trust, thus promoting consumer purchase decisions. For consumers, the number of favorites has always been the focus of their attention, the more favorites a store has, the more it can reduce the cost of searching for information, thus effectively promoting consumer purchases. For stores with different sales segments, quality service subconsciously affects consumers, thus promoting consumer purchases. Higher memory does not significantly promote the purchase of consumers, memory capacity is not easy to be too large, in line with the normal needs of consumers can be, and too large a memory capacity may bring about a significant increase in price beyond the consumer's psychological expectations and thus unable to promote the purchase. The same goes for core count.

Table 2: Regression table

Variable name		The consumer decision model of the neural network is returned				
	Mean regression	0.1	0.3	0.5	0.7	0.9
Intercept term	-3.025***	-0.725	-0..992	-2.30	-5.298 ***	-4.82 ***
memory	-0.056**	-0.040***	-0.032	-0.033	-0.052**	-0.020
Screen size	0.184***	0.093***	0.134*	0.056	0.172 ***	0.211 ***
Core number	-0.036	-0.005	-0.007	0.030	-0.035	-0.10 ***
Listing duration	-0.189***	-0098***	-0.140 ***	-0.182 ***	-0.166 ***	-0.13 ***
price	-0.226***	-0.089***	-0.198***	-0.265***	-0.189***	-0.18***
Preferential treatment	0.042***	0.018	0.020	0.056***	0.054***	0.050*
Evaluation rate	1.245***	0.544*	0.613	0.894*	1.45**	1.472*
Evaluation rate	0.189	-0.052	0.046	0.284*	0.23	-0.098
Image comment rate	0.922***	0.325***	0.680***	0.842***	0.87***	0.574***
Collection number	0.475***	0.162***	0.335***	0.462***	0.547***	0.632***
Comprehensive score	0.660***	0.331*	0.400*	0.658***	1.063***	0.951***

IV. E. Comparison of forecast results

In the actual modeling process need to divide the data to test the prediction ability, in this paper is mainly selected cross-validation, the data is divided into ten equal parts, in order to select one of them as the test set, the rest of the nine as the training set, for each test set of data to calculate the MAE value, the RMSE value and the AIC value, the total number of calculations ten times, and ten times the results of the averaging, to obtain the final results. the final result.

After dividing the data and applying the neural network method, it is necessary to determine the number of hidden layer nodes, and the number of nodes in the hidden layer is too large, which will lead to the model being too complicated, so the method used in this paper is to select 1 to 5 hidden layer nodes, calculate the AIC values of different hidden layer nodes at different quartiles, and thus to determine the number of nodes in the hidden layer that makes the optimal AIC value, and 3 is selected as the number of nodes in the hidden layer after calculation. After calculation, 3 is selected as the number of hidden layer nodes, and the AIC values of the corresponding neural network quantile regression are calculated on the basis of the number of hidden layer nodes of 3. The comparison results are shown in Table 3.

Through the above table as a whole, the neural network quantile regression is superior to the linear quantile regression in terms of the combined situation of the three indicators. As far as each metric is concerned, for the RMSE metric and the MAE metric, the neural network quantile regression outperforms the linear quantile regression at all quantile points. Regarding the AIC indicator of model complexity, the AIC values of the two regressions do not differ much at each quantile point, and the neural network quantile regression is superior to the linear quantile regression at the low quantile point 0.1, 0.3, 0.5, and the high quantile point 0.9, while the linear quantile regression model at the middle and high qualile point 0.7 has better goodness-of-fit. Combining the results of these analyses, the goodness of fit of linear quantile regression is comparable to that of neural network quantile regression, but the neural network is more accurate than linear quantile regression in other metrics.

Table 3: Comparison of MAE, RMSE and aic values of the present

Index		RMSE	MSE	AIC
Linear fractional regression	0.1	1.685	1.292	23.956
	0.3	1.204	0.963	22.354
	0.5	0.996	0.785	21.883
	0.7	1.128	0.914	22.189
	0.9	1782	1.325	21.773
Neural network quantization regression	0.1	1.56	1.22	23.962
	0.3	1.096	0.835	22.48
	0.5	0.944	0.729	22.12
	0.7	1.18	0.84	22.46
	0.9	1.538	1.236	23.79

V. Conclusion

In this paper, the neural network model is combined with the quantile regression model to establish a neural network-based prediction model for consumer purchase decision. Using crawler technology, real data of laptop sales on e-commerce platforms from January to June 2024 are obtained, and the model is empirically tested. Compared with the linear quantile regression model, the neural network quantile regression method has high reliability of estimation results, better prediction ability, and strong nonlinear mapping ability. Moreover, the estimation results of this method have higher precision and better stability. The price factor has a negative and significant effect on consumer purchase. The word-of-mouth factors, such as after-sale rating, product evaluation, logistics evaluation, and after-sale evaluation, show a positive promotion effect. Screen size has a negative and positive influence on consumer purchase respectively.

References

[1] Prabowo, S. H. W., Murdiono, A., Hidayat, R., Rahayu, W. P., & Sutrisno, S. (2019). Digital marketing optimization in artificial intelligence era by applying consumer behavior algorithm. Asian Journal of Entrepreneurship and Family Business, 3(1), 41-48.

[2] Guan, Y. (2023). Consumer behavior analysis and marketing strategy optimization in the digital media environment. Industrial Engineering and Innovation Management, 6(10), 87-94.

[3] Theodorakopoulos, L., & Theodoropoulou, A. (2024). Leveraging big data analytics for understanding consumer behavior in digital marketing: A systematic review. Human Behavior and Emerging Technologies, 2024(1), 3641502.
[4] Zhang, D., & Huang, M. (2022). A Precision Marketing Strategy of e‐Commerce Platform Based on Consumer Behavior Analysis in the Era of Big Data. Mathematical Problems in Engineering, 2022(1), 8580561.
[5] Zhang, G. (2024). Exploring the enterprise marketing strategy innovation in the era of big data. Academic Journal of Business & Management, 6(11), 152-157.
[6] Chen, M. (2024). Research on the Digital Transformation of the Enterprise Marketing Ecosystem Based on Genetic Algorithms. International Journal of High Speed Electronics and Systems, 2540053.
[7] Valecha, H., Varma, A., Khare, I., Sachdeva, A., & Goyal, M. (2018, November). Prediction of consumer behaviour using random forest algorithm. In 2018 5th IEEE Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON) (pp. 1-6). IEEE.
[8] Wang, Y., Li, J., Liu, Q., & Ren, Y. (2015). Prediction of purchase behaviors across heterogeneous social networks. The Journal of Supercomputing, 71, 3320-3336.
[9] Dou, X. (2020, April). Online purchase behavior prediction and analysis using ensemble learning. In 2020 IEEE 5th International conference on cloud computing and big data analytics (ICCCBDA) (pp. 532-536). IEEE.
[10] Gallicchio, C. (2018). Short-term memory of deep RNN. arXiv preprint arXiv:1802.00748.
[11] Rojas, J. S., Rendon, A., & Corrales, J. C. (2019). Consumption behavior analysis of over the top services: Incremental learning or traditional methods?. IEEE Access, 7, 136581-136591.
[12] Ketipov, R., Angelova, V., Doukovska, L., & Schnalle, R. (2023). Predicting user Behavior in e-commerce using machine learning. Cybernetics and Information Technologies, 23(3), 89-101.
[13] Martínez, A., Schmuck, C., Pereverzyev Jr, S., Pirker, C., & Haltmeier, M. (2020). A machine learning framework for customer purchase prediction in the non-contractual setting. European Journal of Operational Research, 281(3), 588-596.
[14] Wang, W., Xiong, W., Wang, J., Tao, L., Li, S., Yi, Y., ... & Li, C. (2023). A user purchase behavior prediction method based on xgboost. Electronics, 12(9), 2047.
[15] Guo, L., Zhang, B., & Zhao, X. (2021). A Consumer Behavior Prediction Model Based on Multivariate Real‐Time Sequence Analysis. Mathematical Problems in Engineering, 2021(1), 6688750.
[16] Nguyen, M. D., & Cho, Y. S. (2019). A hybrid generative model for online user behavior prediction. IEEE Access, 8, 3761-3771.
[17] Xu, J., Wang, J., Tian, Y., Yan, J., Li, X., & Gao, X. (2020). SE-stacking: Improving user purchase behavior prediction by information fusion and ensemble learning. Plos one, 15(11), e0242629.
[18] Xiao, S., & Tong, W. (2021). Prediction of user consumption behavior data based on the combined model of TF-IDF and logistic regression. In Journal of physics: conference series (Vol. 1757, No. 1, p. 012089). IOP Publishing.
[19] XingFen, W., Xiangbin, Y., & Yangchun, M. (2018, December). Research on user consumption behavior prediction based on improved XGBoost algorithm. In 2018 IEEE international conference on big data (Big Data) (pp. 4169-4175). IEEE.
[20] Tian, Y., Lai, Y., & Yang, C. (2022). Research of consumption behavior prediction based on improved DNN. Scientific Programming, 2022(1), 6819525.
[21] Chuang Wan,Wei Zhong,Chenjing Li & Xinyuan Song. (2025) .Interquantile shrinkage and variable selection for longitudinal data in regression models. Science China Mathematics,(prepublish),1-26.
[22] Pham Phuong Thao & Le Thi Thu Huong. (2025) .Unveiling key determinants of higher education expenditure in Vietnam: an adaptive LASSO approach in Tobit regression analysis. Journal of Applied Research in Higher Education,17(2),833-851.
[23] Babak Mohammadi,Mohammed Abdallah,Rachid Oucheikh,Okan Mert Katipoğlu & Majid Cheraghalizadeh. (2025) .Enhancing streamflow drought prediction: integrating wavelet decomposition with deep learning and quantile regression neural network models. Earth Science Informatics,18(2),232-232.
[24] Yasin Altinisik,Roy S. Hessels,Caspar J. Van Lissa & Rebecca M. Kuiper. (2025) .An AIC-type information criterion evaluating theory-based hypotheses for contingency tables. Behavior Research Methods,57(2),70-70.
[25] Lei Shu,Yifan Hao,Yu Chen & Qing Yang. (2025) .SFQRA: Scaled factor-augmented quantile regression with aggregation in conditional mean forecasting. Journal of Multivariate Analysis,207,105405-105405.