Sensors and Materials, Vol. 36, No. 6 (2024) 2629–2641
MYU Tokyo

S & M 3691

Analysis of Campus Catering Data Using Machine Learning

Chien-Min Chen,[1,2] Chen-Sheng Li,[3] and Shih-Pang Tseng[4,5*]

[1]School of Software, Quanzhou University of Information Engineering,
No. 249, Bodong Road, Fengze District, Quanzhou, Fujian, 362000 China
[2]International School of Finance, Fudan University, No. 220, Handan Road, Shanghai, 200433 China
[3]School of Computer Engineering, Jiangsu University of Technology,
No. 1801 Zhongwu Road, Changzhou, Jiangsu, 213001 China
[4]School of Software and Big Data, Changzhou College of Information Technology,
No. 22, Mingxin Middle Road, Changzhou, Jiangsu, 213164 China
[5]School of Information Science and Technology, Sanda University,
No. 2727 Jinhai Road, Pudong District, Shanghai, 201209 China

(Received December 19, 2023 accepted June 24, 2024)

Keywords: machine learning, big data, linear regression, decision tree

At present, sensors and big data create powerful systems capable of real-time monitoring and decision-making. The development of big data in all walks of life is very fast. Big data technology and applications are also gradually accepted by the public, and the data industry is gradually maturing and beginning to enter a rapid development stage. At the same time, the development of the Internet is making data analysis more accurate, and the combination of the two complements each other, contributing to the good development of big data. With the rapid development and wide application of machine learning technology, its application in all walks of life has become increasingly widespread. In this work, we collect the business data of campus restaurants in different time periods to ensure the breadth and depth of the data. The regression algorithm and decision tree algorithm in machine learning are used to integrate and analyze the collected data to reflect the demand tendency of the general public and also the relationship between cost and benefit. We analyze the catering needs of different consumer groups, develop big data applications with greater potential value, and seek long-term development for catering real economy enterprises.

1. Introduction

Sensors are devices that detect and respond to various types of input from the physical environment. These inputs might be light, heat, motion, moisture, pressure, or any other physical quantity. Sensors convert these inputs into electronic signals that can be measured, monitored, and analyzed. In the context of modern technology, sensors are ubiquitous and found in everything from smartphones and cars to industrial machinery and smart campuses. Therefore, these various sensors generate huge amounts of structured/unstructured data.

*Corresponding author: e-mail: tsengshihpang@czcit.edu.cn
https://doi.org/10.18494/SAM4836

ISSN 0914-4935

Big data refers to the vast volumes of data generated by people, devices, and systems every second. This data is characterized by its high volume, velocity, and variety, making traditional data processing software inadequate to handle it. Big data involves not just the collection of massive datasets but also their storage, management, and analysis to uncover patterns, trends, and associations, especially relating to human behavior and interactions.

When combined, sensors and big data create powerful systems capable of real-time monitoring and decision-making. Sensors generate a continuous stream of data, which, when processed through big data technologies, can provide valuable insights. For example, in healthcare, wearable sensors can monitor patient vitals and transmit this data for analysis, leading to the early detection of health issues. In smart cities, sensors on infrastructure can collect data on traffic, weather, and energy usage, which can then be analyzed to optimize city planning and resource allocation.

In the 21st century, the Internet has developed rapidly, the huge amount of data has also ushered in explosive growth, and the big data industry has also made a breakthrough. At present, the generation of big data in all walks of life is very fast. All countries are increasing support, and the scale of the global big data market maintains a high growth trend. Big data technology and applications are also gradually accepted by the public, and the data industry is gradually maturing and beginning to enter a rapid development stage.(1,2) At the same time, the development of the Internet makes data analysis more accurate, and the combination of the two complements each other, contributing to the good development of big data.

Catering profit benefit is the key to the development of enterprises. In the actual operation process, owing to the impact of cost, epidemic prevention and control, store address, and other factors, the profit and benefit of enterprises differ considerably, and the business data obtained will appear complex and chaotic, which cannot accurately reflect the operation of enterprises. In addition, the amount of data obtained from the operation of catering enterprises is relatively large, reaching the level of big data. Therefore, it is difficult to process and apply these data, which has become one of the important factors to judge whether the operation of catering enterprises is healthy. How to accurately fit the relevant cost data, obtain the trace of future operating benefits, and make adjustments to obtain the maximum economic benefits are some of the urgent problems in the catering industry.

With the rapid development and wide application of machine learning technology, its application in all walks of life has become increasingly widespread.(3,4) In the field of campus catering, the application of machine learning is also gradually rising, which is of great significance for providing healthy and delicious catering services.(5) Campus catering is an essential part of college students' daily life. However, the management of campus catering is facing a series of challenges, such as food safety management, supply chain management, and nutritional balance.(6) These problems have a direct impact on students' dietary experience.

The traditional way of food and beverage management is often faced with the problem of a large amount of data accumulation but insufficient utilization,(7) which cannot fully explore valuable information in the data. However, using machine learning technology to analyze and mine campus catering data can help decision makers understand students' dietary preferences and needs more accurately, and improve the quality and efficiency of catering services.(8)

The purpose of this study is to conduct in-depth analysis and content visualization of campus catering data using machine learning technology and provide reference information on food supply, consumption trends, and student needs for campus catering managers.(9) By fully tapping the potential value of campus catering data, we can improve and optimize campus catering services and enhance students' catering experience.

2. Data Source Analysis and Data Preprocessing

With the increase in campus population and economic advancement, the campus catering industry has become a very important and prosperous industry in colleges and universities. At the same time, with the progress of science and technology, the data age has come, and a large number of data mining, analysis, and modeling needs also follow. Therefore, for the campus catering industry, how to use valuable data resources to achieve better management and operation has become one of the concerns of relevant managers and researchers. Combined with the current situation of campus catering, in this study, we analyze the data sources of campus catering.

The catering industry is a growing industry, attracting more and more consumers with its diversified food, high-quality service, and pleasant dining atmosphere. However, in this industry, the data represented by catering data are generated very rapidly, including order information, restaurant ratings, and sales. These data are very important for catering enterprises, because they can help catering enterprises understand the needs and behaviors of consumers and provide them with better services and products. However, these data often have many errors and inconsistencies, which require catering data preprocessing.

2.1 Outlier detection

Outlier detection among restaurant operation data and questionnaire data can improve the quality of data.(10) The operational data are all continuous data, and the data content is provided by the restaurant. Thus, missing values may occur. We can use the drop function in the data frame to delete missing data rows. Through analysis, we found that there are four pieces of missing values in the operation data, and they were deleted, as shown in Table 1.

2.2 Feature selection and eigenvalue mapping

For the questionnaire data, because the data is collected by a specific program, there is no missing value, but among the many attributes, we need to select the attribute that is most in line

Table 1
Abnormal data.

	Material_cost	Sales	Labor_cost	Rental_cost	Other_cost	Profit
Beef		2877.67	198	111.1	30	1815.01
Mutton	726.59	2860.30	198	111.1	30	
Carp	732.86	2974.35		111.1	30	1902.39
Black fish	756.89		198	111.1	30	2028.76

with the analysis requirements. By observing the original questionnaire data, we found that some attributes have little effect on "whether to choose light food", while too much useless data will have a certain impact on the prediction performance of the decision tree. To obtain more accurate prediction performance, some attributes need to be manually deleted from the questionnaire data.

Through filtering, we select the following attributes as the training and testing data of the decision tree, and store the filtered and mapped questionnaire data in the treeexec.csv file, which contains 3297 pieces of data. Owing to space constraints, Tables 2 and 3 only show the first 10 pieces of data.

3. Machine Learning Algorithms

3.1 Linear regression

Linear regression is a machine learning algorithm widely used in prediction and analysis and is widely used in many industries and fields.[11–13] The goal of the algorithm is to predict the expected results (dependent variables) of the input from one or more features (independent

Table 2
Filtered and mapped questionnaire data.

Living_expenses_ situation	Average_ monthly_ consumption	Dining_in_the_ school	Ideal_cost_per_ meal	Thoughts_on_ living_expenses	Most_important_ factor_in_ catering_industry
over 2000	1500 to 2000	300 to 600	11 to 15	mtn	ch
1500 to 2000	over 2000	600 to 1000	11 to 15	je	ch
over 2000	over 2000	over 1000	11 to 15	je	ch
600 to 1000	600 to 1000	600 to 1000	6 to 10	ne	ch
1500 to 2000	1000 to 1500	600 to 1000	11 to 15	mtn	ch
1500 to 2000	1000 to 1500	300 to 600	11 to 15	mtn	ch
over 2000	over 2000	over 1000	over 20	je	ch
1500 to 2000	1500 to 2000	below300	11 to 15	je	ch
1500 to 2000	1000 to 1500	600 to 1000	15 to 20	mtn	ch
1500 to 2000	1500 to 2000	over 1000	11 to 15	je	ch

Table 3
Filtered and mapped questionnaire data.

Dining_options	Understanding_light_ eating_doctrine	Reasons_for_ choosing_light_meals	Food_culture	Choose_light_food
safety	understand	slimming	both	TRUE
safety	common	nutrition	both	FALSE
taste	common	slimming	both	FALSE
taste	not understand	delicious	both	TRUE
taste	not understand	slimming	carnivorism	TRUE
taste	common	slimming	both	TRUE
safety	understand	nutrition	carnivorism	TRUE
safety	common	slimming	carnivorism	TRUE
safety	common	slimming	carnivorism	TRUE
safety	common	slimming	both	FALSE

variables) by finding the best fitting linear equation. The least squares method is the core of the linear regression algorithm. It predicts by finding a line that is best suitable for all existing data. The specific method is to convert the target variable and independent variable of linear regression, and then use the known data to solve the linear equations to obtain the best fitting curve, so as to achieve the optimization goal of "minimizing the sum of squares of residuals".

In machine learning, cost function refers to the difference between the predicted value and the actual value of the model, which is one of the indicators to evaluate the performance of the model. Among them, the gradient descent algorithm is a cost-function-based algorithm and is mainly used to update the weight parameters of the model to minimize the cost function.(14) Specifically, the cost function is expressed by a mathematical formula, which is the sum of squares of the differences between all the predicted values and the actual values divided by the number of samples, and is widely used in linear regression and other tasks. In the gradient descent algorithm, we need to select an appropriate learning rate to find the minimum value of the cost function through iteration. The steps of the gradient descent algorithm include (1) initializing parameters, (2) solving the partial derivative of the cost function, (3) updating the parameters to make the cost function as small as possible, and (4) repeating the above steps until the value of the cost function does not decrease or reaches the predetermined number of iterations.

The known set of data is described as

$$\left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{i=m}, \text{ where in } x^{(i)} = \left[x_1^{(i)} \quad \ldots \quad x_n^{(i)}\right]^T, \\ y^{(i)} = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \ldots \theta_n x_n^{(i)} + \varepsilon, \tag{1}$$

where θ_0 is the intercept, θ_1, θ_2, ..., θ_n is the weight, and ε is the amount of error random variable. In addition, ε is a normal distribution, $N(0, \sigma^2)$. Equation (1) represents a theoretical regression model for which there are two main assumptions. Firstly, dependent variable y and independent variable x have a linear relationship among the dependent variables. Secondly, the value of the independent variable x is fixed and nonrandom in repeated sampling.

Under the above two assumptions, any given value corresponds to the distribution

$$f\left(x^{(i)}\right) = \theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \ldots \theta_n x_n^{(i)}, \tag{2}$$

where $f(x^{(i)})$ represents a line/plane/hyperplane, but since a single data point is extracted from the distribution of Y, it is impossible to be on this line/plane/hyperplane, hence it must contain an error term ε. We hope that the smaller error term ε of σ^2, the better. The observed value of y is to the line/plane/hyperplane described by $f(x^{(i)})$. Thus, the problem is to find a minimal line/plane/hyperplane σ^2, that is, to solve the intercept and weight of $f(x^{(i)})$. In machine learning, intercept and coefficient are usually solved using the loss function and gradient descent method.

We use the mean square error (*MSE*) to construct the loss function $J(\theta)$.

$$J(\theta)=\frac{1}{2m}\sum_{i=1}^{i=m}\left(f\left(x^{(i)}\right)-y^{(i)}\right)^2 \tag{3}$$

$$J(\theta)=\frac{1}{2m}\sum_{i=1}^{i=m}\left(\theta_0+\theta_1x_1^{(i)}+\theta_2x_2^{(i)}+\ldots\theta_nx_n^{(i)}-y^{(i)}\right)^2 \tag{4}$$

Here, the multiplication 1/2*m* is for the derivative. According to the convex optimization principle, this is a quadratic function, namely, the convex function, which exists as a minimum value.

To find the best-fitting line/plane/hyperplane, we use the gradient descent method to find the optimal solution along the fastest direction through continuous iteration. The partial derivatives of each independent variable are to find the direction with the largest slope to change the function. That is the gradient. We can change the values of *x* along the opposite direction of the gradient to advance to the minimum point. After a certain number of iterations, the minimum point can be approximately found in the function.

The partial derivative of the loss function is

$$\frac{\partial J(\theta)}{\partial\theta_0}=\sum_{i=1}^{m}\left(\theta_1x_1^{(i)}+\theta_2x_2^{(i)}+\ldots\theta_nx_n^{(i)}-y^{(i)}\right), \tag{5}$$

$$\frac{\partial J(\theta)}{\partial\theta_p}=\sum_{i=1}^{m}x_p^{(i)}\left(\theta_1x_1^{(i)}+\theta_2x_2^{(i)}+\ldots\theta_nx_n^{(i)}-y^{(i)}\right).(1\le p\le n) \tag{6}$$

According to the partial derivative, the minimum value can be approached gradually by iteration, and the iteration formula is

$$\text{repeat}\left\{\theta_p=\theta_p+\alpha\frac{\partial J(\theta)}{\theta_p}\right\}\text{ until }\left(J(\theta)-J'(\theta)<T\text{ or IterNum}>\text{MaxIter}\right). \tag{7}$$

$J'(\theta)$ means the loss function of the previous iteration. This loop would be terminated when the loss difference between the previous and current iterations is less than the threshold *T* or the iterations is to the maximum number of iterations. α represents the learning rate. The iterative visualization process is shown in Fig. 1, where the specific $\{\theta_p\}_{p=0}^{n}$ calculation process is

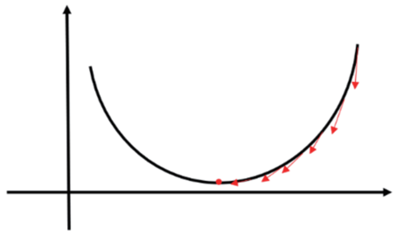

Fig. 1. (Color online) Iteration process.

$$\begin{cases} \theta_0 = \theta_0 + \alpha \sum_{i=1}^{m} \left(\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \ldots \theta_n x_n^{(i)} - y^{(i)}\right), \\ \theta_1 = \theta_1 + \alpha \sum_{i=1}^{m} x_1^{(i)} \left(\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \ldots \theta_n x_n^{(i)} - y^{(i)}\right), \\ \quad \vdots \\ \theta_n = \theta_n + \alpha \sum_{i=1}^{m} x_n^{(i)} \left(\theta_0 + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \ldots \theta_n x_n^{(i)} - y^{(i)}\right). \end{cases} \tag{8}$$

To investigate the effect of the fitted discrimination function and to judge the fit of the training data and the prediction accuracy of the test data, the following root mean square error (*RMSE*) and R^2 are two measurementsare two measurements to investigate the effect of the fitted discrimination function and to judge the fit of the training data and the prediction accuracy of the test data.

The *RMSE* is the difference between the prediction and the actual results. If the *RMSE* is smaller, the error between the predicted value and the true value is smaller. The calculation formula is

$$RMSE = \sqrt{\frac{\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - y^{(i)}\right)^2}{m}}, \tag{9}$$

where m is the total number of test samples.

R^2 describes the fitting degree of the regression line to the observed data. The range of R^2 values is [0,1]. The closer R^2 is to 1, the greater the proportion of the sum of squares in the total sum of squares, the closer the regression line is to each observation point, and the better the fitting degree of the regression line is.

$$R^2 = \frac{\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - E(Y)\right)^2}{\sum_{i=1}^{m}\left(y^{(i)} - E(Y)\right)^2} = 1 - \frac{\sum_{i=1}^{m}\left(f\left(x^{(i)}\right) - y^{(i)}\right)^2}{\sum_{i=1}^{m}\left(y^{(i)} - E(Y)\right)^2} \tag{10}$$

Here, $Y = \left\{y^{(i)}\right\}_{i=1}^{i=m}$ representations are the mathematical expectations.

3.2 Decision tree

A decision tree is a classification model that sequentially evaluates the features of samples to determine the category of the samples.[15,16] It selects the optimal feature on the training sample set to split the original sample set for achieving the best information gain (minimizing the loss function). This process is repeated on the newly generated subsets of samples until all training samples are essentially correctly classified or there are no suitable features left. The entire process can be visualized as a tree structure, with leaf nodes storing well-classified sample sets and branch nodes storing partition conditions based on features.

In the ID3 algorithm for decision trees, information gain is used to select the best feature for data set partitioning.[17,18] Information gain is calculated on the basis of entropy. Entropy is determined by the distribution of each category in the data set.[19,20] In the decision tree construction process, a lower entropy is preferred as it indicates a higher dataset purity. Information gain quantifies the change in entropy after partitioning the data set with a specific feature.

Information entropy is a measure of the overall uncertainty of the source.[21] It is based on the category of samples in sample set S, such as y_1, y_2, ..., y_k. The probability calculation and representation of each type of sample on S can be expressed as $P(y_1)$, $P(y_2)$, ..., $P(y_k)$. The information entropy is defined as

$$\begin{aligned} E(S) &= -P(y_1)\log P(y_1) - P(y_2)\log P(y_2) - \ldots - P(y_k)\log P(y_k) \\ &= -\sum_{j=1}^{k} P(y_j)\log P(y_j), \end{aligned} \tag{11}$$

where $P(y_j)$ is actually the proportion of samples with category Y_i in S. The smaller the value of the information entropy, the smaller the uncertainty and the higher the certainty.

On this basis, the information gain, as a measure of the difference between the two information quantities, further considers the structure of the samples in sample set S, including the categories and attributes of the samples. Different subsets can be divided by different values of attributes. Under this structure, we can calculate the weighted information entropy of sample set S divided by the different values of attribute X_i: $E(S,x_i) = \sum_{t=1}^{r} \frac{|S_t|}{|S|} \times E(S_t)$. After calculating the information entropy and weighted information entropy, we can obtain the information gain, $(S,x_i) = E(S) - E(S,x_i) = E(S) - \sum_{t=1}^{r} \frac{|S_t|}{|S|} \times E(S_t)$, that is, the information gain is equal to the difference between the information entropy and weighted information entropy. The greater the value, the higher the certainty of the information.

Finally, the decision tree can be constructed step by step by selecting conditional attributes for expansion based on the principle of maximum information gain. Suppose $S = \{s_1, s_2, ..., s_n\}$ is the entire sample set, $X = \{x_1, x_2, ..., x_m\}$ is the entire attribute set, and $Y = \{y_1, y_2, ..., y_k\}$ is the sample category; this summarizes the complete process of the ID3 algorithm.

4. Model Construction and Prediction

The linear regression and decision tree are constructed and applied on the campus catering data. The linear regression is used to analyze the transaction data of the restaurant and to predict profits under different operating costs. In addition, the ID3 decision tree is used to analyze the characteristics of consumers who chose the diet food.

4.1 Linear regression for profit

The dataset contains two independent variables, cost and sales. The bivariate linear regression is used to predict the profit value. In addition, each index of the regression model is also calculated for analysis and discussion.The linear function of independent and dependent variables is assumed to be

$$f_3\left(x^{(i)}\right) = \theta_{3,0} + \theta_{3,1}x_1^{(i)} + \theta_{3,2}x_2^{(i)}, \tag{12}$$

where $x_1^{(i)}$ and $x_2^{(i)}$ show the cost and sale of the ith sample, respectively, and $\theta_{3,0}$, $\theta_{3,1}$,and $\theta_{3,2}$ show the offset and weights. The loss function J_3 is set to

$$J_3\left(\theta_3\right) = \frac{1}{2m}\sum_{i=1}^{i=m}\left(f_3\left(x^{(i)}\right) - y^{(i)}\right)^2. \tag{13}$$

We conduct the partial derivative of $J_3(\theta_3)$ for $\theta_{3,0}$, $\theta_{3,1}$, and $\theta_{3,2}$. Next, the gradient descent method is iteratively used to find the minimum. The iterative formula is

$$\begin{cases} \theta_{3,0} = \theta_{3,0} + \alpha\sum_{i=1}^{m}\left(\theta_{3,0} + \theta_{3,1}x_1^{(i)} + \theta_{3,2}x_2^{(i)} - y^{(i)}\right), \\ \theta_{3,1} = \theta_{3,1} + \alpha\sum_{i=1}^{m}x_1^{(i)}\left(\theta_{3,0} + \theta_{3,1}x_1^{(i)} + \theta_{3,2}x_2^{(i)} - y^{(i)}\right), \\ \theta_{3,2} = \theta_{3,2} + \alpha\sum_{i=1}^{m}x_2^{(i)}\left(\theta_{3,0} + \theta_{3,1}x_1^{(i)} + \theta_{3,2}x_2^{(i)} - y^{(i)}\right), \end{cases} \tag{14}$$

until ($J_3(\theta_3) - J'_3(\theta_3) < T$ or IterNum > MaxIter).

Here, $J'_3(\theta_3)$ shows the loss in the previous iteration. When the current iteration loss is less than the previous loss under the threshold $T = 10^{-12}$ or the iterations are larger than 10^{-10}, the iterative process is stopped. $\alpha = 0.01$ shows the learning rate.

For evaluating the fitness of this regression model, the *RMSE* and R^2 index are

$$RMSE_3 = \sqrt{\frac{\sum_{i=1}^{m}\left(f_3\left(x^{(i)}\right) - y^{(i)}\right)^2}{m}}, \tag{15}$$

$$R^2{}_3 = 1 - \frac{\sum_{i=1}^{m}\left(f_3\left(x^{(i)}\right) - y^{(i)}\right)^2}{\sum_{i=1}^{m}\left(y^{(i)} - E(Y)\right)^2}. \tag{16}$$

In the physical case, $RMSE_3 = 3.466$ and $R^2{}_3 = 0.99$. The fitting plane, shown in Fig. 2, can be used to predict the profit under various operation combinations. In Fig. 3, the residuals of variables show the confidence of this regression process.

4.2 Decision tree for consumer characteristics

Each data sample contains 10 characteristics of a consumer, namely, living expenses, average monthly consumption, school meals, ideal cost of each meal, views on living expenses, most important factors in the catering industry, dining choices, understanding the concept of light food, reasons for choosing light food, and food culture. Each sample also corresponds to a classification result, that is, whether the consumer is consuming at the merchant, which is a true or false value.

Suppose the initial sample set is s, the initial attribute set is $X = \{x_1, x_2, ..., x_{10}\}$, and the sample category set is $Y = \{y_1, y_2\}$. To generate each node of the decision tree, we need to select the partition attribute x_1 that can obtain the maximum information gain, even if the weighted

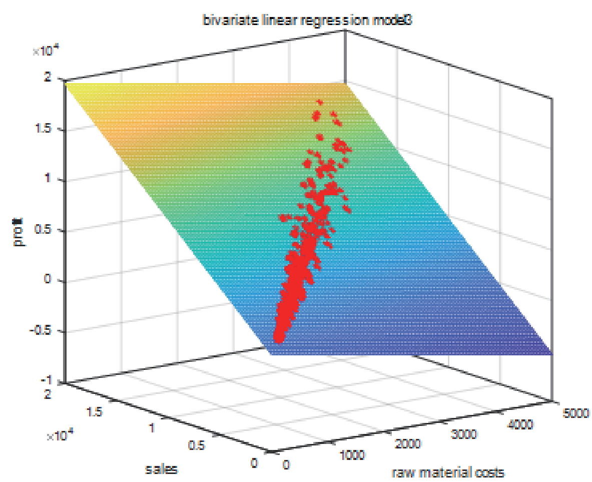

Fig. 2. (Color online) Fitting plane.

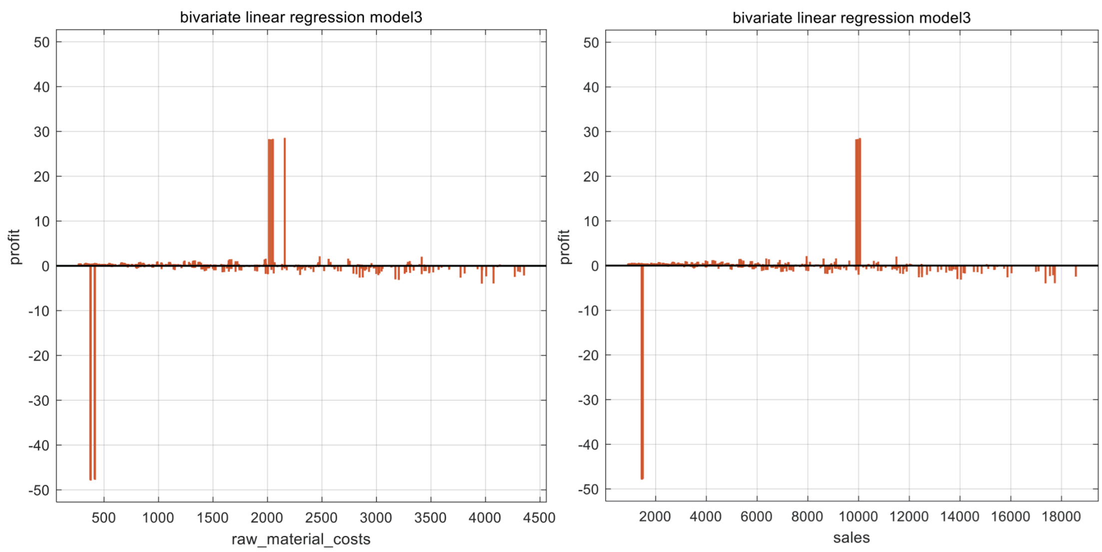

Fig. 3. (Color online) Residuals of variables.

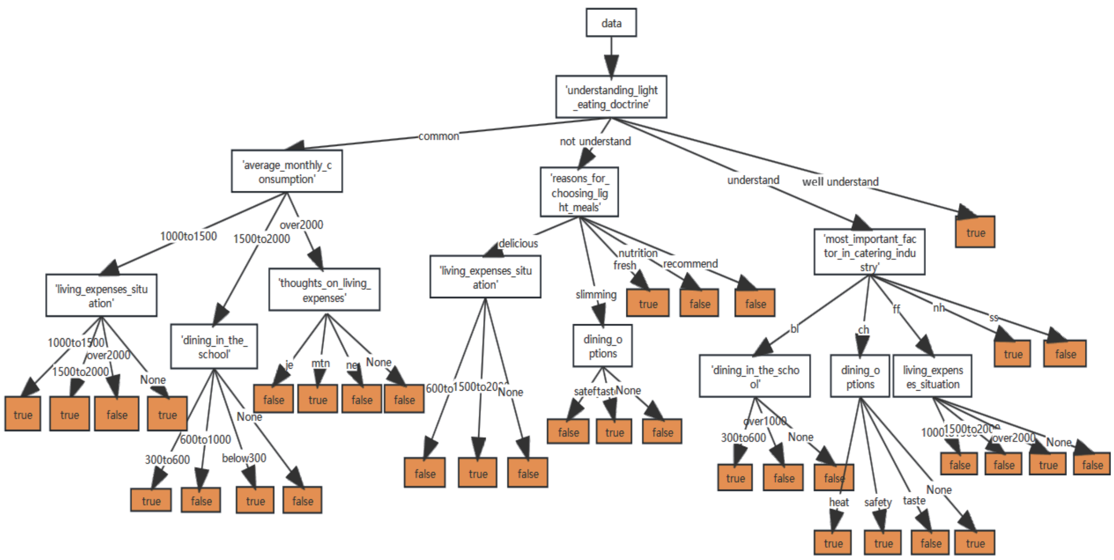

Fig. 4. (Color online) Decision tree visualization.

information entropy of sample set s under different values of attribute x_1 is the smallest. We set the initial partition attribute as x_8 and calculate the weighted information entropy of sample set s divided by x_8. Then, as shown in the formula, we continue to calculate the information entropy divided by other attributes and compare their sizes. Finally, the attribute with minimum conditional entropy is selected as the key feature.

According to the comparison, the key feature (minimum conditional entropy) is x_8. The attribute x_8 is used to split the original sample set into multiple sample subsets. Then, similar steps as those above would be repeated on the sample subsets until the generated decision tree

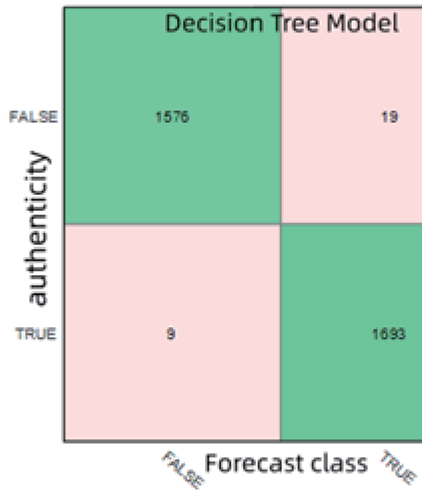

Fig. 5. (Color online) Confusion matrix.

meets the requirements or there are no available attributes to divide. Figure 4 shows the final decision tree of this study. This decision tree can assist the decision making to maximize the profit. To evaluate the confidence of the model, the confusion matrix is calculated, as shown in Fig. 5.

5. Conclusion

The analysis of campus catering data using machine learning provides a valuable technical way of gaining a deeper understanding of the operation mode of campus catering and the consumption behavior of students. Such analysis starts with data preprocessing, including data cleaning to remove noise and outliers, and processing of missing data. These steps ensure that the data used are of good quality and provide a reliable basis for subsequent analysis.

On the basis of this derived intelligence, we can build and train a variety of machine learning models, including linear regression, decision tree, and random forest. These models can reflect students' consumption behavior and catering demand patterns to various degrees. After the model training and prediction, we can not only understand the behavior pattern of campus catering, but also predict future consumption behavior. This enables the school to better manage catering services, such as reasonable human resource allocation, reasonable food procurement, and inventory management.

Finally, the results show that the application of machine learning improves the efficiency of campus catering management and makes the use of resources more reasonable. However, at the same time, we should also be clear that any model is a simplification of reality and cannot predict all situations 100% accurately. Therefore, in practical application, we should use it flexibly and adjust it in accordance with the actual situation.

The analysis of campus catering data based on machine learning reveals various patterns of students' consumption behavior and catering services. It may also be used as a tool to provide more accurate and personalized catering services and improve the quality of students' campus life in the future.

References

1 A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih: J. King Saud Univ. Comput. Inf. Sci. **30** (2018) 431. https://doi.org/10.1016/j.jksuci.2017.06.001
2 N. Abirami R., B. Balusamy, S. Kadry, and A. H. Gandomi: Big Data: Concepts, Technology, and Architecture (John Wiley & Sons, New York, 2021) 1st ed.
3 K. Sharifani and M. Amini: World Inf. Technol. Eng. J. **10** (2023) 3897. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4458723
4 M. Mohammed, M. B. Khan, and E. B. M. Bashier: Machine Learning: Algorithms and Applications. (CRC Press, Boca Raton, 2016) 1st ed.
5 G. Vlontzos and P. M. Pardalos: Int. J. Sustainable Agric. Manage. Inf., **3** (2017) 44. https://doi.org/10.1504/IJSAMI.2017.082921
6 V. Carchiolo, M. Grassia, A. Longheu, M. Malgeri, and G. Mangioni: Proc. 2021 the 6th Int. Conf. Complexity, Future Information Systems and Risk (COMPLEXIS 2021) 117–124.
7 L. Yang, H. Liu, and P. Y. Yen: Machine Learning in the Catering Industry, Encyclopedia of Data Science and Machine Learning (IGI Global, Pennsylvania, 2023).
8 J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng: EURASIP J. Adv. Signal Process. **2016** (2016) 1. https://doi.org/10.1186/s13634-016-0355-x
9 W. Wang, L. Y. Duan, H. Jiang, P. Jing, X. Song, and L. Nie: ACM Trans. Multimedia Comput. Commun. Appl. **17** (2021) 1. https://doi.org/10.1145/3418211
10 A. Boukerche, L. Zheng, and O. Alfandi: ACM Comput. Surv. **53** (2020) 1. https://doi.org/10.1145/3381028
11 D. A. Freedman: Statistical Models: Theory and Practice. (Cambridge University Press, Cambdirge 2009) 2nd ed.
12 X. Yan and X. G. Su: Linear Regression Analysis: Theory and Computing (World Scientific, Singapore, 2009) 1st ed.
13 D. C. Montgomery, E. A. Peck, and G. G. Vining: Introduction to Linear Regression Analysis (John Wiley & Sons, New York, 2021) 6th ed.
14 S. Ruder: arXiv:1609.04747 (2016). https://doi.org/10.48550/arXiv.1609.04747
15 Y. Song and Y. Lu: Shanghai Arch. Psychiatry **27** (2015) 130. https://doi.org/10.11919/j.issn.1002-0829.215044
16 Priyanka and D. Kumar: Int. J. Inf. Decis. Sci. **12** (2020) 246. https://doi.org/10.1504/ijids.2020.10029122
17 A. Rajeshkanna and K. Arunesh: Proc. 2020 Int. Conf. Electronics and Sustainable Communication Systems (IEEE, 2020) 787–790.
18 S. Chanmee and K. Kesorn: Adv. Eng. Inf. **58** (2023) 1474. https://doi.org/10.1016/j.aei.2023.102156
19 B. Charbuty and A. M. Abdulazeez: J. Appl. Sci. Technol. Trends **2** (2021) 20. https://doi.org/10.38094/JASTT20165
20 M. Azad, I. Chikalov, S. Hussain, and M. Moshkov: Entropy **23** (2021) 808. https://doi.org/10.3390/e23070808
21 R. M. Gray: Entropy and Information Theory (Springer, New York, 2011) 1st ed.