AIMS Mathematics

AIMS Mathematics, 9(2): 4161–4177.
DOI: 10.3934/math.2024204
Received: 09 December 2023
Revised: 04 January 2024
Accepted: 09 January 2024
Published: 15 January 2024

http://www.aimspress.com/journal/Math

Research article

An investigation into the determinants of satisfaction concerning varied toll policies on highways using the random forest model

Yonghua Liu[1], Ruikun Duan[1,*], Ke Shen[2], Qingxiong Luan[2], Hanqi Gao[2] and Hao Deng[2]

[1] Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming 650504, China

[2] Development Research Center of Yunnan Transportation Research Institute Co., Kunming 650011, China

* **Correspondence:** Email: 1447328934@qq.com.

Abstract: To address the insufficient consideration of public satisfaction's impact on the assessment of the implementation effectiveness of differentiated toll collection on expressways, a study was conducted using a satisfaction survey questionnaire from differentiated toll road sections in Yunnan Province in 2022. A random forest model (RF) was constructed under a two-category experiment to analyze the factors influencing satisfaction with expressway-differentiated toll policies. Multiple models underwent five-classification and two-classification experiments using the same training and test datasets. Results revealed that the RF model in the binary classification experiment exhibited a good fit. Notably, the satisfaction level with timely and accurate preferential policies emerged as the most critical factor, contributing 20.35% to the overall satisfaction with expressway differentiated toll policies. Independent effect analysis highlighted that the overall satisfaction with the differentiated charging method for empty trucks ranked highest, while satisfaction with the differentiated charging method for road sections was the lowest.

Keywords: differentiated charges; satisfaction evaluation; influencing factors; random forest model; logit model
Mathematics Subject Classification: 62P30

1. Introduction

As China's primary service, leading, and strategic transportation infrastructure, the expressway plays a pivotal role in the country's economic development. In recent years, the construction of highways in China has gradually reached its conclusion, and the highway network is taking its final form. The completion of this network is poised to strengthen communication and connectivity between regions in China, concurrently fostering the rapid growth of the nation's cargo transportation industry and related sectors. Historically, China's highway toll rates have remained relatively fixed, lacking a flexible rate adjustment mechanism. Some of the newly constructed highways adhere to high technical standards and incur elevated costs, contributing to relatively high toll rates [1]. Confronted with elevated highway tolls, drivers in certain regions opt for toll-free parallel roads, leading to diminished transportation efficiency and heightened logistics costs. In response to these challenges, the Ministry of Transport, the Ministry of Finance, and the National Development and Reform Commission collaborated to release a document. With a focus on traffic demand management, the plan involves introducing a diversified toll policy for expressways. The successful implementation of this policy is anticipated to play a pivotal role in cost reduction and efficiency enhancement within the logistics and transportation industry in specific areas [2].

Following the issuance of the "Implementation Plan for Comprehensively Promoting Differentiated Toll Collection of Expressways", numerous provinces and cities across China have sequentially introduced various measures for differentiated toll collection. These measures include differentiation based on road section, time period, vehicle type, and direction, tailored to their specific characteristics. This approach has, to some extent, mitigated traffic congestion, enhanced transportation efficiency, and lowered logistics costs. However, in evaluating the implementation effectiveness of these expressway differentiated charging measures, Ren [3] and others have provided detailed implementation plans but have not conducted a corresponding assessment of the differentiation's actual impact. Wang et al. [4] applied a differentiated toll approach to Hanyi Expressway and assessed its performance based on toll revenue and travel cost. The evaluation yielded the conclusion that the differentiated toll method surpassed the previously employed traditional toll method. In a similar vein, Liu et al. [5] gauged the implementation impact of differentiated toll collection on expressways by establishing technical indicators related to traffic capacity, service level, and accident rates. Their findings indicated that differentiated toll collection outperforms traditional toll methods in terms of traffic capacity, service level, and accident rates. Sang et al. [6] conducted an evaluation of highway differentiation, examining traffic volume change, toll income, average speed change, and total weight change. In summary, there is a limited scholarly assessment of the implementation impact of expressway differentiated charging using public satisfaction-oriented evaluation indicators.

Moreover, within the realm of transportation research focusing on satisfaction, Xiang et al. [7] delved into the factors influencing expressway operation service satisfaction from a public perspective. Grounded in customer satisfaction theory and tailored to the distinctive attributes of China's expressways, they formulated a public satisfaction evaluation index system and model. An empirical study on the Chengdu-Suiyu Expressway served as a case study, where the structural equation model was applied to validate the theoretical model and assess the degree of factor influence.

Wang [8] and colleagues constructed a satisfaction evaluation model for bus transfer preferential policies using Structural Equation Modeling (SEM). They employed the Partial Least

Squares (PLS) method for parameter estimation. Using Suzhou bus transfer preferential policies as an illustration, the satisfaction evaluation verified the soundness of the proposed evaluation model and index system. Their conclusion highlighted that perceived value exerts the most significant impact on passenger satisfaction, followed by perceived quality.

Li et al. [9] developed a satisfaction evaluation index system centered on passenger perception. The CRITIC (Criteria Importance through Inter-Criteria Correlation) method facilitated the objective weighting of the index, while the cloud model-based basic algorithm generated the evaluation cloud and result cloud. Illustrated through digital cloud characteristics and cloud image comparisons, the evaluation results were visually displayed. Applying this methodology to the passenger satisfaction assessment of Shanghai Rail Transit Line 2 validated the effectiveness and rationality of the evaluation method.

Chen et al. [10] utilized the ordered Probit model to establish the functional relationship between bus satisfaction and passenger individual characteristics, travel activity characteristics, and bus service quality attributes. Through marginal effect analysis, the quantitative evaluation of the significant influencing factors on bus satisfaction was achieved. Results indicated that gender, daily bus time, waiting time, transfer convenience, travel information service, waiting environment, fare price, and bus station setting significantly impacted bus service satisfaction. Nevertheless, models such as the cloud model, structural equation model, Probit, and others necessitate assumptions and preset models. Additionally, the outcomes of such models can only confer statistical significance on each influencing factor and overall satisfaction, lacking the ability to quantify the contribution of each factor to overall satisfaction [11–13].

Addressing the aforementioned issues, this paper takes a public perspective by conducting a satisfaction questionnaire survey on the implementation of differentiated toll roads in Yunnan Province. Subsequently, a machine learning model is established based on the collected survey data, effectively overcoming the limitations of previous satisfaction research models like the cloud model, structural equation model, and Probit, which were unable to quantify the contribution of each influencing factor to overall satisfaction. The established machine learning model is then employed to output the contribution of each influencing factor to overall satisfaction. This approach aids in identifying the primary factors influencing the differentiated charging policy, enabling the proposal of a set of practical suggestions aligned with these key factors. The results provide theoretical support for the comprehensive advancement of China's differentiated charging policy. Furthermore, the analysis of primary influencing factors in the differentiated charging policy is crucial for promoting the healthy development of China's expressway system and enhancing expressway transportation efficiency.

2. Research data

2.1. Research scope

In this study, we focus on the differentiated toll road sections in Yunnan Province in the year 2022. The selection criteria for these sections are derived from the pertinent documents, specifically the “Notice of the People’s Government of Yunnan Province on Printing and Issuing Several Policies and Measures for Steady Growth in 2022” (Yun Zhengfa No.7). Taking into account the distinctive features of Yunnan Province, the sections subject to differentiated charges are determined based on the guidelines outlined in Table 1, detailing the specific implementation rules.

Table 1. Specific implementation details of differential charges.

Name of measures	Section name	Specific implementation details
Differentiated charging methods by road section	A, B, C, D, E, F, G	In this context, road sections A, B, C, and D offer a 10% toll concession to buses utilizing ETC payment. Additionally, for road sections E, F, and G, there is a 15% toll concession provided for buses utilizing ETC payment.
Differentiated charging methods by time period	H	Buses and trucks on section H are eligible for a 20% toll discount during the period from 22:00 to 6:00 the following morning.
Differentiated charging methods by vehicle type	I	In section I, a 5% toll discount is applicable to buses using ETC payment. Furthermore, three or four types of buses on section I are entitled to a 10% toll discount.
Differentiated charging methods by direction	J	A 20% toll discount is provided for a specific category of buses traveling from the starting point to the end point of section J.
Differentiated charging methods for empty trucks	K, M, N	Empty trucks on sections K, M, and N are eligible for toll concessions of 40%, 30%, and 20%, respectively.

2.2. Survey content

This survey comprises two main sections:

1) Basic information of surveyed drivers:
   - Vehicle registration place
   - Vehicle type
   - License plate color
   - Education background
   - Monthly income
   - Payment method of tolls
2) Satisfaction evaluation index of surveyed drivers on differentiated charging policy:
   - Degree of understanding of the comprehensive promotion of expressway differentiated charging in Yunnan Province
   - Degree of understanding of the differentiated charging of this expressway
   - Specific measures of the differentiated charging of this expressway
   - Amount of relief
   - Publicity
   - Degree of satisfaction with timely and accurate concessions
   - Congestion after the differentiated charging of this expressway
   - Degree of satisfaction with the service level
   - Overall satisfaction with the differentiated charging policy of this expressway.

2.3. Descriptive statistics of survey data

All experimental protocols for this study received approval from the Development Research Center of Yunnan Academy of Transportation Sciences, China. The methods adhered to relevant guidelines and regulations, and informed consent was obtained from all participants before administering the questionnaire. The datasets utilized and analyzed in the current study are available from the corresponding author upon reasonable request.

Between October and December 2022, a questionnaire survey was conducted on road sections with differentiated charging policies in Yunnan Province. A total of 3,900 questionnaires were distributed across the 13 designated road sections, resulting in 3,866 questionnaires being recovered. Among them, 3,666 were deemed valid, yielding an effective questionnaire rate of 94.83%. The statistical description of the valid questionnaires is presented in Tables 2 and 3.

Table 2. Descriptive statistics of basic information of surveyed respondents.

Variable	Category	Frequency of occurrence	Proportion	Variable	Category	Frequency of occurrence	Proportion
Vehicle ownership	Provincial	2,220	60.56%	Vehicle type	2-axle truck	1,512	41.24%
	Outside the province	1,446	39.44%		3-axle truck	234	6.37%
Vehicle license plate color	Blue card	2,087	56.93%		4-axle truck	219	5.97%
	Yellow card	1,533	41.80%		5-axle truck	43	1.19%
	Green-brand	47	1.27%		6-axle truck	783	21.35%
Monthly profit	Under 3000 yuan	610	16.63%		Class 1 passenger cars	630	17.18%
	3000–5000 yuan	1,348	36.78%		Type 2 bus	47	1.27%
	5000–7000 yuan	1,147	31.30%		3 types of buses	164	4.46%
	7000–9000 yuan	338	9.23%		4 types of buses	35	0.96%
	More than 9000 yuan	228	6.21%	payment methods	Students' expenditure	2957	80.65%
Record of formal schooling	Junior high school and below	885	24.12%		Unit pays	709	19.35%
	High school and technical secondary school	1,564	42.67%				
	Junior college and undergraduate	1,182	32.25%				
	Master's degree and above	35	0.95%				

Table 3. Satisfaction evaluation index of surveyed respondents on differentiated charging policy.

Variable	Category	Frequency of occurrence	Proportion	Variable	Category	Frequency of occurrence	Proportion
The degree of understanding of the province's highway differentiated charges (X1)	Very understanding	785	21.40%	The degree of understanding of the differentiation policy of this expressway (X2)	Very understanding	633	17.28%
	Comparative understanding	548	14.96%		Comparative understanding	1028	28.03%
	General	1687	46.02%		General	1448	39.49%
	Not very understanding	521	14.20%		Not very understanding	482	13.14%
	Ignorance	125	3.41%		Ignorance	77	2.1%
Satisfaction with specific measures (X3)	Very understanding	1366	37.26%	The satisfaction of the amount of relief (X4)	Very satisfactory	1322	36.07%
	Comparative understanding	1636	44.63%		Satisfactory	1629	44.43%
	General	580	15.82%		General	631	17.20%
	Not very understanding	76	2.07%		Dissatisfied	70	1.90%
	Ignorance	8	0.02%		Very Dissatisfied	15	0.40%
Satisfaction of service level (X5)	Very satisfactory	1383	37.72%	Congestion satisfaction (X6)	very satisfactory	594	43.15%
	Satisfactory	1535	41.88%		Satisfactory	643	46.34%
	General	645	17.60%		General	162	9.16%
	Dissatisfied	93	2.55%		Dissatisfied	12	0.96%
	Very Dissatisfied	35	0.96%		Very Dissatisfied	5	0.40%
Propaganda satisfaction (X7)	Very satisfactory	1351	36.86%	Timely and accurate preferential satisfaction (X8)	Very satisfactory	1623	44.27%
	Satisfactory	1643	44.82%		Satisfactory	1477	40.29%
	General	619	16.88%		General	514	14.01%
	Dissatisfied	35	0.96%		Dissatisfied	43	1.19%
	Very dissatisfied	18	0.48%		Very Dissatisfied	9	0.24%
Overall satisfaction (Y)	Very satisfactory	1322	36.07%				
	Satisfactory	1591	43.39%				
	General	716	19.51%				
	Dissatisfied	32	0.88%				
	Very dissatisfied	6	0.16%				

3. Random forest model

3.1. Model choice

Due to the complex and nonlinear relationship between overall satisfaction and the variables associated with differentiated charging policies, the use of a general linear model for analysis can significantly impact the results. Moreover, commonly used models such as the structural equation model and cloud model are prone to subjective human factors. The random forest model (RF) is effective in capturing the nonlinear relationship between overall satisfaction and the relevant variables in differentiated charging policies. Importantly, the RF model does not require any pre-assumptions during calculations and can provide the relative importance of each variable to the dependent variable, facilitating result interpretation [14,15].

3.2. Model theory

Random Forest is an ensemble algorithm consisting of multiple decision trees{h(X,Y_K), K=1, 2,...}, where Y_K is a random variable determining the random extraction of the training set and the random selection of candidate splitting attributes. The detailed process of the Random Forest algorithm is illustrated in Figure 1 as well as the following steps:

1) The original data set is split into training data and test data in an 8:2 ratio.
2) A self-help training set L [14] is created by randomly selecting N samples from the training data set.
3) Using L as the training data, a decision tree T is constructed. For each analysis node, m feature attributes are randomly chosen from M feature attributes as candidate split attributes. Based on the Gini index, one of the m features or attribute variables is selected for splitting. This process is repeated until the tree successfully divides all the test data [14].

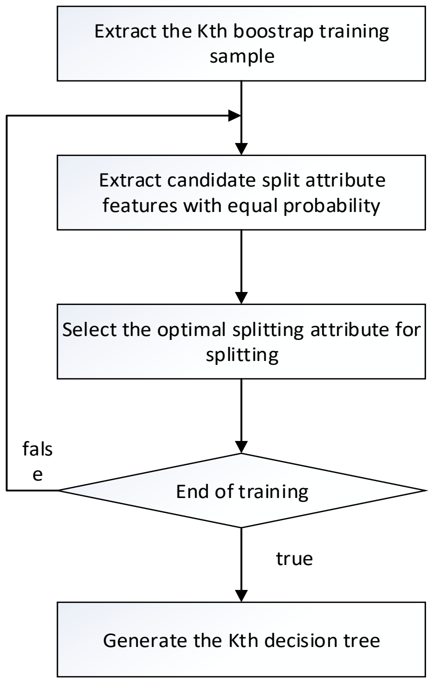

Figure 1. Random Forest construction process.

3.3. RF model interpretability

The explanatory power of the RF model manifests primarily in two dimensions. First, it delineates the contribution of independent variables to dependent variables by highlighting feature importance. Second, it generates partial correlation diagrams through Partial Functional Dependence (PDP), offering insights into the relationship between explanatory variables and explained variables [16].

(1) **Feature Importance.** The calculation concept behind the feature importance of the RF model is straightforward. It involves assessing the contribution each feature makes to each tree within the random forest, averaging these contributions, and then comparing the relative importance

of different features. The specific calculation formula is presented in (1):

$$P_{x_i} = \frac{1}{N}\sum_{n=1}^{N}\sum_{j=1}^{J-1} d_j \, . \tag{1}$$

In the formula, P_{x_i} represents the importance score of the feature Xi, J is the number of nodes in each decision tree, and d_j represents the reduction of squared error loss after the J-th splitting of Xi.

(2) **Partial Functional Dependency.** Partial functional dependencies illuminate the marginal effect of one or two explanatory variables on the explained variables, visually representing these low-cost interactions. This can be computed using the Monte Carlo method, and the specific calculation formula is outlined in (2):

$$f_s(x_s) = \frac{1}{M}\sum_{i=1}^{M} f\left(x_s, x_c^i\right). \tag{2}$$

In the formula, $f_s(x_s)$ is the partial function dependent on the fitting function for x_s. x_s is the explanatory variable under investigation, and x_c^i is the remaining explanatory variable. M is the number of samples.

4. Authentic proof analysis

Utilizing the questionnaire survey data from the differentiated toll road sections in Yunnan Province in 2022, this study employs Random Forest (RF) for the classification and prediction of expressway differentiated toll policy satisfaction evaluation data. To enhance fitting accuracy, we conduct both five-classification and two-classification experiments. The binary classification experiment consolidates five discrete values into two broader categories, wherein ‘very satisfied’ and ‘satisfied’ are mapped to the “satisfactory” class, while ‘general’, ‘unsatisfactory’ and ‘very unsatisfactory’ are mapped to the ‘unsatisfactory’ class [16].

4.1. Determination of model hyperparameters

In machine learning models, certain parameters, known as hyperparameters, must be set before training the model. Incorrectly set hyperparameters can lead to overfitting or underfitting issues, impacting the effectiveness of model training and prediction. Therefore, hyperparameter determination is a crucial aspect of model optimization [17].

The model is implemented using the Scikit-learn library in Python. Following established research practices, this paper employs the grid search method to cross-validate and evaluate all possible combinations of hyperparameters, utilizing the error value as the criterion [17]. The two hyperparameters optimized for the RF model are the number of trees and the maximum eigenvalue. The parameter adjustment process is depicted in Figures 2 and 3. In the binary classification experiment, the model achieves the best fitting effect when the number of trees is set to 181 and the maximum eigenvalue is 2. In the five-classification experiment, the optimal configuration is achieved with 311 trees and a maximum eigenvalue of 2.

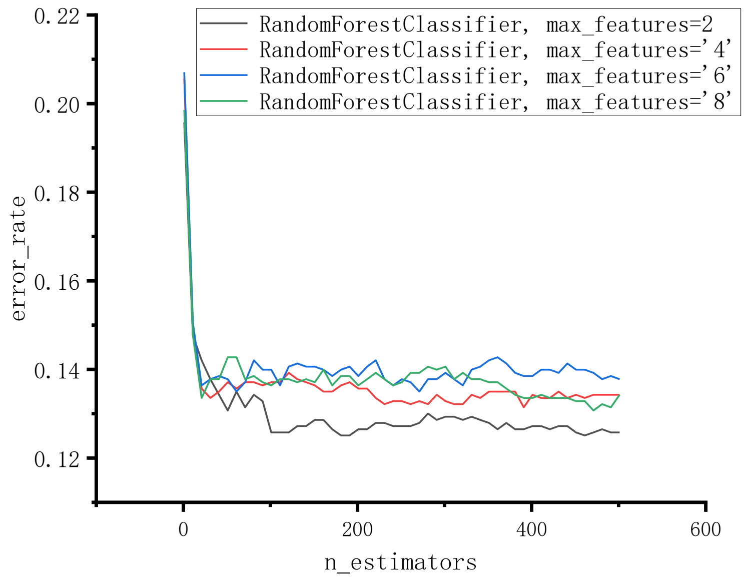

Figure 2. Random forest model with two classification experiment parameter adjustment diagram.

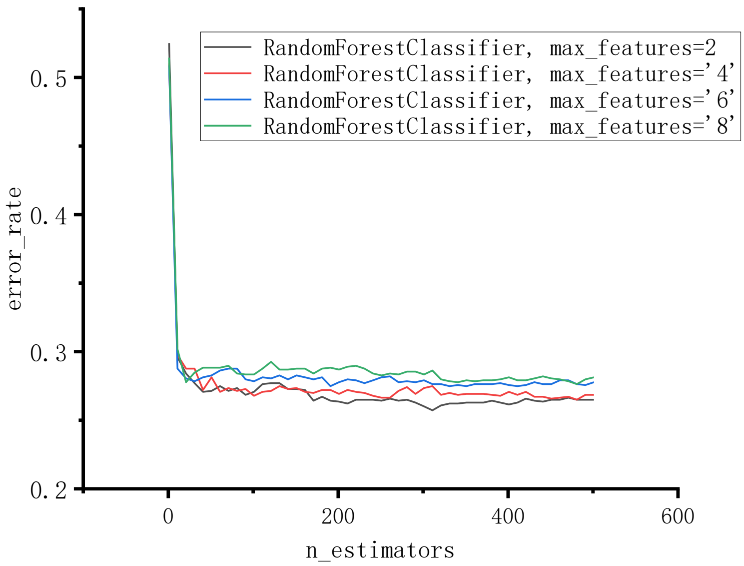

Figure 3. Random forest model with five-category experimental parameter adjustment diagram.

4.2. Overall effect analysis

The overall importance and influence of each variable are presented in Table 4 and Figure 4. Examining Table 4 and Figure 4 reveals that satisfaction with timely and accurate preferential policies is the most crucial factor impacting the overall satisfaction of expressway differentiated charging policies, contributing 20.35%. Following closely are satisfaction with specific measures, satisfaction with the amount of relief, and understanding of the comprehensive promotion of differentiated tolls on this expressway, contributing 14.40%, 14.35%, and 14.24%, respectively. This emphasizes that a significant proportion of drivers prioritize the accuracy of the differentiated toll policy, followed by the timeliness of its implementation and the extent of concessions. Consequently, for the future implementation of differentiated toll collection, to enhance overall satisfaction and ensure the policy is more beneficial, expressway operators should continually enhance the accuracy and timeliness of the differentiated toll collection process. Furthermore, clarifying the range of concessions in differentiated toll collection is essential to enhance understanding among users.

Table 4. Overall results of the model.

Influencing variable categories	Index	Relative importance	Ranking
The satisfaction evaluation of the surveyed drivers on the differentiated charging policy	The understanding of the comprehensive promotion of differentiated charges in Yunnan Province (X1)	6.90%	8
	The degree of understanding of the comprehensive promotion of differentiated tolls on this expressway (X2)	14.24%	4
	Satisfaction with specific measures (X3)	14.40%	2
	The satisfaction of the amount of relief (X4)	14.35%	3
	Satisfaction of service level (X5)	8.09%	7
	Congestion satisfaction (X6)	8.56%	6
	Propaganda satisfaction (X7)	13.12%	5
	Timely and accurate preferential satisfaction (X8)	20.35%	1

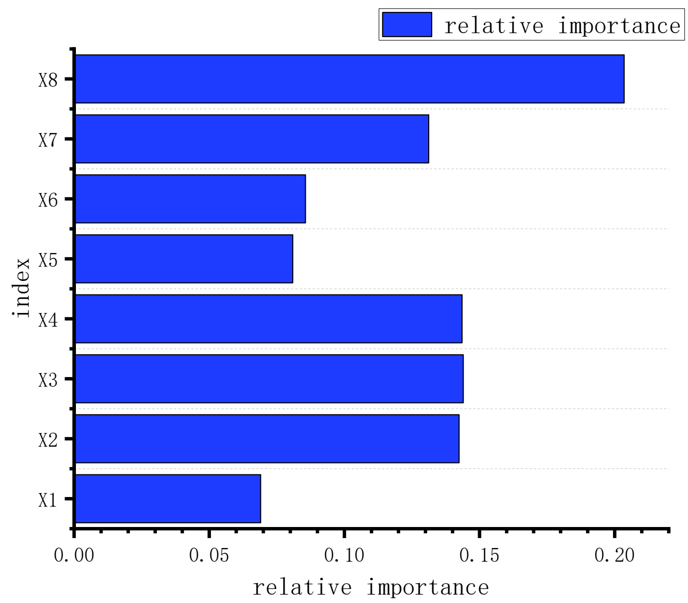

Figure 4. Relative importance output.

4.3. Independent effect analysis

The analysis of overall satisfaction under different differentiated charging methods is conducted using the partial functional dependence of the RF model. Figure 5 illustrates that the K section has the highest overall satisfaction, followed by the M section, while the B, C, and D sections exhibit relatively lower overall satisfaction. This pattern can be attributed to the recent shift in China's freight vehicle charging method from weight-based to vehicle-based, effectively addressing overload and toll station congestion issues to some extent. However, this policy change has drawbacks, where truck drivers opt for highways when their freight vehicles are full, and choose parallel free roads when empty. Implementing differentiated charging for empty trucks can address the limitations of sub-model charging, leading to higher overall satisfaction.

The differentiated charging method for sub-sections, based on a 95% discount for original ETC users, involves fewer users and offers a relatively small preferential amount. Additionally, it may be easily confused with the 95% discount policy for original ETC users, resulting in relatively lower overall satisfaction for this differentiated charging measure.

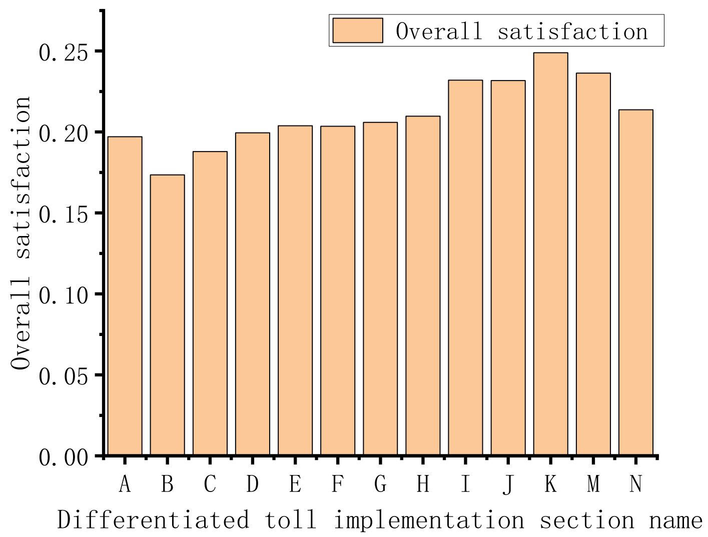

Figure 5. The overall satisfaction map of the differentiated toll implementation section.

4.4. Model analysis

To further validate the superior predictive performance of the random forest model chosen in this study, several commonly employed iterative algorithms, including AdaBoost, gradient boosting iterative decision tree (GBDT), extreme gradient boosting (XGB), and Logit (NL) models, were selected for comparison based on existing research results. These models for comparison were also subjected to five-category and two-category experiments using the same training dataset and test dataset [18].

4.4.1. Evaluation index of binary classification experiment

To conduct a more scientific comparison of model performance, we utilize precision and AUC values as metrics. These performance metrics are calculated based on the confusion matrix [19].

- Confusion matrix.

The confusion matrix serves as the foundation for calculating the performance evaluation of various machine learning models. The columns represent predictions classified as positive and negative cases, respectively, while the rows denote positive and negative cases, respectively. When the prediction is positive and the actual value is also positive, it is considered a true positive, and the corresponding sample count is entered into the appropriate position in the confusion matrix (i.e., the upper-left corner of the matrix). Similarly, the confusion matrix includes counts for false positives, false negatives, and true negatives. The specific confusion matrix is presented in Table 5 [15].

Table 5. Confusion matrix of two-category experiment.

The real situation	Forecasting results	
	Positive example	Counterexample
Positive example	True positives (TP)	False counterexample (FN)
Counterexample	False positives (FP)	True counterexample (TN)

- Precision index.

In accordance with the confusion matrix, the precision index can be easily calculated. The precision index ranges between 0 and 1, with a result closer to 1 indicating a better prediction effect, while a result closer to 0 suggests a poorer prediction effect. The calculation formula is as follow:

$$P = \frac{TP}{TP + FP}\ . \tag{3}$$

- AUC index.

The AUC evaluation index is commonly employed in binary classification machine learning and stands for Area Under the Curve. The ‘Curve’ refers to the ROC (Receiver Operating Characteristic) curve. A higher AUC value, closer to 1, indicates better classification performance of the model. An AUC value of 0.5 suggests that the model’s classification ability is equivalent to random guessing, and when the AUC is less than 0.5, it implies that the model’s classification performance is worse than guessing.

To calculate the AUC index, it is essential to first define the ROC curve. The ROC curve is built on the ‘True Positive Rate’ (TPR) as the vertical axis and the ‘False Positive Rate’ (FPR) as the horizontal axis. These are defined as (4) and (5) and are plotted on the coordinate system to obtain the ROC curve. The AUC value refers to the area enclosed by the abscissa below the ROC curve [19]:

$$FPR = \frac{FP}{FP + TN}\ . \tag{4}$$

$$TPR = \frac{FP}{TP + FN}\ . \tag{5}$$

4.4.2. Evaluation index of five-category experiment

The evaluation of the five-category model is based on the confusion matrix, although the matrix is no longer presented in Table 5; instead, it is shown in Table 6.

Table 6. Confusion matrix of five-category experiment.

	The prediction classification is 1	The prediction classification is 2	The prediction classification is 3	The prediction classification is 4	The predicted classification is 5
The real classification is 1	TP	FP	FP	FP	FP
The real classification is 2	FP	TP	FP	FP	FP
The real classification is 3	FP	FP	TP	FP	FP
The real classification is 4	FP	FP	FP	TP	FP
The real classification is 5	FP	FP	FP	FP	TP

It is evident from Table 6 that in the confusion matrix of the five-category experiment, the correctly classified samples (where the predicted value aligns with the true value) are located along the diagonal. Similar to the precision index calculation in binary classification, the precision index for each category is calculated separately in the five-category evaluation. Specifically, for a given category A and category B, the accuracy calculation is as shown in formulas (6) and (7) [19]:

$$P_A = \frac{TP_A}{TP_A + FP_A}. \tag{6}$$

$$P_B = \frac{TP_B}{TP_B + FP_B}. \tag{7}$$

To comprehensively evaluate the prediction effect of the five classifications, it is crucial to consider the prediction performance across all categories. Taking the precision index as an example, we select the following two evaluation indices to comprehensively assess the prediction effect of the five classifications [19].

1) Macro Average Precision (Macro-P):

The macro average precision treats each category equally, calculating the total precision index as the arithmetic mean of all classification precision indices. The macro average precision rate ranges between 0 and 1, with a result closer to 1 indicating better prediction performance, while a result closer to 0 suggests poorer performance. The calculation formula is shown in (9) [17]:

2) Micro-Average Precision (Weighted-P):

The micro-average precision rate calculates the sum of true positives and false positives for all categories and then computes the overall precision rate. It considers the weight of each category and is more sensitive to larger categories. The value of the micro-average precision rate is between 0 and 1, with a result closer to 1 indicating better prediction performance, while a result closer to 0 suggests poorer performance. The calculation formula is shown in (10) [17].

From Tables 7 and 8, and Figures 6 and 7, it can be observed that in this experiment, after multiple model parameter tuning on the original dataset, the RF model demonstrates the best performance in two-class prediction, with precision and AUC values of 0.85 and 0.76, respectively.

The AdaBoost model exhibits the least ideal performance in two-class classification. For the five-classification prediction, the RF model also performs the best, with Macro-Average and Weighted-Average values of 0.68 and 0.73, respectively. The five-classification effects of the AdaBoost model and the NL model are not ideal [15]. Therefore, we select the RF model under the two-classification condition to evaluate the satisfaction of the expressway differentiated charging policy, ensuring high accuracy.

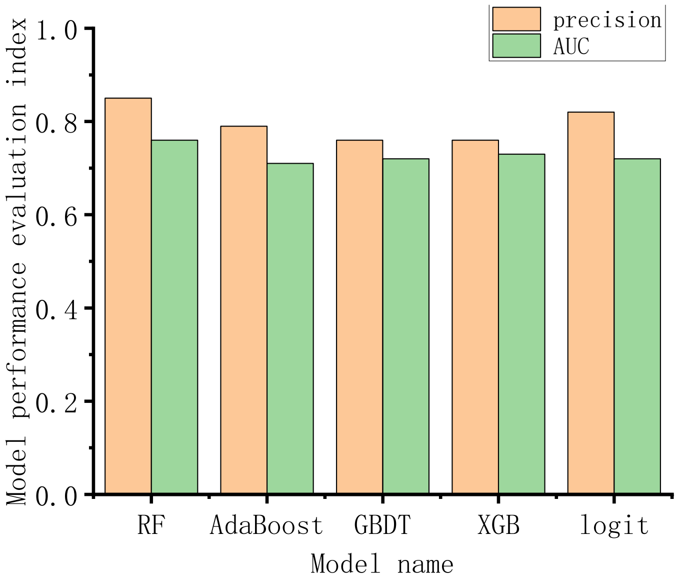

Figure 6. Performance comparison of binary classification experimental model.

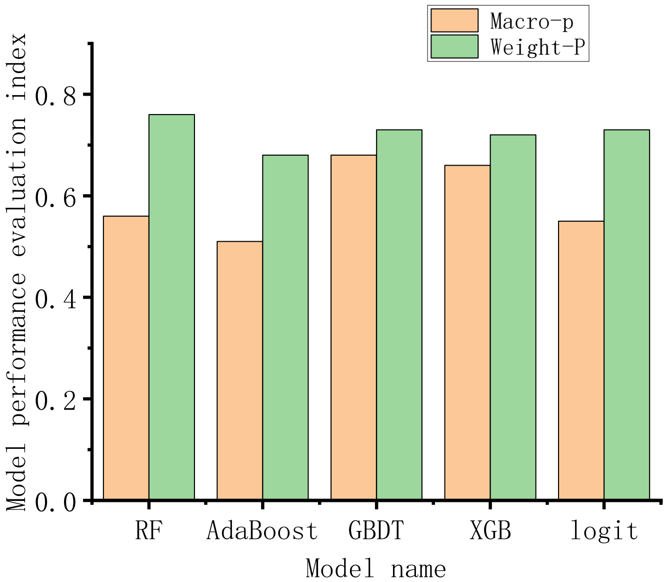

Figure 7. Comparison of performance of five classification experimental models.

Table 7. The performance index results of each model in the binary classification experiment.

	RF	AdaBoost	GBDT	XGB	logit
Precision	0.85	0.79	0.76	0.76	0.82
AUC	0.76	0.71	0.72	0.73	0.72

Table 8. Five classification experiment results of each model performance index.

	RF	AdaBoost	GBDT	XGB	logit
macro-p	0.56	0.51	0.68	0.66	0.55
Weighted-p	0.76	0.68	0.73	0.73	0.73

5. Conclusions

In this study, based on the 2022 survey data from differentiated toll sections in Yunnan Province, we successfully classified and predicted expressway differentiated toll policy satisfaction. The results highlighted the superiority of the two-category experiment over the five-category experiment, with the Random Forest (RF) model exhibiting the best performance, achieving precision values and AUC values of 0.85 and 0.76, respectively.

The comprehensive analysis underscored the pivotal role of satisfaction with timely and accurate preferential policies, contributing 20.35% to the overall satisfaction of expressway differentiated charging policies. Subsequently, satisfaction with specific measures, relief amount, and understanding of the comprehensive promotion of differentiated tolls on the expressway followed closely, contributing 14.40%, 14.35%, and 14.24%, respectively. For stakeholders, such as high-speed operation and management enterprises and the government, future implementation of differentiated charging policies must prioritize improving the accuracy and timeliness of the process to enhance overall satisfaction and gain public support.

Furthermore, the independent effect analysis revealed that overall satisfaction with the differentiated charging method for empty trucks was the highest. Combining this method with the existing sub-model charging approach effectively mitigated the drawbacks of sub-model charging, fostering a fairer charging system in China. Conversely, overall satisfaction with the differentiated toll collection method was the lowest. Hence, future implementations of differentiated toll collection should adopt a more diversified approach beyond the existing 95% discount for original ETC users.

In terms of research limitations and future prospects, we acknowledge the lack of specificity in analyzing various differentiated charging methods in Yunnan Province. Future research should consider a more targeted approach, focusing on one or two differentiated charging methods for a comprehensive understanding of user satisfaction. This approach would offer more instructive insights for the implementation of diverse differentiated charging methods for highways.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Entrepreneurial Practice Project—Application of System Dynamics in Highway Financial Evaluation (# 2021106740022). Horizontal Project: Highway Section Passing Load and Intelligent Regulation and Control Data Fusion Analysis Technology Service Project (# 2023530103003765).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. Y. Cao, *Research on pricing strategy of differential toll collection of expressway freight trucks*, Ph.D. Thesis, Chongqing: Chongqing Jiaotong University, 2022.
2. K. Shen, L. Duan, Z. Li, C. Yan, Calculation and analysis on differentiated charges for highway different trucks, *China Transp. Rev.*, **44** (2022), 141–148.
3. Y. Ren, J. Lu, S. Wu, Q. Xiang, Variable pricing evaluation and the evaluation index system of freeway, *J. Trans. Eng. Inform.*, **1** (2009), 17–21. https://doi.org/10.3969/j.issn.1672-4747.2009.01.004
4. L. Wang, G. Feng, S. Wu, Y. Li, Study on the differentiated toll rates of different sections of Wuhan-Yichang expressway, *J. Highw. Transp. Res. Dev.*, **39** (2022), 183–190. https://doi.org/10.3969/j.issn.1002-0268.2022.01.024
5. D. Liu, *Research on differential tolling mechanism of expressway trucks*, Ph.D. Thesis, Beijing: Beijing Jiaotong University, 2020.
6. M. Sang, Study on the effect evaluation of the reform of Expressway freight vehicle toll standard, *Price: Theory Pract.*, **10** (2021), 45–48.
7. P. Xiang, L. Gao, F. Jia, Influencing Factors of expressway operation service satisfaction from the public perspective, *J. Chongqing Jiaotong Univ. (Natl. Sci.)*, **12** (2022), 26–32.
8. R. Wang, P. Du, Satisfaction model of transfer policy based on PLS-SEM, *J. Transp. Syst. Eng. Inform. Tech.*, **A01** (2018), 10–15.
9. L. Li, X. Guo, J. Fu, B, Wu, Evaluation approach of passenger satisfaction for urban rail transit based on cloud model, *J. Tongji Univ. (Natl. Sci.)*, **47** (2019), 378–385.
10. J. Chen, J. Fan, X. Li, S. Zhu, Influencing factors of the satisfaction of passengers with cars for bus in small and medium cities based on ordered probit model, *J. Chongqing Jiaotong Univ. (Natl. Sci.)*, **3** (2022), 45–52.
11. J. He, S. Zhu, Research on the comprehensive evaluation of rail transit and feeder transit transfer based on extensible cloud model, *J. Railway Sci. Eng.*, **8** (2021), 2183–2190. https://doi.org/10.19713/j.cnki.43-1423/u.T20200983
12. R. Li, J. Chen, Z. Fu, P. Yong, Passenger satisfaction analysis of customized bus based on structural equation model, *Sci. Tech. Eng.*, **25** (2020), 10499–10503.
13. K. Jiang, Q. Shao, Analysis of passengers' willingness to use intelligent facilities in airports based on technology acceptance model, *Sci. Tech. Eng.*, **23** (2023), 5777–5784.
14. Z. Fu, P. Tao, J. Chen, D. Chen, J. Di, The model of travelers' parking choice intention from the perspective of built environment, *J. Kunming Univ. Sci. Tech. (Natl. Sci.)*, **48** (2023), 142–150.

15. X. Wang, C. Shao, L. Guan, C. Yin, Exploring influences of built environment on car ownership based on a machine learning method, *J. Transp. Syst. Eng. Inform. Tech.*, **4** (2020), 173–177. https://doi.org/10.16097/j.cnki.1009-6744.2020.04.025
16. X. Fan, *Research on customer satisfaction prediction and influencing factors of O2O local service based on machine learning*, Ph.D. Thesis, Beijing: Beijing University of Posts and Telecommunications, 2021.
17. Z. Fu, Y. Gao, J. Chen, Q. Chen, Impact model of COVID-19 and built environment on bus passenger flow, J. *Transp. Syst. Eng. Inform. Technol.*, **23** (2023), 207–215.
18. C. Ding, P. Chen, J. Jiao, Non-linear effects of the built environment on automobile involved pedestrian crash frequency: a machine learning approach, *Accident Anal. Prev.*, **112** (2018), 116–126. https://doi.org/10.1016/j.aap.2017.12.026
19. Z. Zhou, *Machine learning*, Beijing: Tsinghua University Press, 2016.

AIMS Press

©2024 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)