

Electric Vehicle Sales Analysis Model Based on User Purchase Intention Analysis

Ying Zhang^{1,*,#}, Yibing Chen^{2,#}, Peisen Huang^{3,#}

¹College Of Economics&Management, Huazhong Agricultural University, Wuhan, Hubei China

²College Of Resources&Environment, Huazhong Agricultural University, Wuhan, Hubei China

³College Of Engineering, Huazhong Agricultural University, Wuhan, Hubei China

*Corresponding author: daphne8840@163.com

#These authors contributed equally.

Abstract: New energy vehicles are a major innovation field in China's 12th Five-Year Plan. In order to study the sales strategy of target customers of electric vehicles, this paper establishes the sales prediction model of electric vehicles. First, the initial decision tree model of customer mining of different brands of electric vehicles is constructed. The model was optimized by changing the minimum sample number of leaf nodes and pruning. At the same time, we do cross validation error analysis on three decision tree models, and finally get a stable and sensitive customer mining model. In the evaluation of the model, four indicators TP, FN, FP and TN in the confusion matrix were used to calculate the classification accuracy, recall rate, precision rate and F1 score, and the calculation results showed that the model had good classification and prediction effect.

Keywords: Automobile sales, decision tree model, purchase intention analysis

1. Introduction

The automobile industry occupies an important position in the national economy, and the new energy automobile industry is a strategic emerging industry supported by the state. Vigorously developing new energy vehicles such as electric vehicles can effectively solve the energy and environmental problems, and has a broad market prospect. However, as an emerging product, compared with traditional cars[1], consumers have some doubts in battery issues and other aspects, and the market sales need to be made scientifically. In this paper, customer mining models of different brands of ELECTRIC vehicles are established and evaluated. Based on the sales data of customers of different brands of electric vehicles, the intention of target customers to buy electric vehicles is analyzed[2].

2. Establishment of decision tree model

2.1 Introduction to decision tree classification process

Decision tree is a top-down recursive method. Attribute values are compared among nodes in the decision tree and the branches down from the node are judged according to different attribute values, and conclusions are obtained in the decision-making leaf node[3].

In this case, a multi-variable attribute classification model should be established. The algorithm has several main indicators, which are as follows[4]:

1) Information entropy: Information entropy is the information contained in a set of data, a measure of probability. The more ordered a set of data, the lower the entropy of information. The formula is:

$$H(D) = -\sum_{i=1}^n \frac{d_i}{\text{sum}} \log_2 \frac{d_i}{\text{sum}} \quad (1)$$

D_i is the sample value of the attribute, and sum is the sum

2) Information gain: the effective reduction of information entropy. The higher this value is, the more information entropy the target attribute loses in the reference attribute, and the higher the attribute should be in the upper decision tree[5].

3) Split information: The split information measure is used to consider the number and size of branches when an attribute is split.

4) Gain ratio: Gain ratio is defined as the ratio of information gain to split information.

Figure1 shows the flowchart of establishing multivariate attribute classification.

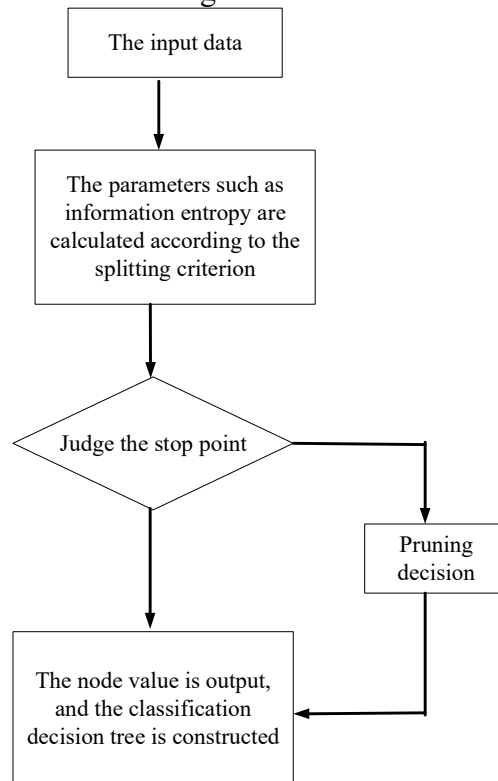


Figure 1 Flowchart for establishing multivariate attribute classification

2.2 Evaluation indicators of the model

(1) The four values of the confusion matrix are shown in Table 1.

Table 1 Four values of the confusion matrix

indicators	meaning
True Positive(TP)	Predict the number of positive classes
False Negative(FN)	Predict the number of positive classes as negative classes
False Positive(FP)	Predict the number of negative classes as positive classes
True Negative(TN)	Predict the number of negative classes

In this paper, we define a positive class for users to buy cars, and a negative class for users not to buy cars.

(2) Accuracy of classification

Refers to the proportion of correctly classified samples to total samples.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

(3) Recall rate (R)

Predict the correct proportion of the sample that is actually positive. $R = \frac{TP}{TP + FN}$

(4) Accuracy (P)

The percentage of the sample whose prediction is positive is correct. $P = \frac{TP}{TP + FP}$

(5) F1 Score

The F1 score is the harmonic mean of recall and precision. The range of F1 is between [0,1], and the closer it is to 1, the better the classification effect is. $F_1 = \frac{2TP}{2TP + FP + FN}$

3. Solution of model

3.1 Decision tree solution

The initial model constructed is used to generate a decision tree, as shown in Figure 2[6].

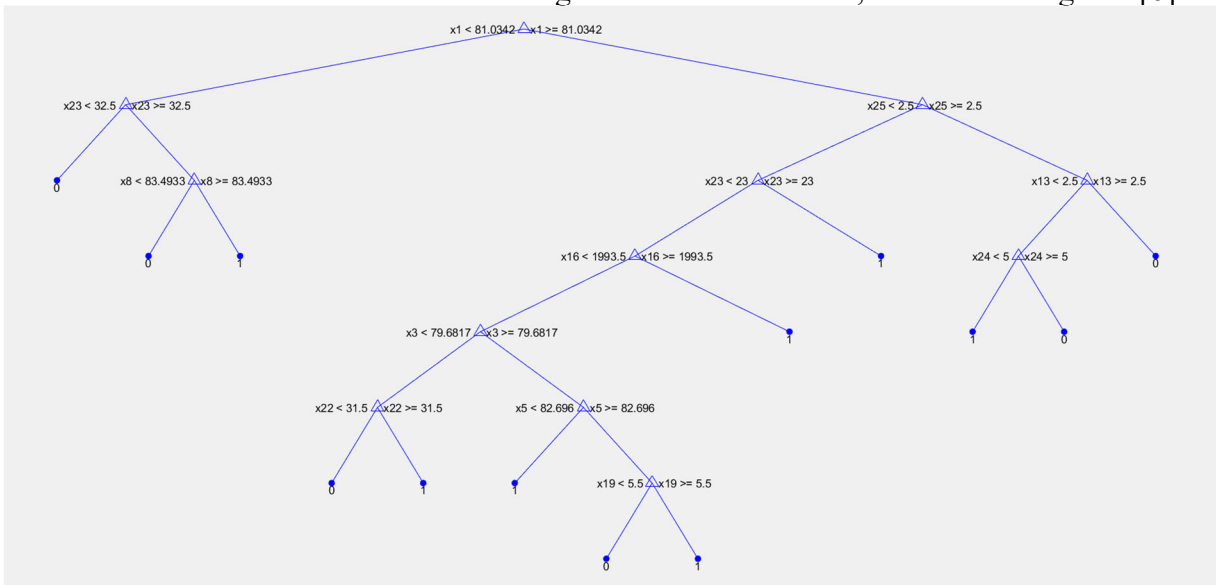


Figure 2 Initial model spanning tree

Optimization of decision tree

Optimization was carried out by changing the minimum sample number of leaf nodes, and the cross validation error of the above decision tree model under different minimum sample numbers of leaf nodes was cyclically tested, as shown in Figure 3[7].

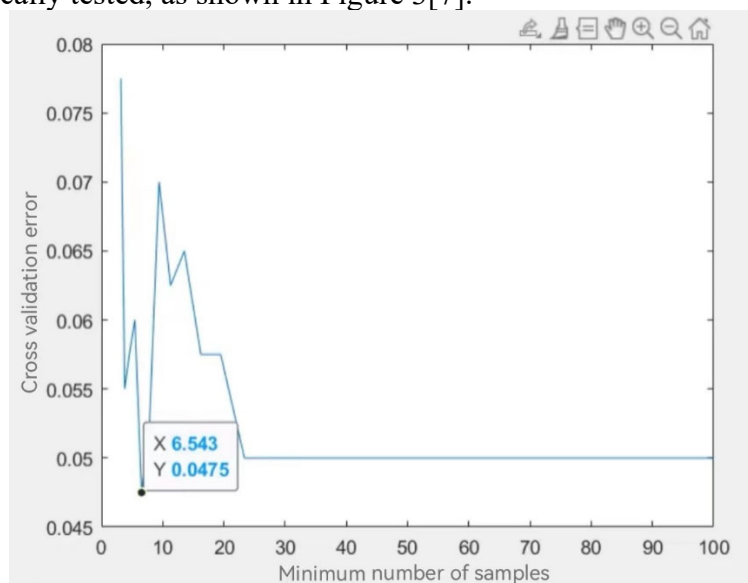


Figure 3 Cross validation error diagram

Obviously, when the minimum sample number of leaf nodes is about 7, the cross validation error can reach the minimum value[8]. By setting the minimum sample value of leaf nodes, we get the optimized decision tree as shown in Figure 4.

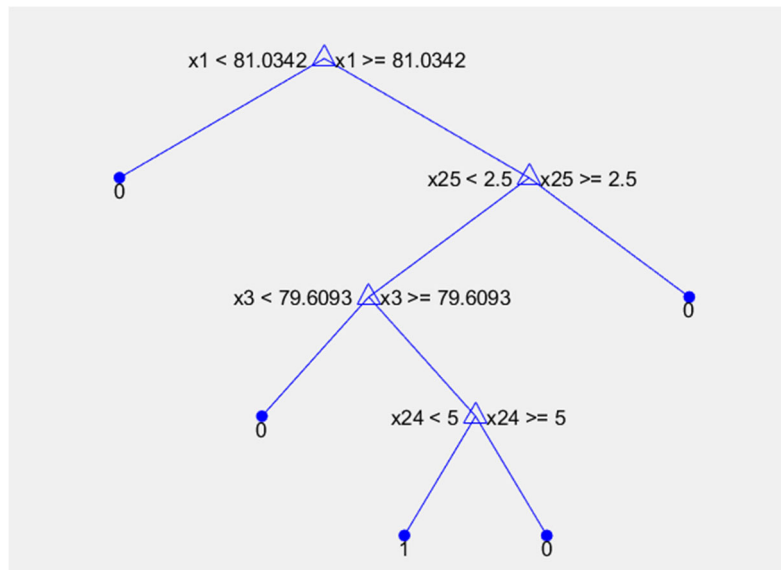


Figure 4 Optimized decision tree

Next, we need to carry out Pruning optimization of the initial decision tree. We use Pessimistic Pruning method[9], which is one of the post-pruning methods. It uses the training set to generate a decision tree and prunes with the training set without the need for an independent pruning set[10]. The basic idea of the pessimistic pruning method is that if the error rate can be reduced by replacing the original subtree with a leaf node, the leaf node should be replaced with the original subtree. The decision tree after pruning is shown in Figure 5.

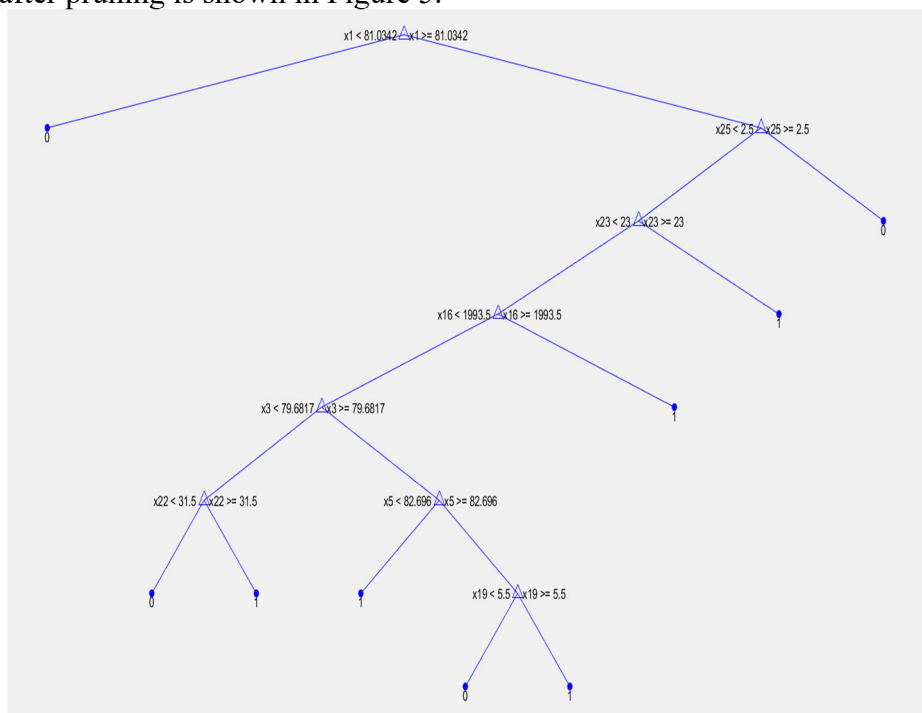


Figure 5 Decision tree after pruning

The cross validation errors and cross validation errors of the three decision trees are calculated respectively

The calculation table is shown in Table 2.

Table 2 Cross validation errors

Decision tree model	Cross validation error
Primitive decision tree	0.0675
Leaf node sample number optimization decision tree	0.0550
Prune the decision tree	0.0725

It can be seen from the above table that the cross validation error of the leaf node sample number optimization decision tree is the smallest, which indicates that its prediction accuracy is more accurate than the other two decision trees. Therefore, we finally select this decision tree as the decision tree model of the car. The model is as follows, and the pruned decision tree is shown in Figure 6.

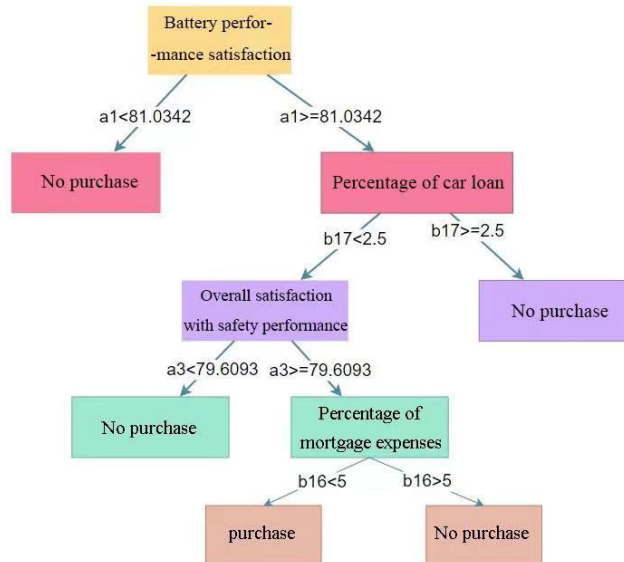


Figure 6 Decision tree after pruning

3.2 Analysis of the model

Users attach great importance to battery performance. If the battery durability index of this electric car is poor and charging is not convenient, users' satisfaction with A1 will be less than 81.0342, so they will not buy this car. Secondly, the percentage of car loan expenditure is also a major influencing factor. If the target user already has a car loan expenditure of more than 2.5, the user will not buy the electric car. The third factor is the user's overall satisfaction with the safety performance. If the user's satisfaction with the safety is less than 79.6093, he will not buy the car. When the above three conditions are met, the proportion of mortgage expenditure is also a factor to be considered. When the proportion of mortgage expenditure is greater than 5, the user is most likely not to buy the car.

To sum up, it can be analyzed from the decision-making effect of all kinds of satisfaction related to electric vehicle performance that target users attach great importance to the battery technical performance and safety performance of this electric vehicle. From the analysis of the user's own situation, the car loan is the main factor to decide whether they buy a car.

3.3 Test of model

Table 3 shows the model test table.

Table 3 Model test

Real results	buy (is)	TP 41	FN 15
	Don't buy (negative)	FP 6	TN 870
		Purchase (positive category)	Do not buy (negative category)
Predicted results			

The final classification accuracy of automobile models obtained from the evaluation index is 97.30%, and the proportion of correct prediction is 63.3% in the actual purchase samples, and 77.8% in the predicted purchase samples. The F score of 0.7 is close to 1, indicating that the classification effect is good.

4. Conclusion

In order to study the sales strategy of target customers of electric vehicles, this paper establishes the sales prediction model of electric vehicles. First, the initial decision tree model of customer mining of different brands of electric vehicles is constructed. The model was optimized by changing the minimum sample number of leaf nodes and pruning. At the same time, we do cross validation error analysis on three decision tree models, and finally get a stable and sensitive customer mining model. In the evaluation of the model, four indicators TP, FN, FP and TN in the confusion matrix were used to calculate the classification accuracy, recall rate, precision rate and F1 score, and the calculation results showed that the model had good classification and prediction effect. In subsequent studies, the correlation between different user habits and vehicle sales will be considered to provide better solutions for the optimization of electric vehicles.

References

- [1] Jiang Qiyuan, Xie Jinxing, YE Jun. *Mathematical Models* (4th Ed.). Beijing: Higher Education Press, 2011.
- [2] Guo Xianguang. How to Use SPSS software for Principal Component Analysis [J]. *Statistics and Information Forum*,1998(02):61-65.
- [3] Li M R. Analysis of sub-health status and influencing factors of college students based on Logistic and decision tree model [D]. Beijing Jiaotong University,2015.
- [4] Yang Hongbao, Gan Hongcheng. *Logistics engineering and management*,2017,39(01):115-118.
- [5] Luan Lihua, Ji Genlin. Research on Decision Tree Classification Technology [J]. *Computer Engineering*,2004(09):94-96+105.
- [6] He Chunai. Research on faW Precise marketing strategy under the background of big data [J]. *Modern Marketing (Business Edition)*,2020(11):170-171.
- [7] Tian Xiaohui. Research on marketing Strategy of New energy Vehicles [J]. *Time Automobile*,2020(20):98-99.
- [8] Larry P. Vellequette. The New Abnormal: Auto Sales Had Wild ride in 2021[J]. *Automotive News*,202,96(7020).
- [9] Ken Bredemeier. Biden Promotes Sale of Electric Vehicles[J]. *Voice of America News / FIND*,2021.
- [10] J.D. Power and LMC Automotive: Low Inventory Continues to Stall U.S. Auto Sales[J]. *Wireless News*,2021.