World Electric Vehicle Journal

MDPI

Article

Research on the Purchase Intention of Electric Vehicles Based on Customer Evaluation and Personal Information

Jian Chen [1], Zhenshuo Zhang [1], Chenyu Zhao [1], Shuai Zhang [1], Wenfei Guo [2], Cunhao Lu [1,*] and Xiaoguang Sun [3,4,*]

[1] School of Mechanical Engineering, Yangzhou University, Yangzhou 225127, China; jian.chen@yzu.edu.cn (J.C.); 221202232@stu.yzu.edu.cn (Z.Z.); 221202235@stu.yzu.edu.cn (C.Z.); 221204434@stu.yzu.edu.cn (S.Z.)
[2] Housing and Construction Bureau of Zhangdian District, Zibo 255000, China; wenfei_guo@163.com
[3] Guangdong International Cooperation Base of Science and Technology for GBA Smart Logistics, Jinan University, Zhuhai 519070, China
[4] School of Intelligent Systems Science and Engineering, Jinan University, Zhuhai 519070, China
* Correspondence: lch_ok@yzu.edu.cn (C.L.); xgsun@jnu.edu.cn (X.S.)

Abstract: With the continuous development of electric vehicle (EV) technology, there is an increasing need to analyze the factors influencing customers' purchase intentions. According to the data of customers' vehicle experience evaluation and personal information, this paper develops the analysis models of influencing factors using the analysis of variance algorithm (ANOVA) and Kruskal–Wallis algorithm. Then, the purchase intention model for EVs is proposed using the random forest method. Finally, the optimization model for the EV sales plan was built. The results show that the main factors influencing customers' purchases are different for different vehicle brands. However, the customer's evaluation of the vehicle experience has a greater influence on the customer's purchase. Compared to other prediction models, the random forest model has the highest accuracy. For 3 EV brands, the prediction accuracies are 97.8%, 98.9%, and 97.6%. In addition, this paper predicts the purchase intentions of 15 customers. By optimizing the sales plans for 3 EV brands, the predicted purchase rate of 15 customers increased from 40% to 53%. The research work contributes to the sales of electric vehicles, the accurate positioning of customers, and the identification of more potential customers.

Keywords: electric vehicles; vehicle purchase forecast; sales plan optimization; ANOVA algorithm; Kruskal–Wallis algorithm; random forest method; single objective linear programming model

check for updates

Citation: Chen, J.; Zhang, Z.; Zhao, C.; Zhang, S.; Guo, W.; Lu, C.; Sun, X. Research on the Purchase Intention of Electric Vehicles Based on Customer Evaluation and Personal Information. *World Electr. Veh. J.* **2024**, *15*, 9. https://doi.org/10.3390/wevj15010009

Academic Editors: Ayman EL-Refaie and Joeri Van Mierlo

Received: 8 November 2023
Revised: 15 December 2023
Accepted: 26 December 2023
Published: 27 December 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With increasing global environmental problems [1], people's awareness of environmental protection [2] is strengthened, and battery technology continues to advance. The new energy electric vehicles (EVs) have occupied part of the automotive industry market [3]. Compared with traditional fuel vehicles, EVs have the advantages of zero emissions, low noise, low maintenance costs, and intelligent functions [4]. How to further promote the adoption of EVs and increase sales of EVs is currently the focus of EV enterprises.

At present, there are two main policies to improve the popularity of EVs: First are government policies, such as emission regulations, regional restrictions, a higher fuel tax, and electric vehicle subsidies, which have been implemented in some cities [5]. Even some governments in developed countries plan to stop the sale of fuel vehicles in the next few years. Second is the new vehicle technology, such as driverless technology, vehicle networking technology, and intelligent cockpits [6,7]. Compared with fuel vehicles, these new technologies are more suitable for application in EVs, which can better meet the needs of modern people for vehicle functions. However, there are also some factors that affect the adoption of EVs, such as range anxiety, battery safety, the popularity of charging facilities, and so on [8].

In order to improve the adoption of EVs, many researchers have analyzed the influencing factors of EV sales and established many EV sales or purchase prediction models. Most studies focus on macro factors, such as national or regional policies, overall EV technology, people's average education level, average income, etc., and their sales or purchase models are only applicable to the whole EV industry [9]. However, there is little research on microfactors, such as customers' personal information and EV brands.

This paper analyzes the influence factors of EV purchase from a micro perspective. Based on the customer information and evaluations of three brands of EVs, the influence factor analysis model, predictive purchase model, and sales plan optimization model are established. The research work can be helpful for the sales of different brands of EVs. The innovations in this paper are as follows:

(1) Aimed at the micro-purchase factors of EVs, two influence factor analysis models are established, which determine the main factors influencing customers to purchase EVs from different brands.
(2) Based on analyzing the influencing factors, the predictive purchase model of EVs is established, which can quickly predict the purchase intention of customers.
(3) Based on the above research, the sales plan optimization model is established so as to propose a better sales plan.

The contents of the following chapters are as follows: Section 2 is a review of current EV purchase prediction research; Section 3 is the data processing of personal information and evaluation of customers of three EV brands; and Section 4 is the analysis of the influencing factors of EV purchase, the establishment of the purchase prediction model, and the optimization model of the sales plan. Section 5 is the result analysis, and Section 6 is the discussion.

2. Related Works

At present, the prediction research for EV purchases is mainly based on a questionnaire survey. The various influencing factors have been analyzed [10], and a variety of prediction models have been established. Descriptive statistics and principal component analysis were used to investigate the important factors of EV purchase intention in China, which showed that the intrinsic essence of the products and the cost were the most important, while the government policy was moderate [11]. The survey and statistical methods were applied to study the factors influencing the plug-in EV (PEV) purchase intention of adult drivers in the USA. They found that highly educated consumers were more interested in PEVs, and interest in PEVs is slightly higher than interest in EVs [12]. Based on an adapted stage model of self-regulated behavior change, a two-month follow-up survey was conducted among people interested in buying EVs. It was found that stage transitions were preceded by increases in goal intentions and implementation intentions in the week prior to the transition. Intent prediction was different for different people [13]. Following a survey in China and the US, consumer preferences for conventional vehicles, PEVs, and EVs were modeled. They found that EVs were less popular in the US than in China. Despite subsidies for EVs, US consumers prefer low-range PEVs to EVs. The adoption of EVs in China would be earlier than in the US [14]. A conceptual framework was developed to examine the influence of consumer innovativeness on EV preferences. The results indicated that adopters' innovativeness and attitudes toward the functional performance of EVs significantly affect EV preferences [15]. The structural equation model (SEM) was used to analyze the factors influencing EV purchase intentions in China. They pointed out that government policies have a great influence on EV purchase intentions [16]. The SEM was used to compare the effect of environmental performance, price value, and range confidence on consumers' EV purchase intentions. The results showed that EV environmental performance was a stronger predictor of attitude and thus purchase intention than price value and range confidence [17]. The univariate time series model and multivariate time series model were proposed to predict the sales volume of EVs in China. They focused on short-term (12 months) and long-term (60 months) predictions, which showed relatively high accuracy [5]. PEV and

EV purchase intentions in the US were studied, which showed that consumers preferred PEV to EV, mainly because of the range anxiety of EV [18]. A choice experiment was conducted to evaluate whether personal carbon trading (PCT) influences individual decisions to adopt EVs. The results showed that PCT can effectively promote the adoption of EVs. Government subsidies played a more important role in EV purchase intention than free parking and vehicle tax exemption [19]. A model of EV purchase intention was proposed, taking into account a number of demographic characteristics and attitudinal factors. It was found that attitudinal factors such as network externalities, government subsidies, vehicle performance, and demographic characteristics such as gender, age, and marital status have a significant impact on EV purchase intention [20]. A personality-perception-intention framework was proposed to study consumers' EV adoption behavior. It was shown that consumer perception and personality play an important role in EV purchase intention; personal innovativeness and environmental concern significantly also have a significant impact [21]. The correlation analysis and hierarchical multiple regression analyses were applied to study the socio-psychological effect on EV purchase intention. They found that low-carbon awareness has a slight moderating effect, while subjective norms and government policies have stronger effects [22]. An online questionnaire was conducted to investigate the variations and determinants of EV purchase intentions in three countries. The results showed that Chinese citizens were more willing to purchase EVs than citizens of Brazil and Russia, mainly because of social networks and government policies [9]. Using big data and text mining technologies, Chinese consumers' preferences for EVs were investigated through their online behaviors, which found that EV prices, car classification, and powertrain have a great influence on consumers' EV selection [23]. Based on the structural equation model (SEM), the purchase intention model and post-satisfaction model of EVs in Japan were proposed. They found that environmental awareness had a direct effect on purchase intention and a non-direct effect on post-purchase satisfaction [24]. Based on the theory of planned behavior (TPB), a questionnaire survey was conducted among potential consumers of EVs in Beijing. In addition, a structural equation model (SEM) was proposed to analyze the factors influencing EV purchase intention. The results showed that attitude, perceived behavioral control, cognitive status, product perception, and monetary incentive policy were important for consumers to purchase EVs [25]. The determinants of Chinese citizens' intentions to purchase EVs were studied in depth through an online survey. It was found that people with a wide social network and friends who already own EVs were more likely to purchase an EV. Age and education also had a limited effect [26]. The survey and statistical methods were used to investigate the changes in factors affecting PEV purchase intention over time. The results showed that the intention to purchase a PEV increased over time [27]. SEM was used to study customers' perceived value of EVs and found that the concept of "mianzi" has no significant effect on purchase intention, while the price factor has a direct effect [28]. The data mining method combined with deep earning technologies was used to study the purchasing reasons for EVs and found that EVs, demographic characteristics, and national policies were the main reasons [29]. A partial least squares structural equation model (PLS-SEM) of purchase intention for electric two-wheelers was built. It was found that perceived economic benefits were the most important factor, and women were more inclined to purchase electric two-wheelers than men in India [30]. The SEM was used to estimate the effect of positive anticipated emotion (PAE), negative anticipated emotion (NAE), and moral norms with TPB on EV purchase intention in China. The results showed that PAE has the greatest effect on EV purchase intention, followed by attitude, NAE, and perceived behavioral control (PBC) [31]. To analyze consumers' stated preferences, a rank-ordered logit model was constructed to provide 5-year fuel cost and total cost of ownership information on EV stated preferences. The results indicated that providing information can increase consumers' EV purchase intentions [32].

To summarize, the current research mainly analyzes a variety of EV purchase factors, including national and regional policies, people's cognition, EV prices, etc., and establishes

a variety of EV purchase and sales forecasting models. The current research has macro-guiding significance for government policy formulation and EV industry development. However, there is little research on customer personal factors and EV brand factors. This kind of analysis can help EV companies accurately target potential customers, provide accurate services, and optimize sales plans. In this study, customer personal factors and EV brand factors are analyzed, and the influence factor analysis model, predictive purchase model, and sales plan optimization model are established.

3. Data Processing

In this paper, the research data come from three EV brands in China [33]: Brand 1 is a joint venture brand; Brand 2 is an independent brand; and Brand 3 is a new brand. The total data include data A (customer experience score data) and data B (customer personal information data), as shown in Table 1.

Table 1. Implications of customer data indexes.

Data A		Data B			
Index	**Implication**	**Index**	**Implication**	**Index**	**Implication**
A_1	Battery performance	B_1	Residence	B_{10}	Working years
A_2	Comfort	B_2	City residency time	B_{11}	Kind of working unit
A_3	Economy	B_3	Residence zone	B_{12}	Job title
A_4	Security	B_4	Driving years	B_{13}	Household income
A_5	Power	B_5	Number of family members	B_{14}	Personal income
A_6	Vehicle handling	B_6	Marital status	B_{15}	Household disposable annual income
A_7	Exterior and interior	B_7	Number of children	B_{16}	Annual proportion of house loan
A_8	Configuration and quality	B_8	Year of birth	B_{17}	Annual proportion of car loans
\	\	B_9	Educational background	\	\

Due to the presence of anomalous data, the original data must be processed, the anomalous data must be deleted, and the blank data must be filled in.

3.1. Customer Experience Score Data (Data A)

Data A is customer experience scores from 1600 customers for 3 EV brands. The lowest score is 0, and the highest is 100. After examining data A, it was determined that there are some scores given by customers that are inconsistent with the scoring interval. In this paper, the box plot method is used to examine data A and eliminate unreasonable data. The results are shown in Figure 1.

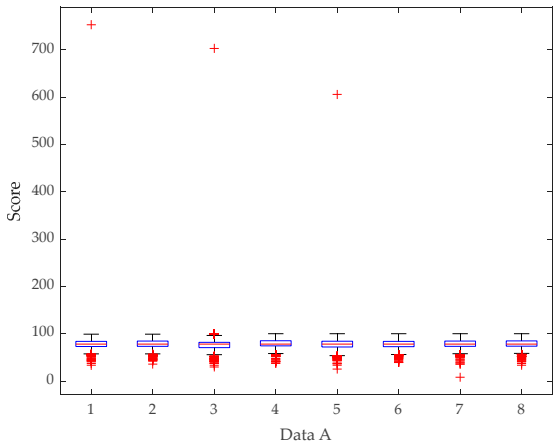


Figure 1. Data A filtered by the box plot method. Red "+" in the figure represents the satisfaction scores in data A.

In Figure 1, the abnormal data are found in columns A_1, A_3, A_5, and A_7. Then, data greater than 100 or less than 0 were replaced by linear interpolation, y_{imod}:

$$y_{imod} = y_{i-1} + \frac{(x_i - x_{i-1})(y_{i+1} - y_{i-1})}{x_{i+1} - x_{1-1}} \quad (1)$$

where x_i is the corresponding serial number according to data A_i and is arranged in ascending order, and y_i is the abnormal score corresponding to x_i. After correcting the data of all customers, relatively correct data can be obtained.

3.2. Customer Personal Information Data (Data B)

Data *B* has some missing data and logical errors; therefore, a logical analysis is carried out. Data groups with logical relationships in data *B* include groups B_4 and B_8, groups B_{13}, B_{15}, B_{16}, and B_{17}, and groups B_5, B_6, and B_7, where b_i is the value of the *i*-th index in data B.

For groups B_4 and B_8, the customer cannot obtain a driver's license until adulthood. Therefore, b_4 driving years should be less than or equal to the difference between the current time and the time from the b_8 birth year to adulthood:

$$b_4 \leq 2023 - (b_8 + 20) \quad (2)$$

For groups B_{13}, B_{15}, B_{16}, and B_{17}, the criterion is that income must be greater than or equal to expenses:

$$b_{13} \geq b_{15} + \frac{b_{16} + b_{17}}{100} \times b_{13} \quad (3)$$

b_{15} does not satisfy Equation (3) and must be excluded. The excluded data are filled in using the family annual income rate and the individual annual income rate calculated from the normal data:

$$K = \frac{b_{13}}{b_{14}} \quad (4)$$

where b_{13} is the total family annual income and b_{14} is the total individual annual income. K is the magnification ratio corresponding to b_{13} data and b_{14} data. By using the ergodic cycle method, the average rate $\overline{K}$ of all normal data is solved as follows:

$$\overline{K} = \frac{1}{m} \sum_{\lambda=1}^{m} K \quad (5)$$

where m is the number of the total normal data. Combining $\overline{K}$, personal income and total family income, the abnormal value of b_{15} that needs to be replaced is calculated:

$$b_{15} = \frac{b_{13}(\overline{K} - 1)(b_{16} + b_{17})}{100} \quad (6)$$

For groups B_5, B_6, and B_7, definitions of their values are shown in Table 2.

Table 2. Data definitions for B_5, B_6, and B_7.

Index	Value	Implication
B_5	1, 2, 3...	Family size
B_6	1	Unmarried, living alone
	2	Unmarried, living with parents
	3	Married, no children, not living with parents
	4	Married, no children, living with parents
	5	Married, children, not living with parents
	6	Married, children, not living with parents
	7	Divorced/widowed
	8	Others
B_7	0, 1, 2, 3...	Number of children

Provided that the customer does not become pregnant out of wedlock, set the logical relationship between family size (B_5), marital status (B_6), and the number of children (B_7) as a criterion; B_7 is set as a variable. Combined with Table 2, the values of B_5, B_6, and B_7 are adjusted, as shown in Table 3.

Table 3. Logical relationships of B_5, B_6, and B_7.

Raw Data			Processing
B_7	B_6	B_5	
1	≤4	Arbitrary	Change B_7 to 0
	5	3	Normal
	6	≤3	Change B_5 to 5
		$4 < B_5 < 9$	Normal
2	<4	Arbitrary	Change B_7 to 0
	5	4	Normal
	6	≤4	Change B_5 to 6
		$4 < B_5 < 7$	Normal
3	5	<5	Change B_5 to 5
		5	Normal
	6	6	Normal
Not filled	Arbitrary	Arbitrary	Change B_7 to 0

4. Methods

After the customer data processing is completed, the influencing factors of the EV purchase are analyzed to determine the important influencing factors. Then, based on the important factors, the probability prediction model of the customer's EV purchase is established. Finally, the sales plan optimization model is established. The flow chart of these models is shown in Figure 2.

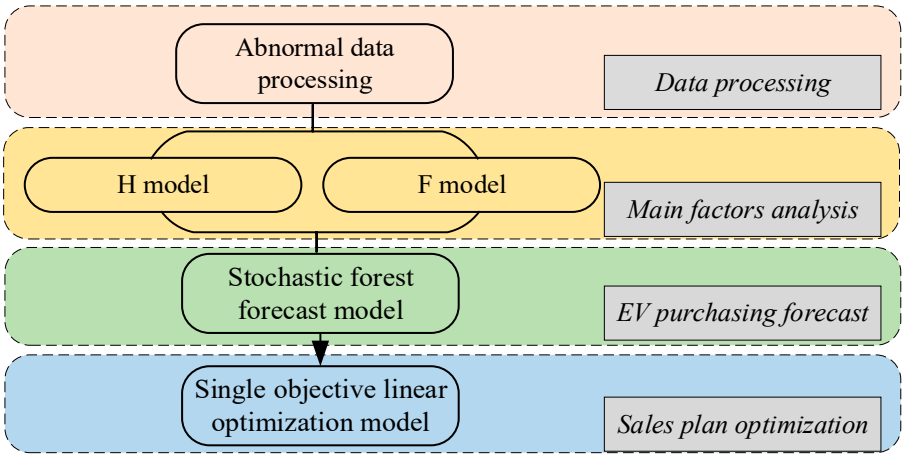


Figure 2. Flow chart of the methods.

In Figure 2, the analysis of variance (ANOVA) algorithm (F model) and the Kruskal–Wallis algorithm (H model) are used to analyze the influencing factors, and the important factors for the purchase of EVs by each brand are determined. The EV purchase prediction model adopts the random forest method, which uses 70% of the data for training and 40% of the data for verification. Based on the purchase prediction model, the sales plan optimization model is established, which can provide better sales plans for different EV brands.

4.1. ANOVA Analysis Model

Based on the ANOVA algorithm [34], the influencing factors analysis model (F model) is established:

$$F_r = \frac{MSE}{MSB} \tag{7}$$

where F_r is the ratio of intra-group and inter-group variances, and MSB is the mean square error between groups, that is, the sum of squares between groups divided by the degrees of freedom between groups. MSE is the mean square error within a group, which is the sum of squares divided by the group degree.

$$\begin{cases} MSB = \frac{SSB}{dfB} \\ MSE = \frac{SSE}{dfE} \end{cases} \tag{8}$$

where dfB is the number of inter-group degrees of freedom, i.e., the number of constraints on inter-group independence, calculated as the total number of groups minus 1. dfE is the number of constraints for the error within the group, i.e., the total number of observations minus the number of groups.

4.2. Kruskal–Wallis Analysis Model

In addition to the above model, this paper also establishes the Kruskal–Wallis analysis model [35] (H model). Assuming that there are k independent score data, the total number of samples N is as follows:

$$N = \sum_{q=1}^{k} n_q \tag{9}$$

where n_q is the size of the samples in group q. The sum of the collected data is arranged in ascending order, and statistic H_r is used to compare whether the population median of different samples is significantly different:

$$H_r = \frac{12}{N(N+1)} \sum_{q=1}^{k} \frac{{R_q}^2}{n_q} - 3(N+1) \tag{10}$$

where q is the number of samples and R_q is the rank sum of the n_q observations of group q ($X_1, \ldots, X_q$).

4.3. Random Forecast Purchase Forecast Model

Based on the above analysis results, the random forest method [36] is adopted to establish a car purchase prediction model, and the algorithm flow is shown in Figure 3.

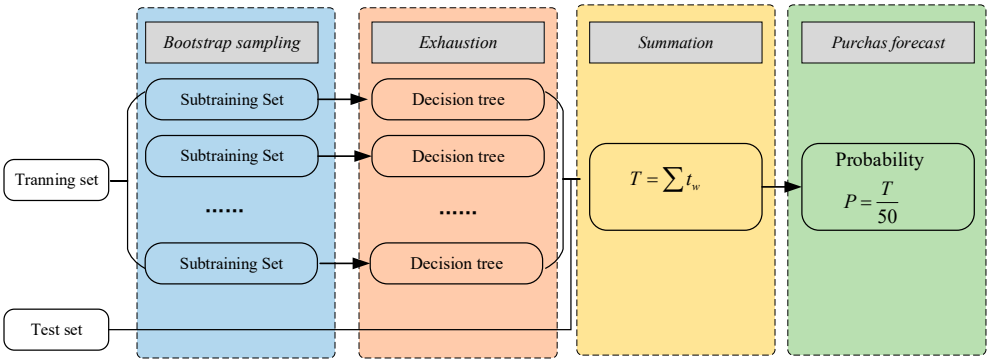


Figure 3. Flow chart of the random forest algorithm.

Taking 70% of the major factor data selected above as the training set and 30% as the test set, the bootstrap sampling model of the random forest model is established. The specific steps are as follows: Assume the set S contains u different samples (x_1, x_2, . . ., x_u). If a sample is taken from S and put back a total of u times to form a new set S^*, then the probability of S^* not containing sample x_s (s = 1, 2, . . ., u) is as follows:

$$P_{sample} = (1 - \frac{1}{u})^u \tag{11}$$

When $u \to \infty$, there is, as follows:

$$\lim_{u\to\infty} P_{sample} = \lim_{u\to\infty} (1 - \frac{1}{u})^u = e^{-1} \approx 0.368 \tag{12}$$

Although the total number of samples in S^* is equal to the total number of samples in S, S^* may contain duplicate samples. When duplicate samples are removed, S^* contains only about 63.2% of the samples in S. Then, using the exhaustive method [37], all values of each feature in the sample are traversed to find the best segmentation variable and segmentation point. The quality of the segmentation variables and segmentation points is measured by the impurity of the nodes after segmentation, which is the weighted sum $G(r_o, t_{cd})$ of the impurity of the child node:

$$G(r_o, t_{od}) = \frac{\alpha_l}{N_s} y'_{bag}(x_l) + \frac{\alpha_r}{N_s} y'_{bag}(x_r) \tag{13}$$

$$y'_{bag}(x_e) = \frac{1}{M} \sum_{h=1}^{N} y'^{f}(x) \tag{14}$$

where r_o is the shred variable, t_{cd} is the value of r_o, $y'_{bag}(x_e)$ is the impurity of the node, and α_l, α_r, and N_s are the number of training samples of the left sub-node after segmentation, the number of training samples of the right sub-node, and the number of all training samples of the current node, respectively. x_l and x_r are the sets of training samples of the left and right sub-nodes, and M is the number of decision trees. After calculating the best segmentation variables and segmentation points in each decision tree, the training set and test set can be brought into the model for calculation. Finally, a prediction of the purchase can be obtained.

For ease of description, $f(y_j)$ is used to represent the entire correlation between the input and output results of the random forest method. Therefore, the probability that customers will purchase electric vehicles can be predicted as follows:

$$P_{forecast} = f(y_j) \tag{15}$$

4.4. Single Objective Linear Optimization Model of the Sales Plan

Based on the above analysis, customer experience score data (data A) has a significant impact on customer purchase intention. Usually, the service cost of EV sales companies is directly related to data A. Increasing the service cost can improve scores in data A, thereby increasing customer purchase intention. However, it is costly and difficult for automobile sales companies to comprehensively improve all data A scores. Therefore, based on the above prediction model and existing data, an optimization model for short-term sales plans is proposed. The goal is to slightly improve the scores of some items in data A, i.e., to increase the customer's purchase intention with less service cost. For sake of simplicity, some assumptions are made as follows.

(1) Assuming that the customer experience score data are equivalent to the service cost of the automobile sales company, increasing the customer experience score data score means increasing the service cost.

(2) Considering the short-term improvement, it is assumed that the scores of data A increase in integer form, with a maximum increase of 5 for each item and a maximum score of 100.

Based on the predicted purchase model and the above assumptions, a single objective optimization model is established with the goal of minimizing the increase in data A [38]. This model is constrained by a customer purchase probability greater than 0.5. By investing less in service costs, more purchase customers can be attracted.

(1) Objective function

Since the improvement of the service score is proportional to its difficulty, it is necessary to find the point where the increase in the score is minimum and the purchase probability is maximum until the predicted purchase is achieved. To reduce the service difficulty as much as possible, the satisfaction score should increase as little as possible, so the objective equation is as follows:

$$\min g = \sum_{j=1}^{8} v_j \tag{16}$$

where v_j is the increased service score; j is [1,8].

(2) Constraint condition

Since the limitation of the satisfaction score is 100, all increased satisfaction scores should be less than 100:

$$y_j + v_j \leq 100 \tag{17}$$

where y_j is the initial satisfaction score.

$$\begin{cases} v_j \leq 5 \\ v_j \in \text{data A} \\ v_j \in N_a \end{cases} \tag{18}$$

where v_j is the j-th satisfaction score of brand i. For the sake of optimization, v_j is specified as an integer. v_j can only be increased in the satisfaction score of data A, which means that v_j belongs to data A. Moreover, the maximum v_j is set to 5.

According to Equation (15), the probability of customers purchasing electric vehicles is obtained as follows:

$$P_{forecast} = f(y_j + v_j) \tag{19}$$

At the same time, if $P_{forecast}$ is greater than 0.5, the customer will buy the vehicle, and if $P_{forecast}$ is less than 0.5, the customer will not buy, so the constraint equation is as follows:

$$P_{forecast} \geq 0.5 \tag{20}$$

Finally, the final optimization model is as follows:

$$\min g = \sum_{j=1}^{8} v_j$$

Subject to:

$$\begin{cases} v_j \leq 5 \\ y_j + v_j \leq 100 \\ v_j \in \text{data A} \\ P_{forecast} = f(y_j + v_j) \\ P_{forecast} \geq 0.5 \end{cases} \tag{21}$$

5. Results and Discussion

According to the above models, the influencing factors of EVs purchase are analyzed, customer purchase intention is predicted, and the sales plan is optimized.

5.1. Analysis of Influencing Factors

When the F model is used, the tested data should conform to a normal distribution. Therefore, the built-in normplot function of MATLAB is used to compare the data groups, as shown in Figure 4. The results show that all the data groups used in this paper are closed with the standard plus distribution, which means that all the data groups conform to the plus distribution.

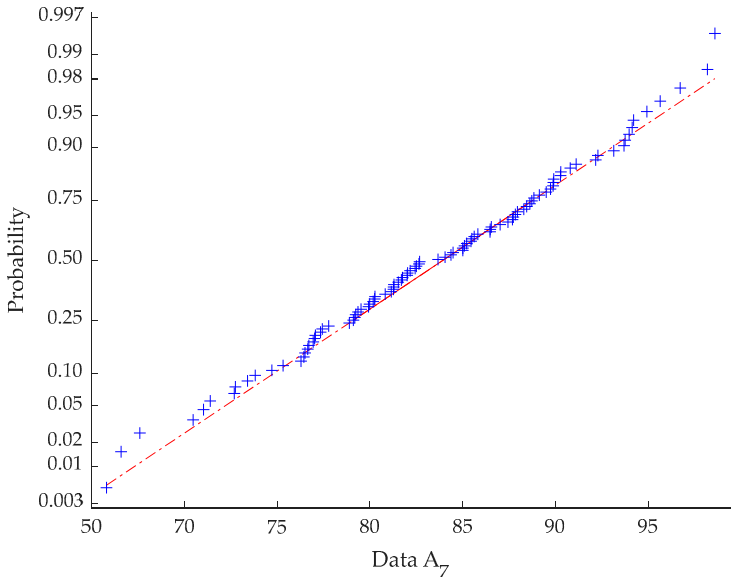
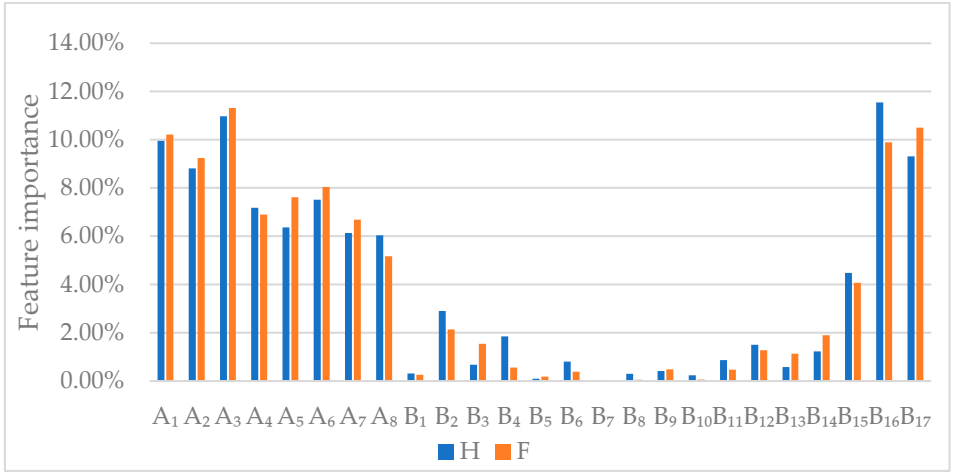


Figure 4. Data group 25 normal distribution probability graph. Blue "+" in the figure represent the probability of values in data A_7. The dotted red line reflects the degree to which the data conform to a normal distribution.

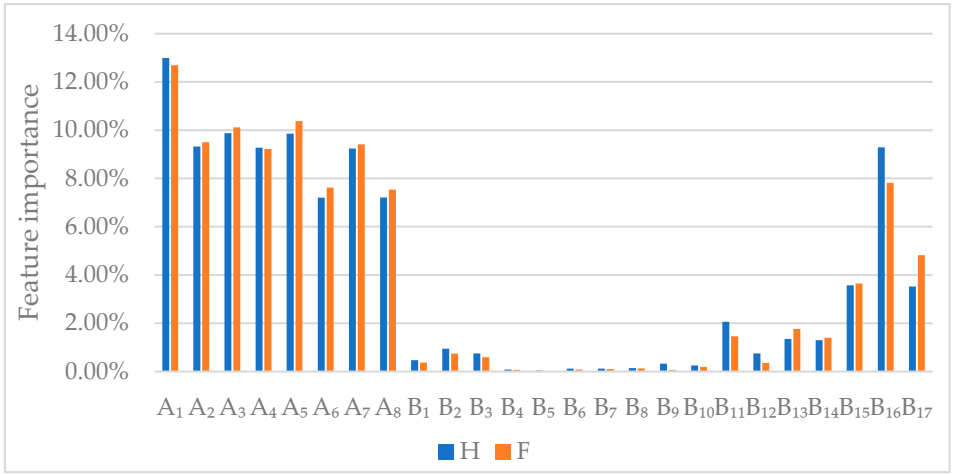
The F and H models were used to analyze the influence factor of the customer data for the three brands, and the results are shown in Figure 5.

As shown in Figure 5, the analysis results of the influencing factors of the F and H models are basically consistent and mutually verified. Overall, for the three brands, data A has a greater influence, while data B has a lesser influence, which means that customer experience score has a great influence on EV purchase. In addition, B_5, B_7, and B_8 (family size, number of children, and year of birth) have little impact. However, there are certain differences in the influence factors of different brands, among which B_{16} and B_{17} have a greater influence on the purchase of Brand 1 and Brand 2, and a relatively small influence on Brand 3. For further analysis, based on the results of the F model with 4% as the benchmark, the main influencing factors are screened out according to the proportion shown in Table 4.

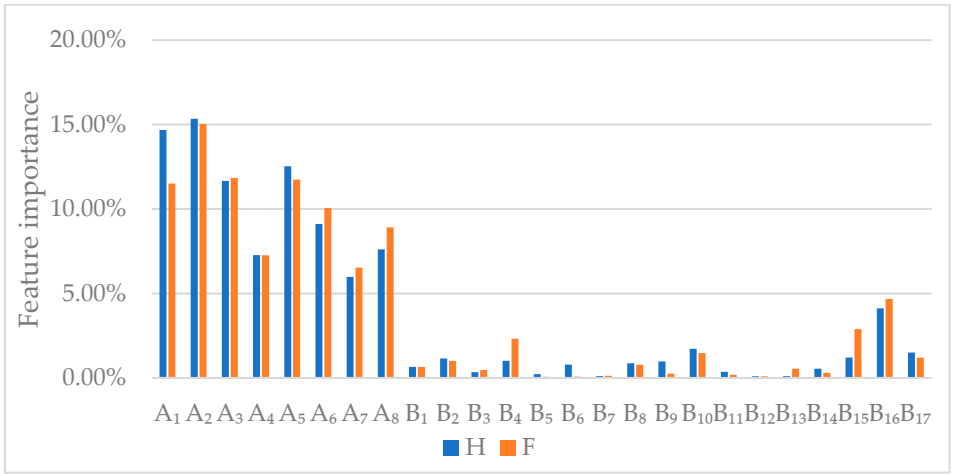
It can be seen from Table 4 that the scores of battery technical performance, comfort, economy, safety performance, and power performance in data A have a great influence on the purchase of three brands. For Brand 1 (joint venture brand), customers are more concerned about the vehicle economy (A_3); for Brand 2 (independent brand), customers are more concerned about the battery technical performance (A_1); for Brand 3 (new power brand), customers are more concerned about the vehicle comfort performance (A_2). In addition, car loans and mortgage loans have more influence on Brand 1 and less influence on Brand 3.



(**a**)



(**b**)



(**c**)

Figure 5. Analysis of influencing factors for (**a**) Brand 1; (**b**) Brand 2; and (**c**) Brand 3.

Table 4. Main influencing factors of car purchases.

Brand 1	Brand 2	Brand 3
A_3	A_1	A_2
B_{17}	A_5	A_3
A_1	A_3	A_5
B_{16}	A_2	A_1
A_2	A_4	A_6
A_6	A_7	A_8
A_5	B_{16}	A_4
A_4	A_6	A_7
A_7	A_8	B_{16}
A_8	B_{17}	
B_{15}		

5.2. EV Purchase Prediction

The main influencing factors screened above are input into the purchase prediction model, in which 70% of the data is used as the training set and 30% of the data is used as the test set. For the results, the probability of [0, 0.5) means "No", and the probability of [0.5, 1] means "Buy". Since the number of decision trees in the model is different, the prediction error will be different, as shown in Figure 6. It can be seen that the overall error decreases as the number of decisions increases. When the number of decision trees is greater than 40, the change tends to be smooth. In this paper, the decision tree number of the model is set to 50 with an error of 0.04. The prediction accuracy of the training set and the test set is shown in Table 5. It can be seen that the accuracy of the purchase prediction model is high. The lowest accuracy of the training set of three brands is 98.9%; the highest is 99.4%. The accuracy of the test set is 97.8%, 98.9%, and 97.6%, respectively. The accuracy of the purchase model is related to the amount of data. Brand 2 has the most data and Brand 1 has the least, so the accuracy of Brand 2 is higher than the other two brands. Overall, the maximum error of the predictive purchase model of three brands is not more than 2.4%, which meets the research requirements.

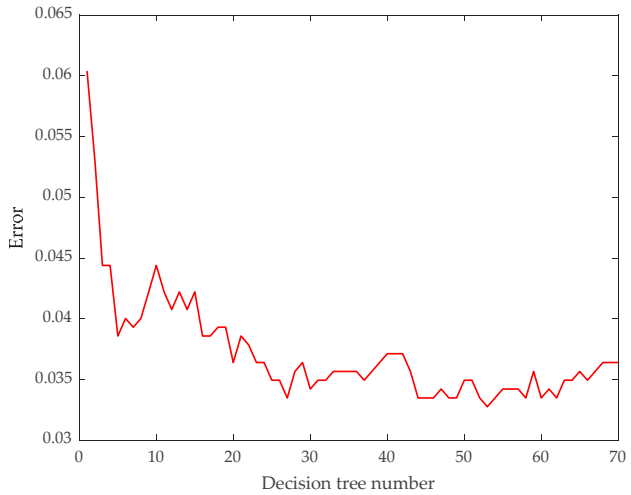


Figure 6. Influence of decision tree number on error.

Table 5. Prediction accuracy.

	Training Set Accuracy	Test Set Accuracy	Maximum Number of Splits	Number of Learners
Brand 1	99.40%	97.80%	389	50
Brand 2	99.70%	98.90%	1272	50
Brand 3	98.90%	97.60%	136	50

In order to further verify the random forest purchase prediction model, this paper simultaneously compares the decision tree prediction model, the BP neural network prediction model, the SVM prediction model, and the KNN prediction model. Take Brand 1 as the research object; the comparison of model results is shown in Figure 7. As shown in the figure, the random forest prediction model has the highest accuracy compared to the other models.

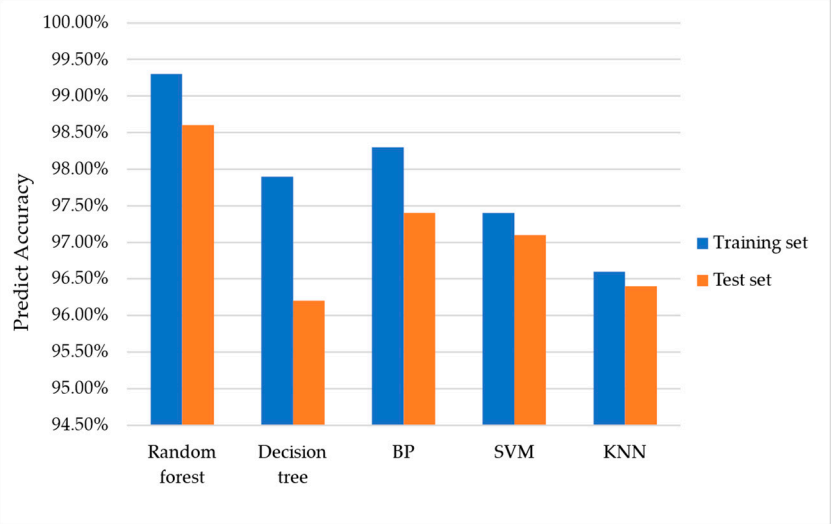


Figure 7. Comparison of the results of different prediction models.

In this paper, the established random forest prediction model is used to simulate 15 customer data intentions (5 customers for each brand). The results are shown in Table 6. Among the 15 customers, 6 are predicted to buy and 9 are predicted not to buy. For Brand 1, customers 1 and 2 will buy; for Brand 2, customers 6 and 7 will buy; and for Brand 3, customers 11 and 12 will buy.

Table 6. Prediction results of customers purchasing EVs.

Customer ID	Brand	$P_{forecast}$	Forecast
1	1	0.840	Buy
2	1	0.510	Buy
3	1	0.000	No
4	1	0.285	No
5	1	0.060	No
6	2	0.660	Buy
7	2	0.520	Buy
8	2	0.000	No
9	2	0.000	No
10	2	0.272	No
11	3	0.520	Buy
12	3	0.920	Buy
13	3	0.000	No
14	3	0.000	No
15	3	0.459	No

5.3. Sales Plan Optimization

For the above 9 customers who would not buy EVs, Brand 1 is customers 3, 4, and 5; Brand 2 is customers 8, 9, and 10; and Brand 3 is customers 13, 14, and 15. The total customer purchase rate (purchase rate = number of purchasing customers/total number of customers) is 40%. Based on the sales strategy optimization model, some scores of data A for three brands are slightly increased. The optimized plan is shown in Table 7. For Brand

1, the scores of A1, A3, and A5 in data A are increased by 1, 3, and 1, respectively, with a total increase of 5; for Brand 2, A1 and A3 are improved by 3 and 1, respectively, with a total improvement of 4; and for Brand 3, A2 is increased by 5, with a total increase of 5. It can be seen that different brands have different optimization plans.

Table 7. Optimized sales plan.

Brand	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	g
1	1	0	3	0	1	0	0	0	5
2	3	0	1	0	0	0	0	0	4
3	0	5	0	0	0	0	0	0	5

According to the optimization plan, the customer's purchase probability is recalculated, and the results are shown in Table 8. It can be seen that after the optimization plan, the purchase probability of customers has increased, with the purchase probability of customer 10 increasing from 0.275 to 0.52 and that of customer 15 increasing from 0.459 to 0.52. The total number of customers purchasing has increased from 6 to 8, with a purchase rate of 53%, an increase of 13% compared to before.

Table 8. Comparison of customer purchase probability before and after optimization.

	Customer ID	3	4	5	8	9	10	13	14	15
$P_{forecast}$	Before optimization	0	0.285	0.06	0	0	0.272	0	0	0.459
	After optimization	0.16	0.4	0.133	0.2	0.14	0.52	0.05	0.32	0.52

Based on the above research, it can be found that different electric vehicle companies have different focuses in the sales process. If key points of the sales process can be found, it will be beneficial to attract more potential customers, provide customers with more detailed and accurate services, improve customer satisfaction, and thereby increase car sales. For example, when targeting the sales of electric vehicles for Brand 1, according to Table 6, when focusing on improving the scores of A_1, A_3, and A_5, focus on introducing the battery performance, economy, and power of the company's electric vehicles to customers during the sales process; this will significantly increase customer purchase intention and satisfaction. Similarly, for the sales of electric vehicles under Brand 2, if the focus is on introducing the battery performance and economic viability of the company's electric vehicles to customers during the sales process, it will significantly increase their purchase intention and satisfaction. And, for the sales of electric vehicles under the third brand, during the sales process, the focus should be on introducing the comfort advantages of the company's electric vehicles to customers; this will significantly increase customer purchase intention and satisfaction. In summary, for the above three electric vehicle brands or other electric vehicle companies, using the model established in this paper for sales strategy planning during the sales process can, to some extent, increase the number of electric vehicle customers and sales volume.

6. Conclusions

Based on the customer experience score and personal information data, this paper establishes a whole process of data processing—influencing factor analysis, customer purchase prediction, and sales plan optimization—to further reveal the relationship between EVs and customers. Specific conclusions are as follows:

(1) According to the logical relationships, the original data are processed, and the abnormal data are eliminated. The missing data are filled in by linear interpolation, average magnification, and other methods, and the reliability of the data is greatly improved.
(2) The F and H models are used to analyze the influencing factors in the data. It was found that the results of the two test methods are consistent and mutually verified. For

different EV brands, the main influencing factors of customer purchase are different, but on the whole, the customer experience score has a greater impact on EV purchase.

(3) Based on the main influencing factors tested above, a customer purchase prediction model is established using the random forest method. The results show that the model is more accurate than other purchase prediction models, with the accuracy of all three brands exceeding 97%. In addition, 15 customer purchase intentions are predicted, and 6 customers will buy EVs.

(4) Based on the customer prediction model, the single objective linear optimization model of the sales plan is established. The results showed that, with the optimized sales plan, the predicted purchase rate of test customers increased from 40% to 53%.

The model established in this paper uses F-tests and H-tests to extract the important factors affecting automobile sales, and a random forest model is established to predict customer purchases. At the same time, according to the prediction model, the optimization model of the sales plan is established. It can predict and optimize automobile sales from a more intelligent aspect and provide a more scientific sales model for automobile sales. The research work in this paper can theoretically help EV companies attract more potential customers and provide accurate services.

However, there may be certain limitations to the data used in this paper. For example, the uneven distribution of the sales data from different electric vehicle companies collected in this paper manifests this specific unevenness. The amount of data collected between each company is uneven, with sufficient data for Brands 1 and 2 and poor data for Brand 3. The distribution of purchase situations among different customers is uneven, with the majority of customers not making purchases and only a small number of customers making purchases, resulting in a significant difference in the amount of purchase data and non-purchase data. In some cases, this may have a certain impact on the prediction accuracy. At the same time, the model used in this paper can be extended in the future; the prediction model can be applied to the sales of different products; for example, housing, electronic products, and other products can use the model established in this paper to find the best sales strategy. In the future, more rich data will be collected, and the model established in this paper will be improved with higher accuracy.

Author Contributions: Conceptualization, J.C.; methodology, Z.Z.; software, C.Z.; validation, S.Z.; formal analysis, C.L.; investigation, X.S.; resources, W.G.; data curation, X.S.; writing—original draft preparation, Z.Z.; writing—review and editing, J.C.; visualization, C.Z.; supervision, S.Z.; project administration, J.C.; funding acquisition, J.C. and X.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant Number: 52005435), the Natural Science Foundation of Jiangsu Province (Grant Number: BK20190873), the Undergraduate Education Reform Project of Yangzhou University (Special Funding for Mathematical Contest in Modeling) (grant number: xkjs2023049), and the Shuangchuang Program of Jiangsu Province of China (Grant Number: (2019)30756).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Wenfei Guo is an employee of the Housing and Construction Bureau of Zhangdian District. Xiaoguang Sun is an employee of Guangdong International Cooperation Base of Science and Technology for GBA Smart Logistics. The paper reflects the views of the scientists and not the company. The authors declare no conflicts of interest.

References

1. Hartig, T.; Kahn, P.H. Living in Cities, Naturally. *Science* **2016**, *352*, 938–940. [CrossRef] [PubMed]
2. O'Brien, K. Political Agency: The Key to Tackling Climate Change. *Science* **2015**, *350*, 1170–1171. [CrossRef] [PubMed]

3. Brown, S.; Pyke, D.; Steenhof, P. Electric Vehicles: The Role and Importance of Standards in an Emerging Market. *Energy Policy* **2010**, *38*, 3797–3806. [CrossRef]
4. Yang, L.-F.; Xu, J.-H.; Neuhäusler, P. Electric Vehicle Technology in China: An Exploratory Patent Analysis. *World Pat. Inf.* **2013**, *35*, 305–312. [CrossRef]
5. Zhang, Y.; Zhong, M.; Geng, N.; Jiang, Y. Forecasting Electric Vehicles Sales with Univariate and Multivariate Time Series Models: The Case of China. *PLoS ONE* **2017**, *12*, e0176729. [CrossRef] [PubMed]
6. Tang, M.; Zhuang, W.; Li, B.; Liu, H.; Song, Z.; Yin, G. Energy-Optimal Routing for Electric Vehicles Using Deep Re-inforcement Learning with Transformer. *Appl. Energy* **2023**, *350*, 121711. [CrossRef]
7. Moy, K.; Ganapathi, D.; Geslin, A.; Chueh, W.; Onori, S. Synthetic Duty Cycles from Real-World Autonomous Electric Vehicle Driving. *Cell Rep. Phys. Sci.* **2023**, *4*, 101536. [CrossRef]
8. Bigerna, S.; Micheli, S. Attitudes toward Electric Vehicles: The Case of Perugia Using a Fuzzy Set Analysis. *Sustainability* **2018**, *10*, 3999. [CrossRef]
9. Habich-Sobiegalla, S.; Kostka, G.; Anzinger, N. Electric Vehicle Purchase Intentions of Chinese, Russian and Brazilian Citizens: An International Comparative Study. *J. Clean. Prod.* **2018**, *205*, 188–200. [CrossRef]
10. Higueras-Castillo, E.; Guillén, A.; Herrera, L.-J.; Liébana-Cabanillas, F. Adoption of Electric Vehicles: Which Factors are Really Important? *Int. J. Sustain. Transp.* **2020**, *15*, 799–813. [CrossRef]
11. Wang, Z.; Wang, C.; Hao, Y. Influencing Factors of Private Purchasing Intentions of New Energy Vehicles in China. *J. Renew. Sustain. Energy* **2013**, *5*, 063133. [CrossRef]
12. Carley, S.; Krause, R.M.; Lane, B.W.; Graham, J.D. Intent to Purchase a Plug-in Electric Vehicle: A Survey of Early Im-pressions in Large Us Cites. *Transp. Res. Part D Transp. Environ.* **2013**, *18*, 39–45. [CrossRef]
13. Klöckner, C.A. The Dynamics of Purchasing an Electric Vehicle–a Prospective Longitudinal Study of the Decision-Making Process. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *24*, 103–116. [CrossRef]
14. Helveston, J.P.; Liu, Y.; Feit, E.M.; Fuchs, E.; Klampfl, E.; Michalek, J.J. Will Subsidies Drive Electric Vehicle Adoption? MEASURING Consumer Preferences in the U.S. and China. *Transp. Res. Part A Policy Pract.* **2015**, *73*, 96–112. [CrossRef]
15. Morton, C.; Anable, J.; Nelson, J.D. Exploring Consumer Preferences Towards Electric Vehicles: The Influence of Consumer Innovativeness. *Res. Transp. Bus. Manag.* **2016**, *18*, 18–28. [CrossRef]
16. Wang, Z.; Zhao, C.; Yin, J.; Zhang, B. Purchasing Intentions of Chinese Citizens on New Energy Vehicles: How Should One Respond to Current Preferential Policy? *J. Clean. Prod.* **2017**, *161*, 1000–1010. [CrossRef]
17. Degirmenci, K.; Breitner, M.H. Consumer Purchase Intentions for Electric Vehicles: Is Green more Important than Price and Range? *Transp. Res. Part D Transp. Environ.* **2017**, *51*, 250–260. [CrossRef]
18. Lane, B.W.; Dumortier, J.; Carley, S.; Siddiki, S.; Clark-Sutton, K.; Graham, J.D. All Plug-In Electric Vehicles are not the Same: Predictors of Preference for a Plug-In Hybrid Versus a Battery-Electric Vehicle. *Transp. Res. Part D Transp. Environ.* **2018**, *65*, 1–13. [CrossRef]
19. Li, W.; Long, R.; Chen, H.; Yang, T.; Geng, J.; Yang, M. Effects of Personal Carbon Trading on the Decision to Adopt Battery Electric Vehicles: Analysis Based on a Choice Experiment in Jiangsu, China. *Appl. Energy* **2018**, *209*, 478–488. [CrossRef]
20. Lin, B.; Wu, W. Why People want to Buy Electric Vehicle: An Empirical Study in First-Tier Cities of China. *Energy Policy* **2018**, *112*, 233–241. [CrossRef]
21. He, X.; Zhan, W.; Hu, Y. Consumer Purchase Intention of Electric Vehicles in China: The Roles of Perception and Personality. *J. Clean. Prod.* **2018**, *204*, 1060–1069. [CrossRef]
22. Du, H.; Liu, D.; Sovacool, B.K.; Wang, Y.; Ma, S.; Li, R.Y.M. Who Buys New Energy Vehicles in China? Assessing Social-Psychological Predictors of Purchasing Awareness, Intention, and Policy. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *58*, 56–69. [CrossRef]
23. Ma, S.-C.; Fan, Y.; Guo, J.-F.; Xu, J.-H.; Zhu, J. Analysing Online Behaviour to Determine Chinese Consumers' Preferences for Electric Vehicles. *J. Clean. Prod.* **2019**, *229*, 244–255. [CrossRef]
24. Okada, T.; Tamaki, T.; Managi, S. Effect of Environmental Awareness on Purchase Intention and Satisfaction Pertaining to Electric Vehicles in Japan. *Transp. Res. Part D Transp. Environ.* **2019**, *67*, 503–513. [CrossRef]
25. Huang, X.; Ge, J. Electric vehicle development in Beijing: An analysis of consumer purchase intention. *J. Clean. Prod.* **2019**, *216*, 361–372. [CrossRef]
26. Habich-Sobiegalla, S.; Kostka, G.; Anzinger, N. Citizens' electric vehicle purchase intentions in China: An analysis of micro-level and macro-level factors. *Transp. Policy* **2019**, *79*, 223–233. [CrossRef]
27. Carley, S.; Siddiki, S.; Nicholson-Crotty, S. Evolution of Plug-in Electric Vehicle Demand: Assessing Consumer Per-ceptions and Intent to Purchase over Time. *Transp. Res. Part D Transp. Environ.* **2019**, *70*, 94–111. [CrossRef]
28. Chen, K.; Ren, C.; Gu, R.; Zhang, P. Exploring Purchase Intentions of New Energy Vehicles: From the Perspective of Frugality and the Concept of "Mianzi". *J. Clean. Prod.* **2019**, *230*, 700–708. [CrossRef]
29. Wang, L.; Fu, Z.-L.; Guo, W.; Liang, R.-Y.; Shao, H.-Y. What Influences Sales Market of New Energy Vehicles in China? Empirical Study Based on Survey of Consumers' Purchase Reasons. *Energy Policy* **2020**, *142*, 111484. [CrossRef]
30. Jayasingh, S.; Girija, T.; Arunkumar, S. Factors Influencing Consumers' Purchase Intention Towards Electric Two-Wheelers. *Sustainability* **2021**, *13*, 12851. [CrossRef]

31. He, Z.; Zhou, Y.; Wang, J.; Shen, W.; Li, W.; Lu, W. Influence of Emotion on Purchase Intention of Electric Vehicles: A Comparative Study of Consumers with Different Income Levels. *Curr. Psychol.* **2022**, *42*, 21704–21719. [CrossRef] [PubMed]
32. Ji, D.; Gan, H. Effects of Providing Total Cost of Ownership Information on Below-40 Young Consumers' Intent to Purchase an Electric Vehicle: A Case Study in China. *Energy Policy* **2022**, *165*, 112954. [CrossRef]
33. HSBMMC. *"Huashu Cup" National College Student Mathematical Modeling Competition, Question C*, 2021st ed.; Big Data and Mathematical Model Professional Committee of China Future Research Association: Tianjin, China, 2021.
34. Ozcelik, B.; Erzurumlu, T. Comparison of the Warpage Optimization in the Plastic Injection Molding Using Anova, Neural Network Model and Genetic Algorithm. *J. Mater. Process. Technol.* **2006**, *171*, 437–445. [CrossRef]
35. Guo, S.; Zhong, S.; Zhang, A. Privacy-Preserving Kruskal–Wallis Test. *Comput. Methods Programs Biomed.* **2013**, *112*, 135–145. [CrossRef]
36. Hoarau, A.; Martin, A.; Dubois, J.-C.; Le Gall, Y. Evidential Random Forests. *Expert Syst. Appl.* **2023**, *230*, 120652. [CrossRef]
37. Wang, Z.; Liang, Z.; Zeng, R.; Yuan, H.; Srinivasan, R.S. Identifying the Optimal Heterogeneous Ensemble Learning Model for Building Energy Prediction Using the Exhaustive Search Method. *Energy Build.* **2023**, *281*, 112763. [CrossRef]
38. Yang, Q.; Guo, X.; Zheng, H.; Dong, W. Single- and Multi-Objective Optimization of an Aircraft Hot-Air Anti-Icing System Based on Reduced Order Method. *Appl. Therm. Eng.* **2023**, *219*, 119543. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.