

Research Article

Prediction and Analysis of Train Passenger Load Factor of High-Speed Railway Based on LightGBM Algorithm

Bing Wang ¹, Peixiu Wu ¹, Quanchao Chen ^{1,2,3} and Shaoquan Ni ^{1,2,3}

¹School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, China

²National and Local Joint Engineering Laboratory of Comprehensive Intelligent Transportation, Southwest JiaoTong University, Chengdu, China

³National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Chengdu, China

Correspondence should be addressed to Shaoquan Ni; shaoquanni@163.com

Received 3 April 2021; Revised 9 May 2021; Accepted 2 June 2021; Published 15 June 2021

Academic Editor: Chi-Hua Chen

Copyright © 2021 Bing Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the prediction accuracy of train passenger load factor of high-speed railway and meet the demand of different levels of passenger load factor prediction and analysis, the influence factor of the train passenger load factor is analyzed in depth. Taking into account the weather factor, train attribute, and passenger flow time sequence, this paper proposed a forecasting method of train passenger load factor of high-speed railway based on LightGBM algorithm of machine learning. Considering the difference of the influence factor of the passenger load factor of a single train and group trains, a single train passenger load factor prediction model based on the weather factor and passenger flow time sequence and a group of trains' passenger load factor prediction model based on the weather factor, the train attribute, and passenger flow time sequence factor were constructed, respectively. Taking the train passenger load factor data of high-speed railway in a certain area as an example, the feasibility and effectiveness of the proposed method were verified and compared. It is verified that LightGBM algorithm of machine learning proposed in this paper has higher prediction accuracy than the traditional models, and its scientific and accurate prediction can provide an important reference for the calculation of passenger ticket revenue, operation benefit analysis, etc.

1. Introduction

High-speed railway has become the main transportation mode for passengers' mode of transportation for passengers. According to the relevant statistical data, in 2019, the passenger volume of the national railway reached 3.57 billion, of which the passenger volume of multiple unit train was 2.29 billion, accounting for 64.15%. High-speed railway is a significant driver of railway passenger operation revenue and passenger flow growth, and its profit and loss analysis is critical to train operation and operation decision, and the passenger load factor is used as a direct measure of train operation efficiency and a momentous basis for calculating passenger ticket revenue. Scientific and accurate prediction of the train passenger load factor can provide significant reference for train operation scheme, ticket revenue calculation, operational benefit analysis, and so on and so forth [1].

Passenger load factor prediction of passenger trains is usually based on historical data of passenger tickets; the traditional method is to input the train information into the electronic form, use manual to process, classify, and estimate the passenger load factor, and to form a decision table [2]. Nevertheless, there are some problems such as substantial error and inconsistent decision information. At present, there are quite a few research studies on the passenger load factor prediction of high-speed railway multiple units. Different scholars use a variety of model methods to predict, such as multiple regression model, time series model, neural network model, decision tree model, gray theory model, and integrated learning algorithm model. Aiming at the competitive relationship between high-speed train and air transport, Wang et al. [3] proposed a prediction model of the passenger load factor of the high-speed railway trains based on Adaboost-CART algorithm from the perspective of the

impact of air fare level and dynamic fluctuation on high-speed railway trains passenger flow. From the perspective of train attributes, Zhang et al. [4] proposed a classification and prediction model of the passenger train load factor based on random forest algorithm. On the basis of analyzing the influencing factors of the train passenger load factor, Xu and Nie [5] established a BP neural network prediction model of train passenger load factor, considering the factors of train attributes and the operation period. Based on two single prediction models, ARIMA model and BP neural network, Zhang and Bai [6] constructed a linear combination prediction model of railway passenger load factor according to the principle of minimum sum of square errors [7]. In the research of machine learning prediction, quite a few studies use machine learning algorithms for short-term traffic flow prediction [1, 8, 9] and use LightGBM and XGBoost for prediction and classification [10–15]. Dong et al. [16] established a short-term traffic flow prediction model based on XGBoost algorithm. After analyzing the vehicle speed, road, and weather features in the course of the operation of the bus, Wang et al. [17] established a prediction model of bus travel time based on LightGBM algorithm. Huang et al. [18] constructed deep belief neural network based on multitask learning to predict traffic volume [19]. Zhang et al. [2] constructed a short-term traffic flow prediction model based on the fusion algorithm of XGBoost and LightGBM.

In the study of the passenger load factor or passenger volume forecast, various models are mainly used to predict the historical passenger load factor or passenger volume or passenger volume, and the rules of generating target variables are rarely obtained according to the attributes of trains [20–23]. In this paper, for high-speed railway trains, consider the influence factors such as train attributes, historical weather, and passenger flow sequence, and a single train passenger load factor prediction model and a group train passenger load factor prediction model based on the LightGBM algorithm are proposed, which can provide decision-making basis for ticket revenue calculation and operation benefit analysis.

1.1. Influence Factors. The passenger load factor of high-speed railway is an index reflecting the utilization degree of passenger carrying capacity. It is the ratio of passenger turnover to the total number of passenger kilometers, which is expressed as a percentage of the average number of passenger per kilometer [24]. The passenger load factor data comes from the analysis and statistics of passenger flow; in essence, the passenger flow determines the passenger load factor, and the influencing factors of passenger travel choice are the factors that affect the train occupancy rate. Hence, the analysis of influencing factors of the passenger load factor is the analysis of passenger flow travel choice.

From the perspective of demand and the macro-distribution of passenger flow, the spatial distribution of passenger flow is determined by the regional economic development level, population, and function orientation along the high-speed railway. In a period of time, the macrofactors such as regional economic development level

are relatively stable, so the spatial distribution of passenger flow is also at a relatively stable level. Passenger travel has evident time preference for specific travel behavior, and the departure and arrival time of a train will directly affect the train's load factor. Simultaneously, travel time and weather will affect the choice of the travel mode. For different time's nodes, on weekdays, most passengers travel mainly for business; nevertheless, on weekends and holidays, most passengers travel for tourism, family visits, etc. Therefore, the difference in travel time will also affect the load factor of the train.

From the perspective of transportation supply, firstly, the distribution of passenger flow direction is unbalanced, so the train operation direction will affect the train passenger load factor; the number of trains running between OD, namely, the service frequency of trains between OD is one of the main factors affecting the choice of passengers; furthermore, the departure and arrival time, running mileage, station of the way, train capacity, and type of the train will all affect the choice of passenger travel, thus affecting the load factor of the train [21].

By and large, the influencing factors of the train passenger load factor can be divided into internal factors and external factors, as shown in Figure 1. In a period of time, the regional economic level, population, and function orientation of city is relatively stable; furthermore, the main factors affecting the passenger load factor of high-speed railway are the direction of train operation, service frequency of OD, train attributes, weather, and travel time [14]. To this, this paper mainly considers attributes of the train, weather, and travel time.

2. Prediction Model

The passenger load factor prediction model of high-speed railway is mainly composed of data acquisition, data processing, feature engineering, model training, and prediction.

- (1) Data acquisition: the historical weather data, train attribute data, and passenger load factor data are obtained from various ways, and the original data is formed by data fusion.
- (2) Data processing: the types and dimensions of data variables are inconsistent; therefore, it is necessary to transform the categories and features of the original data before data modeling so that the data can meet the requirements of algorithm data structure. And, data processing mainly includes data transformation and data cleaning, such as removing the character "C" from the maximum and minimum temperature of the weather data. For the train with few times that are temporarily operating or have been suspended, it should be deleted (high-speed railway below 60 times in this paper have been deleted); as there is a great difference in the change of the passenger load factor in each operation period, this thesis mainly selects weekday data for research.
- (3) Feature engineering: average temperature feature, decomposition date feature, and multiclass

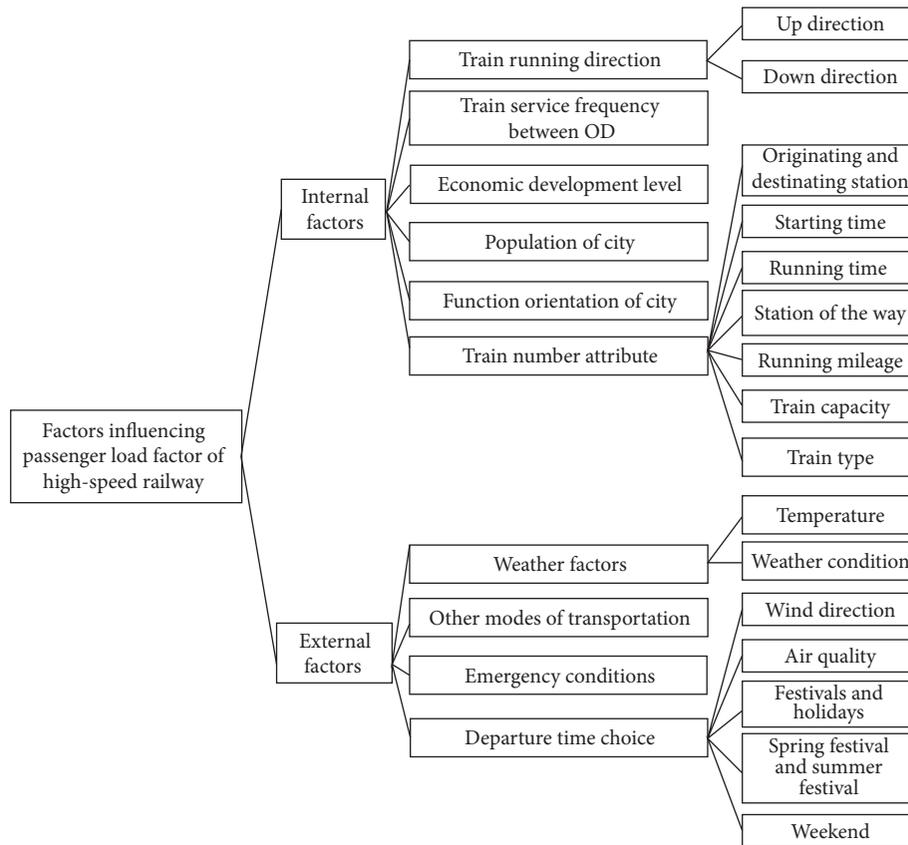


FIGURE 1: Influencing factors of train passenger load factor.

classification variable code are constructed. For train attribute, weather, and time characteristic data, the following characteristic engineering processing is carried out:

- ① The minimum and maximum temperatures are replaced by the characteristic average temperature (Avgtemperature)
- ② Constructing the features of passenger flow time sequence, new features DayOfWeek, WeekOfYear, Month, and Day are constructed by decomposing date features, and it denotes the days of a week, the week of a year, the months of a year, and the days of a month, respectively
- ③ LabelEncoder features mileage, capacity, weather, wind turbines (WIntensity), departure time (DepaTime), run time (OperTime), and turn discrete variables into multiclass continuous numerical variables

All feature engineering obtained after data preprocessing, including historical weather and train attributes of the passenger rate data, are shown in Table 1.

- (4) Model training and prediction: after feature engineering, the sample data set is constructed and divided into the training set and the test set. In the meantime, the model is trained and tested. The framework of the prediction model is shown in Figure 2.

For a single train, the attributes of the train are fixed, and the main factors affecting its passenger load factor are weather and time series characteristics. For group trains, the attributes of the train are one of the cardinal factors affecting the passenger load factor, therefore, the prediction of passenger load factor prediction of group trains, train attributes, weather features, and time-series characteristics need to be considered.

3. Case Analysis

To verify the effectiveness of the single train passenger load factor prediction model and group train passenger load factor prediction model based on LightGBM algorithm proposed in this paper, the passenger load factor data of all down directions from station A of a high-speed railway in the area as an example is taken. And, the load factor data of high-speed railway comes from all trains that departed from station A in the downward direction from October 9, 2017, to September 30, 2019, (provided by Station A); historical weather data is obtained from the 2345 weather forecast network through *Python* (<http://tianqi.2345.com/>). The target data covers 722 days of historical data of station A from October 9, 2017, to September 30, 2019; train attribute data includes the train number/type, departure station, terminal station, departure time, arrival time, traveling time, stop station, and ticket price (<https://www.12306.cn>). Based on machine LightGBM algorithm, this paper forecasts the

TABLE 1: Features of train passenger load factor data.

Feature	Variable type
Mileage	Multiple categorical variable
Capacity	Multiple categorical variable
Avgtemperature	Continuous numerical value
Weather	Multiple categorical variable
WIntensity	Multiple categorical variable
DepaTime	Multiple categorical variable
AQI	Multiple categorical variable
OperTime	Multiple categorical variable
DayOfWeek	Multiple categorical variable
WeekOfYear	Multiple categorical variable
Month	Multiple categorical variable
Day	Multiple categorical variable
StatNumber (number of intermediate stations)	Continuous numerical value
Load factor	Continuous numerical value

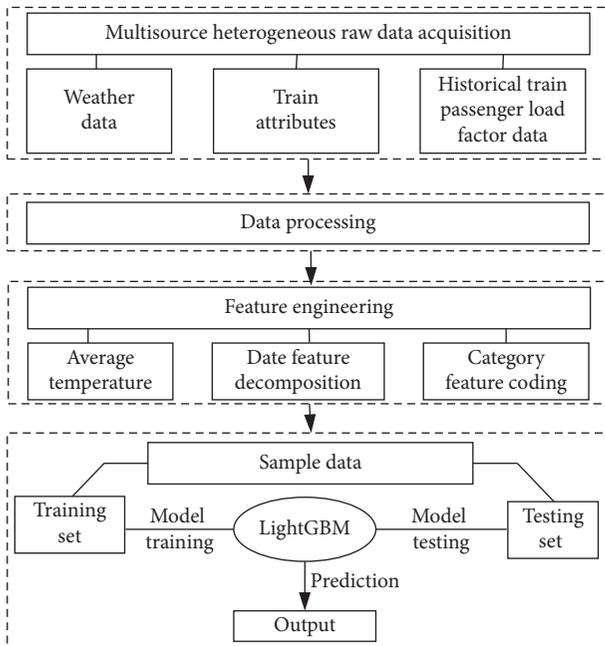


FIGURE 2: Overall framework of prediction model.

data of train passenger load factor, and the train load factor data of the first 600 days is taken as the training set and the last 67 days is taken as the test set.

3.1. Passenger Load Factor Prediction of a Single Train Based on LightGBM Algorithm

3.1.1. Model Train. The passenger load factor of a train is predicted based on the LightGBM algorithm and, in comparison, with XGBoost and ARIMA algorithm. During the training in the training set, use LightGBM. `Cv()` function to optimization of 10-fold cross validation parameters; set the learning_rate to 0.01 and adjust the parameters {n_estimators, num_leaves, bagging_fraction, bagging_freq, feature_fraction, max_bin, min_data_in_leaf, lambda_11, lambda_12, min_split_gain, max_depth} in turn; finally, a

set of optimal parameters is obtained, and then, fine adjustment is made. MAE is used as the index of performance evaluation in the training process:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (1)$$

where the actual value of train passenger load factor is y_i , and \hat{y}_i is the forecast value.

3.1.2. Result Analysis. After the optimal parameters are trained by the training set model, the visual fitting process is shown in Figure 3 and compared with XGBoost algorithm.

It can be seen from Figure 3 that the model cannot be well identified and fitted at the mutation point of the passenger load factor, but it can fit other relatively stable points well. Meanwhile, considering that the passenger load factor sequence is a kind of time sequence, in order to further verify the effectiveness of the model constructed in this paper, the ARIMA model is selected for the comparison test.

For the ARIMA (p, d, q) model, through ADF unit root test, Ljung-Box test, ACF chart of autocorrelation coefficient, and PACF chart of partial autocorrelation coefficient combined with AIC and BIC minimum as the target order, ARIMA $(7, 8)$ is determined as the final model, and the visualization of the fitting of the model is shown in Figure 4, where the red sequence is the true values and the yellow sequence is the fitting values.

The trained model is used to predict the test set of LightGBM and XGBoost and rolling prediction ARIMA, and then, the MAE of three models in the training set and test set is obtained, as shown in Table 2. Therefore, LightGBM has the best prediction performance; the ARIMA model has the worst fitting, the lowest prediction accuracy, and the rolling prediction needs to increase the actual value in each step and then retrain, which is not suitable for multistep prediction. The LightGBM and XGBoost prediction results are compared with the true values, as shown in Figure 5.

The LightGBM model is used to predict the passenger load factor of the selected train and visualize the importance of its features, as shown in Figure 6.

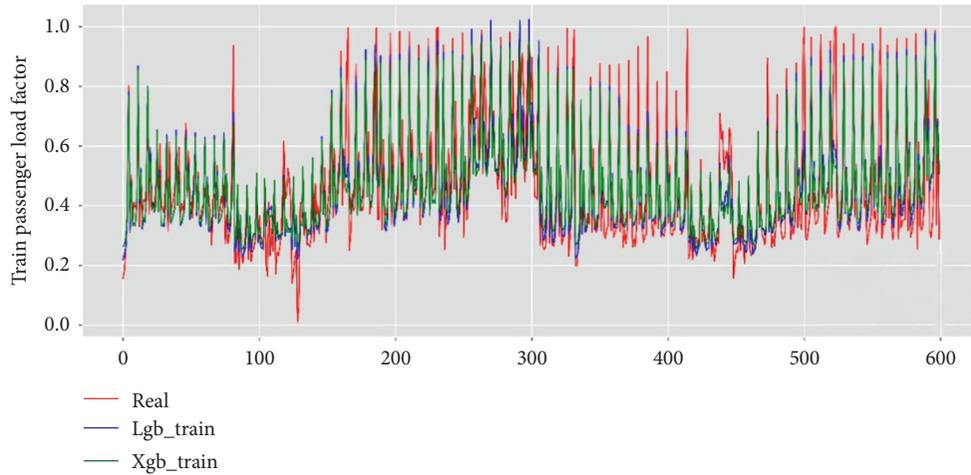


FIGURE 3: Comparison between LightGBM and XGBoost.

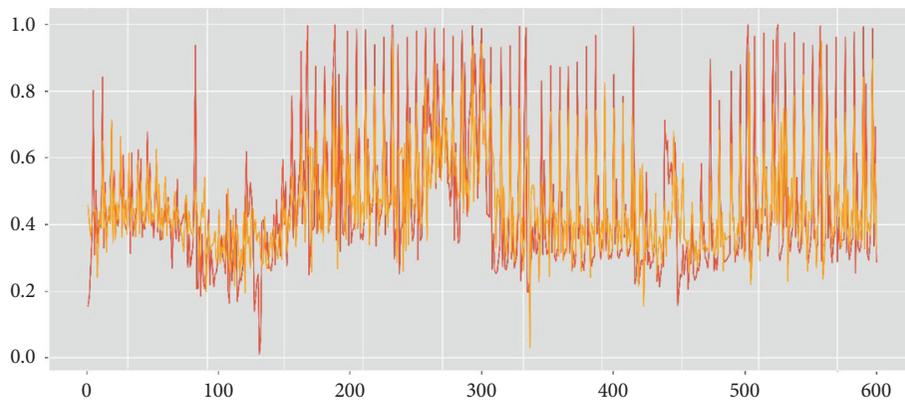


FIGURE 4: Comparison of ARIMA model predicted value and real value.

TABLE 2: Performance comparison of LightGBM, XGBoost, and ARIMA models.

Algorithm	Train set MAE	Test set MAE
LightGBM	0.071622	0.082692
XGBoost	0.080532	0.093274
ARIMA (7, 8)	0.113674	0.135241

It can be seen from Figure 6 that the characteristics of the passenger rate of the train are sorted by importance are WeekOfYear, DayOfWeek, Avgtemperature, Day, Weather, Month, and AQIlevel (air quality level), of which the most important is WeekOfYear and the least important feature is AQIlevel.

3.2. Group Train Passenger Load Factor Prediction Based on LightGBM Algorithm. Before the prediction of the passenger load factor of group train, the histogram of train passenger load factor to be predicted is drawn to check the distribution of its value. The x -axis represents the 0 to 1 passenger load factor divided into 100 cells and the y -axis represents the statistics for the cells, as shown in Figure 7; according to the histogram, the passenger load factor of the sample is a long tail distribution, and it has a large imbalance.

LightGBM algorithm can set the parameters of data acquisition in the course of training, compared with other traditional machine learning algorithms; it ensures that the data acquisition of training keeps the original proportion, which is more suitable for dealing with the issue of unbalanced sample distribution. LightGBM algorithm has designed the parameter `class_Weight`, passing the value “balanced” to this parameter, and it will automatically calculate various weights according to the classification label value. The problem of unbalanced sample distribution can be adjusted, which is helpful to the convergence of training when the samples are unbalanced. Therefore, to ensure that the passenger load factor of group train is predicted under small error, the prediction of passenger load factor of the group train with 10 classifications is constructed in [7], as shown in Figure 8, and good classification results are obtained.

3.2.1. Model Train. The data processed by feature engineering are divided into the training set and test set, and the data of test set is the last month. In the given parameter space, the cross-validation function “`LightGBM.cv()`” of LightGBM official website is used to optimize the parameters. The parameters that need to be optimized for

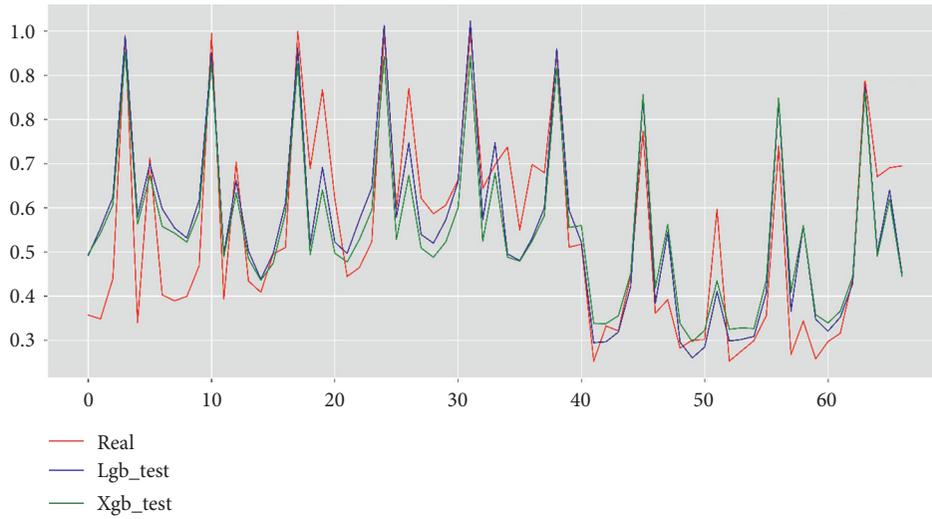


FIGURE 5: Prediction results of LightGBM and XGBoost.

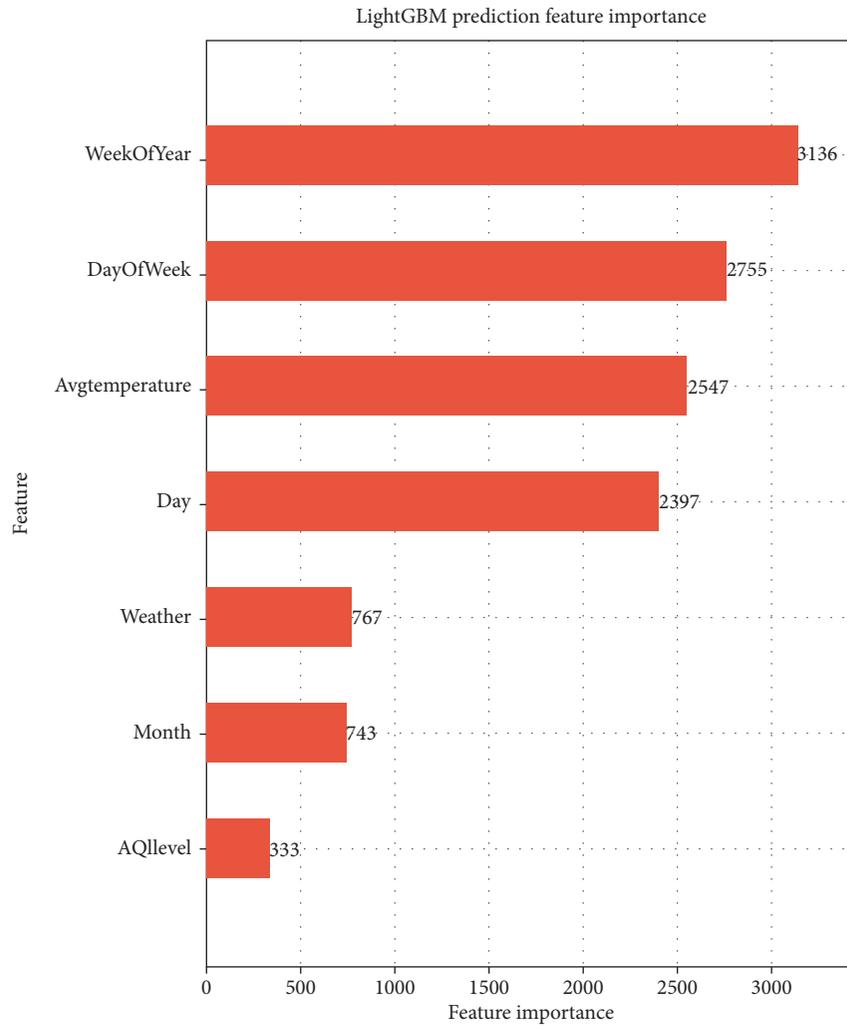


FIGURE 6: Importance of prediction model features.

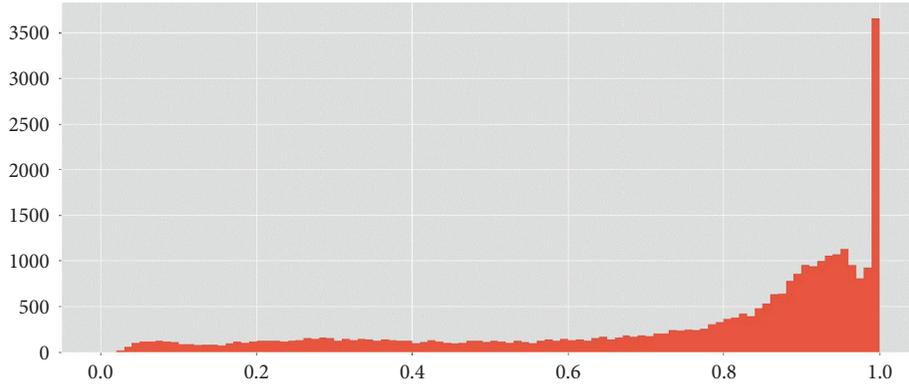


FIGURE 7: Distribution of load factor data.

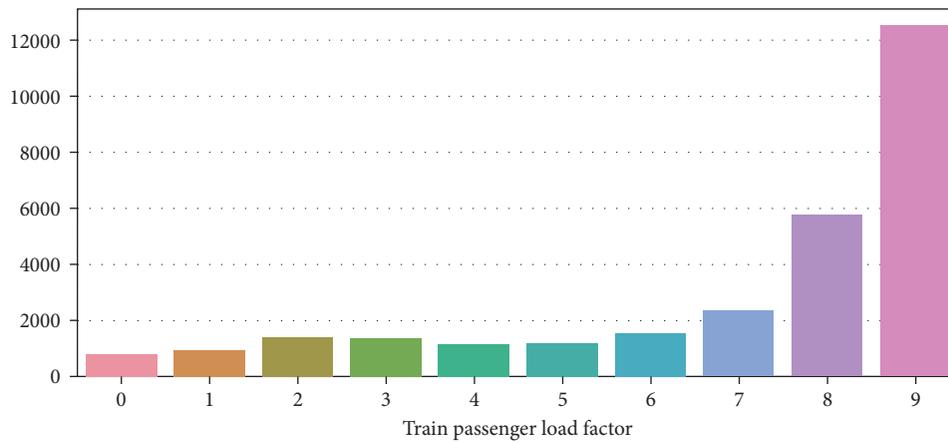


FIGURE 8: 10 classifications of load factor data.

LightGBM function are {"num_leaves", "max_bin", "min_data_in_leaf", "feature_fraction", "bagging_fraction", "bagging_freq", "lambda_l1", "lambda_l2", "min_split_gain", "learning_rate"}. After the optimal combination of the parameters is obtained, the optimal model is retrained in the training set, and finally, the model is used to predict the test set.

For binary classification, 1 is used as a positive example and 0 as a negative example, and the four classifications are defined as follows:

- (1) TP (true positive) indicates the number of samples that the actual values and the predictions are positive examples
- (2) FP (false positive) indicates the number of samples that the actual values are negative example but predicted to be positive
- (3) FN (false negative) indicates the number of samples that the actual values are positive example but predicted to be negative
- (4) TN (true negative) indicates the number of samples that the actual values and the predictions are negative examples

For multiclass classification, suppose there are i categories. A similar dichotomous confusion matrix approach

can be used to obtain TP, FP, FN, and TN, for each category, which are recorded as TP_i , FP_i , FN_i , and TN_i , respectively. And, for the multiclass classification model, different issues have different evaluation indexes. In the official document of machine learning classification evaluation, Sklearn. metrics. `f1_score` set the parameter `average = "weighted"`, which can address the problem of unbalanced sample evaluation indexes in multiclassification.

It is denoted as Weighted_F1, and the formula is as follows:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}, \quad (2)$$

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}, \quad (3)$$

$$\text{Precision}_w = \frac{\sum_{i=1}^L \text{Precision}_i \times \omega_i}{L}, \quad (4)$$

$$\text{Recall}_w = \frac{\sum_{i=1}^L \text{Recall}_i \times \omega_i}{L}, \quad (5)$$

$$\text{Weight_F1} = \frac{2 \cdot \text{Precision}_w \cdot \text{Recall}_w}{\text{Precision}_w + \text{Recall}_w}, \quad (6)$$

TABLE 3: The accuracy of LightGBM, XGBoost, and RandomForest forecast classification.

Algorithm	Evaluation index (Weighted_F1)
LightGBM	0.7927
XGBoost	0.7718
RandomForest	0.7160

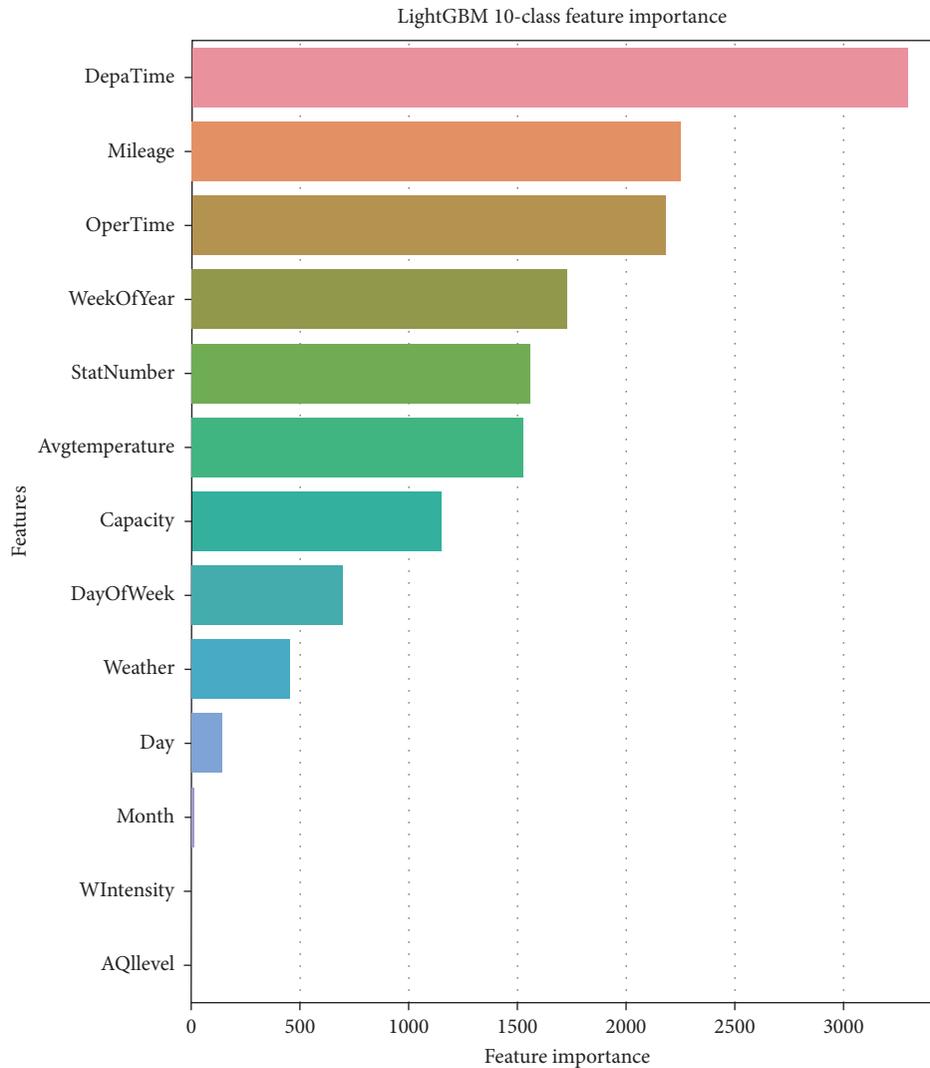


FIGURE 9: The importance of 10 classification model features.

where TP_i , FP_i , FN_i , and TN_i refer to the calculated value of category i in L multiclass classification and ω_i means the proportion of category i in L multiclassifications. Formula (2) means the ratio of the number of TP_i samples to the number of samples predicted to be positive. That is, the accuracy rate of the calculated value of category i is for the prediction result, which means the probability of actually being category i in all the samples predicted as category i . Formula (3) is expressed as the ratio of the number of samples of TP_i to the number of samples of actual positive examples, which means the recall rate. The precision is for the prediction results, which means the probability of actually being category i among all the

samples predicted as category i . Formula (4) is expressed as the ratio of the weight of the accuracy rate to the multicategory. Formula (5) represents the ratio of the weight of the recall rate to the multicategory. Formula (6) is to solve the problem of sample evaluation index imbalance in multicategory.

3.2.2. Result Analysis. The 10 classification model is based on machine learning LightGBM algorithm; the optimal model is optimized after cross validation and the parameter adjustment. For machine learning, the feature_importance function in LightGBM algorithm can calculate, output, and

visualize the importance of each feature. Simultaneously, to verify the prediction model of the group trains' passenger load factor based on the unbalanced 10 classification samples, the XGBoost and RandomForest algorithm are used to compare the prediction results. In XGBoost algorithm, One-Hot Encoder is needed for the features of the categorical variables, and the parameter tuning of RandomForest and XGBoost algorithm is based on *Python* machine learning GridSearchCV function. The prediction results of this paper are shown in Table 3.

It can be seen from Table 3 that the classification consequence of LightGBM algorithm is the best; visualize the importance of the features of LightGBM optimal model, as shown in Figure 9.

It can be seen from Figure 9 that, in the classification and prediction model of the train passenger load factor, the top five important features are DepaTime, Mileage, OperTime, WeekOfYear, and StatNumber, in which DepaTime, Mileage, OperTime, and StatNumber are the train attribute features and WeekOfYear is the time sequence feature of the train passenger load factor. In addition, Capacity is the least important in the features of train attributes, Month is insignificant in the time-series characteristics, and WIntensity and AQLlevel have little influence on a train passenger load factor.

4. Conclusion

In this paper, consider the factors such as train attributes, historical weather, and passenger flow time sequence that affect the passenger load factor of high-speed railway trains; a single train passenger load factor prediction model and a group train passenger load factor prediction model based on LightGBM algorithm are constructed for different prediction requirements and compared with XGBoost, RandomForest and ARIMA algorithm; the feasibility and effectiveness of the prediction model constructed in this paper are verified.

By analyzing the importance of the passenger load factor features' output by the machine learning LightGBM algorithm, the influencing factors of passenger load factor of high-speed railway trains in the region can be obtained. For a train, the crucial factors that affect the passenger load factor are WeekOfYear, DayOfWeek, and average temperature, which are the features of passenger flow time sequence. For high-speed railway trains in a certain area, the main factors affecting the passenger load factor are the attributes of the train, followed by departure time, mileage, and operation time.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Key R&D Program of China (2017YFB1200702), National Natural Science Foundation of China (project nos. 52072314 and 71971182), Sichuan Science and Technology Program (project nos. 2020YFH0035, 2020YJ0268, 2020YJ0256, 2020JDRC0032, 2021YFQ0001, and 2021YFH0175), Chengdu Science and Technology Plan Research Program (project nos. 2019-YF05-01493-SN, 2020-RK00-00036-ZF, and 2020-RK00-00035-ZF), and Science and Technology Plan of China Railway Corporation (project no. P2018T001 and 2019KY10).

References

- [1] J. F. Xu, T. Tang, J. Yan, and Z. Liu, "Prediction of short-term traffic flow based on ensemble learning mechanism," *Journal of Transportation Systems Engineering and Information Technology*, vol. 16, no. 04, pp. 185–190+198, 2016.
- [2] M. Zhang, X. Fei, and H. Liu Zhen, "Short-term traffic flow prediction based on combination model of xgboost-lightgbm," in *Proceedings of the 2018 international conference on sensor networks and signal processing (SNSP)*, pp. 322–327, IEEE, Xi'an, China, October 2018.
- [3] F. J. Wang, F. J. Wang, Y. C. Wang, and C. Bian, "Bus travel time prediction based on light gradient boosting machine algorithm," *Journal of Transportation Systems Engineering and Information Technology*, vol. 19, no. 02, pp. 116–121, 2019.
- [4] Y. Zhang, J. S. Zhu, M. Feng, X. Y. Lu, X. R. Jia, and W. W. Wang, "A study on the model for classifying and predicting occupancy rates of passenger trains," *Rail Way Transport and Economy*, vol. 40, no. 03, pp. 39–45, 2018.
- [5] G. Y. Xu and L. Nie, "Research on prediction method for percentage of passenger seats utilization per multiple unit train," *Comprehensive Transportation*, vol. 37, no. 12, pp. 63–67, 2015.
- [6] C. L. Zhang and Y. P. Bai, "Application of ARIMA time series and BP neural network combination model in railway passenger rate," *Mathematics in Practice and Theory*, vol. 48, no. 21, pp. 105–113, 2018.
- [7] J. S. Jeng-Shyang Pan, W. Qingxiang Feng, Q. X. Lijun Yan, and Y. L. J. Jar-Ferr Yang, "Neighborhood feature line segment for image classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 3, pp. 387–398, 2015.
- [8] G. Fu, G. Q. Han, F. Lu, and Z. X. Xu, "Short-term traffic flow forecasting model based on support vector machine regression," *Journal of South China University of Technology (Natural Science Edition)*, vol. 41, no. 09, pp. 71–76, 2013.
- [9] D. H. Wang, Y. Zhang, and Y. Zhao, "LightGBM: an effective miRNA classification method in breast cancer patients," in *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics (ICCB 2017)*, pp. 7–11, Association for Computing Machinery, New York, NY, USA, October 2017.
- [10] Q. C. Chen, D. Wen, X. Q. Li et al., "Empirical mode decomposition based long short-term memory neural network forecasting model for the short-term metro passenger flow," *PLoS One*, vol. 14, no. 9, 2019.
- [11] M. Gumus and M. S. Kiran, "Crude oil price forecasting using XGBoost," in *Proceedings of the 2017 International Conference*

- on *Computer Science and Engineering (UBMK)*, pp. 1100–1103, IEEE, Antalya, Turkey, October 2017.
- [12] G. L. Ke, M. Qi, T. Finley et al., “LightGBM: a highly efficient gradient boosting decision tree,” in *Proceedings of the Advances in Neural Information Processing Systems 30(NIPS 2017)*, pp. 1–9, San Diego, CA, USA, May 2017.
 - [13] X. L. Sun, M. X. Liu, and Z. Q. Sima, “A novel cryptocurrency price trend forecasting model based on LightGBM,” *Finance Research Letters*, vol. 32, 2020.
 - [14] Z. Tian, G. Sun, D. Chen, Z. Qiu, and Y. Ma, “Method for determining the valid travel route of railways based on generalised cost under the syncretic railway network,” *Journal of Advanced Transportation*, vol. 2020, Article ID 8287648, 12 pages, 2020.
 - [15] M. M. Wang, *Research on Short-Term Traffic Flow Forecasting Method Based on Machine Learning*, Chang’an University, Xi’an, China, 2017.
 - [16] X. C. Dong, T. Lei, S. T. Jin et al., “Short-term traffic flow prediction based on XGBoost,” in *Proceedings of 2018 IEEE 7th Data Driven Control and Learning Systems Conference*, pp. 854–859, IEEE, New York, NY, USA, May 2019.
 - [17] Y. Wang, W. Fang, L. Wang, and B. Xue, “Seating rate prediction for EMU sleeping train based on adaboost-CART model,” *China Railway*, vol. 10, pp. 34–38, 2019.
 - [18] W. Huang, G. Song, H. Hong, and K. Xie, “Deep architecture for traffic flow prediction: deep belief networks with multitask learning,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2191–2201, 2014.
 - [19] Y. E. Liu, Y. Z. Zhang, and C. H. Chen, *Review on Deep Learning in Intelligent Transportation Systems*, Springer, Berlin, Germany, 2020.
 - [20] L. M. Hou and G. F. Ma, “Forecast of railway passenger traffic based on a grey linear regression combined model,” *Computer Simulation*, vol. 28, no. 7, pp. 1–3, 2011.
 - [21] S. Liu, C. S. Yin, D. J. Chen, H. X. Lv, and Q. P. Zhang, “Cascading failure in multiple critical infrastructure interdependent networks of syncretic railway system,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 202114 pages, 2021.
 - [22] Z. Wang, Y. H. Wang, L. M. Jia et al., “The application of improved BP neural network in the prediction of railway passenger volume time serial,” *China Railway Science*, vol. 26, no. 2, pp. 127–131, 2015.
 - [23] G. Y. Xu, *High-speed Train Load Factor Forecast and Breakeven Analysis*, Beijing Jiaotong University, Beijing, China, 2016.
 - [24] L. B. Deng and F. Shi, “Evaluation index system of passenger train operation plan,” *China Railway Science*, vol. 27, no. 3, pp. 106–110, 2006.