



Analysis of on-time arrival prediction and influencing factors of logistics trucks

Xi Xi

701, Unit 3, Building 134, Quanshan District National Pipe Network Storage and Transportation Co., LTD., Xuzhou City, Jiangsu Province

1664564887@qq.com

Abstract. The present study focuses on the 2020 truck transportation data in India, conducting preprocessing operations on the data to construct variables such as driver experience, latitude and longitude, and seasonal time. By building a logistic regression model based on the influencing factors, this study explores the impact of various variables on truck delays. The research model reveals that whether the truck driver is a formal employee, driver experience, and transportation distance significantly affect the likelihood of delays. Finally, this study concludes with suggestions and recommendations for the development of the transportation industry in India.

Keywords: Logistics; Transport; Trucks.

1 Introduction

Transport is a critical infrastructure required for the development process, accounting for a major share of India's energy consumption, especially for petroleum products. Consumption is likely to grow further as economic and population growth, rapid industrialization, urbanization, and agricultural development increase freight and passenger transport, and higher real incomes stimulate leisurely-related tourism. In the short term, at least a decade or so, there are only limited possibilities for fuel conversion. As a result, the industry may need to import a lot of oil, adding to the foreign exchange burden. [8]

Nowadays, several Indian cities are considering how to improve the quality of their transport services to meet the growing demand for urban transport. [2] Although the nature of transport problems varies among developing countries, the current situation of low income and uneven development is still the main problem facing the transport industry in these countries. [7]

As of March 2023, the Ministry of Road Transport and Highways has reported the cost of land acquisition has also increased, with the average rate increasing from 0.92 rupees crore per hectare in 2013 to 313 rupees crore per hectare in 2020. [3] In recent years, India's road transport industry has developed rapidly, but the freight turnover is still only half of the railway. [5]

© The Author(s) 2023

A. Bhunia et al. (eds.), *Proceedings of the 2023 International Conference on Finance, Trade and Business Management (FTBM 2023)*, Advances in Economics, Business and Management Research 264, https://doi.org/10.2991/978-94-6463-298-9_36

2 Variable Selection

This article aggregates a data set called ‘delivery truck trip data’ (year: 2020), and to organize and analyze it, and establish a model to predict the truck-like delays.

The total number of data collected in this data set is 6880, and the study variables covered (in units per transport trip; Most variables are numeric variables, while categorical variables are marked with *) as follows in Table 1:

Table 1. Main Variables Selected in this Study

Real-time position	Destination address
Latitude and longitude of the current location	Total distance traveled
Latitude and longitude of arrival	Departure time
Latitude and longitude of departure point	Time of arrival
The time of arrival agreed upon before the departure of each transport trip	Minimum distance requirement
Departure point	Logistics company name
Temporary employment Act or Contract Act *	Is it on time *
The number and model of the vehicle	No delay *
Scheduled time of trip	Actual time of arrival

3 Empirical analysis

3.1 Data preprocessing

(A) Missing value processing

This paper first analyzes the missing values of the experimental data set and summarizes the missing variables as shown in Table 2 below:

Table 2. Statistical table of missing variables

Variable name	Number of variables	Number of missing value	Missing value ratio
Real-time position	5916	964	14.0%
Longitude of current position	5927	953	13.9%
Current position latitude	5927	953	13.9%
Minimum distance requirement	2820	4060	59.0%

There are four variables with missing values. As some of these variables will not be used in the subsequent analysis, they will not be treated. However, for the variable "Longitude of current position," which is a key variable for the analysis, missing values will be filled using the mean value.

(B) Data distribution characteristics

(1) In this data set, since ‘total distance’ is the most valuable numerical variable, the descriptive statistical variables of the ‘total distance’ variable according to this data trend are shown in Table 3 below:

Table 3. Descriptive statistical variation table of total distance

Statistical variable name	Number	Range	Minimum	Maximum	Mean	Standard deviation	Variance
Value	6880	6879	320	7199	3759.5	1986.2	3945106.7

(2) Research on the distribution of delayed cities: The longitude and latitude distribution characteristics of the departure point and the departure point where the delay occurred were collected, as shown in Figure 1 below. The darker the color of the scatter points, the more the number of delays. According to statistics, the delay rate in this data set is as high as 63%.

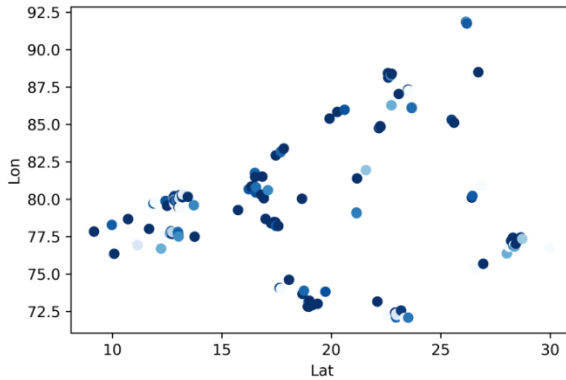


Fig. 1. Departure delay scatter plot

(3) Quality driver screening: This paper defines the ratio of ‘delay times/driving times’ as ‘driver delay percentage’. The lower the ratio, the higher the driver's qualification. A bar chart of the percentage is drawn as shown in Figure 2 below:

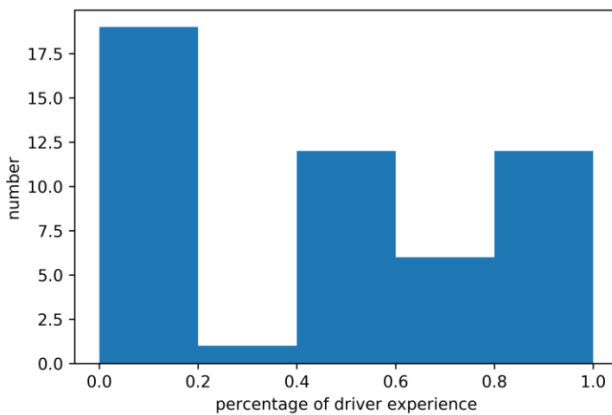


Fig. 2. Plot of the percentage of delays versus the number of drivers

3.2 Logistic regression

(A) Principles

Logistic regression model is used for data classification. The principle involves linear regression equation in linear regression model, whose expression is shown in Formula 1:

$$Y = k_0 + k_1x_1 + k_2x_2 + \dots + k_nx_n \tag{1}$$

Formula 1. Linear regression equation

In addition, the logistic regression model can perform a nonlinear transformation of the output of linear regression to obtain a probability value between 0 and 1, and the probability of predicting 1 can be calculated by the following Formula 2:

$$p = \frac{1}{1+e^{-y}} \tag{2}$$

Formula 2. Function of Sigmoid

Where, y is a linear regression equation. When $P < 0.05$, it indicates that there is a significant correlation between the current independent variable and the dependent variable; otherwise, there is no significant correlation.

(B) Variable selection

In order to improve the accuracy of logistic regression model, it is crucial to select appropriate variables. ‘Delay or not’ is now selected as the dependent variable, and the independent variables of this regression model are shown in Table 4 below:

Table 4. Summary of logistic regression variables

Longitude of departure	Total distance
Latitude of departure	Percentage of delay
Destination longitude	Place of arrival
Destination latitude	Departure point
Temporary worker regulations or contract regulations or not	Driver experience

The explanatory variables listed in Table 4 are explained below:

‘Driver experience’ variable: that is, the number of times the driver's name appeared before the current time sample is calculated by using the for loop statement;

(C) Regression results and analysis (summarized in Table 5 below)

Table 5. Summary of significant results of logistic regression

Variable name	Regression coefficient positive and negative	The value of P
Total distance	0.0002	0.000
Temporary worker regulations or contract regulations or not	1.8853	0.000
Percentage of delay	7.5381	0.000
Driver experience	-0.0537	0.020

Finally, this regression result is explained:

(1) Delay percentage: the regression coefficient is positive, and the regression effect is positive. The results show that driver delay percentage is a static indicator of driver experience, and the higher the value of this indicator, the higher the delay rate.

(2) Driver experience: the regression coefficient is negative, and the regression effect is negative. The results show that the higher the driver experience, the lower the delay probability.

(3) Total distance: the regression coefficient is positive, and the regression effect is positive. The results show that the longer the total distance, the higher the delay probability.

(4) Temporary worker regulations or contract regulations or not: the regression coefficient is positive, and the regression impact is positive. The results show that the probability of transportation delay in contract form is high, while that in temporary form is low.

(D) Regression forecasting

Since the dependent variable 'delay or not' in this logistic regression model is a binary variable, the prediction effect of this independent variable is analyzed.

Each piece of data in the data set is divided into the training set and the test set at the ratio of 8:2, and the prediction accuracy of the training set is 0.7747, and the accuracy of the test set is 0.7647. It can be seen that the accuracy difference between the test set and the training set is small, and the prediction level is relatively stable.

4 Conclusion

Based on the relevant conclusions of the logistic regression conducted in Chapter 3, the following recommendations are made for the development of road transportation in India:

First of all, according to the variable range of 'total distance transported' in the research data and the results of logistic regression, it can be found that the relevant transportation part can formulate more specific road regulations, requiring the minimum distance of one-way transportation to be controlled above 275 kilometers. Therefore, in order to ensure the effectiveness of transportation, reduce the probability of delay, and strengthen the development of domestic transportation industry.

Secondly, in the transportation industry, the two forms of employment of temporary workers and regular employees who sign labor contracts should be reasonable checks and balances. As shown in the results of Chapter 3, although temporary drivers are employed by specific transport companies, they are under pressure to maintain long-term employment opportunities and are more inclined to achieve high performance in order to obtain lifelong employment rights. Therefore, the Indian government should make efforts in the corresponding employment policy, so that transport companies can properly adjust the proportion of employees to be transferred to the regular, and expand the benefits and benefits of temporary workers. In addition, regular employees should be regularly evaluated and screened for their performance and experience, and the salary increase system and dismissal system should be rectified according

to the performance within a specific period, so as to mobilize the enthusiasm of regular employees, so as to reduce the probability of delay and create more profits for the company.

Finally, when it comes to India's transport sector, the relevant authorities should appropriately rectify the more backward means of transport such as trailers, improve their technology, explore new technologies to replace them, or seek international cooperation. For example, road infrastructure cooperation has now become a priority area of China-India transport infrastructure cooperation. ^[1] In early 2005, for example, India conceived a plan to build a dedicated freight corridor to separate freight and passenger traffic on the busiest routes, further reducing freight unit costs and increasing rail's share of the freight market. ^[4] In addition, India is also involved in the International North-South Transport Corridor project, which includes a multi-modal transport network of shipping, rail and road, which can effectively promote connectivity and trade between India and Russia and even Northern Europe, and can integrate into other transport networks in the region to form a new hub. ^[6]

Reference

1. Yang Wenwu, Jia Jia. (2016) China-India Transport infrastructure cooperation in the post-financial crisis era. *South Asian Studies Quarterly*, No.164.
2. Yuan C. (2005) Urban Public transport in India. *A Glimpse Abroad*.
3. Zhu Changli. (1996) Transportation in India. *Southeast Asia*.
4. Ranjan Kumar Jain. (2014) India develops rail heavy haul transport network. *Foreign railway vehicles*.
5. Yang Xueying. (2009) Overview of transportation management system in India. *Overseas Windows*.
6. Yang Wenwu, Yi Xin. (2020) Current status and problems of India's land and sea transport infrastructure connectivity to the outside world. *South Asian Studies Quarterly*, No.18.
7. John Pucher, Nisha Korattyswaropam, Neha Mittal, Neenu Ittyerah. (2005) Urban transport crisis in India. *Transport Policy*.
8. Jyoti Parikh, R. Ramanathan. (1999) Transport sector in India: an analysis in the context of sustainable development. *Transport Policy*.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

