Article

Analysis of the Accident Propensity of Chinese Bus Drivers: The Influence of Poor Driving Records and Demographic Factors

Lili Zheng, Xinyu He, Tongqiang Ding *, Yanlin Li and Zhengfeng Xiao

Transportation College, Jilin University, No. 5988 Renmin Street, Changchun 130022, China
* Correspondence: dingtq@jlu.edu.cn

Abstract: Previous studies have shown that bus drivers are a major contributing factor to bus accidents. The aim of this study is to explore the factors that contribute to the presence of accident propensity among bus drivers, as well as the relative importance of each influencing factor and the mechanism of influence. To this end, a C5.0 decision tree model was developed to determine the relative importance as well as rank the importance of the impact of poor driving records and demographic factors on accident propensity, and a binary logistic regression model was developed to analyze the relationship between accident propensity and the different values of each essential influencing factor. Based on our results, we found that: (1) the number of violations had the most significant effect on bus drivers' accident propensity, followed by age, driving age, and number of alarms; (2) violations and alarms are positively related to bus driver accident propensity; age and driving age are inversely related to bus driver accident propensity; and (3) men have a higher accident risk probability than women. This study's findings will help bus companies and traffic management authorities to implement more targeted improvements to their bus driver management programs.

Keywords: bus driver; accident propensity; poor driving record; demographic factor; C5.0 decision tree; binary logistic regression

MSC: 62P30; 62F03

1. Introduction

Public transportation has always been a crucial component of the transportation system as a more practical and effective means of transportation, and ensuring the safety of public transportation operations plays a vital role in maintaining the efficient operation of the urban passenger transportation system. Buses are generally regarded as being safer than other forms of transportation [1–3], but it can be challenging to prevent accidents involving public transportation due to the complexity of urban road traffic. Many serious bus accidents have occurred recently in various nations, raising widespread concern and significant public worries. According to the Federal Motor Carrier Safety Administration, over 60,000 buses are involved in traffic accidents each year in the United States (FMCSA). In China, 1024 bus-related accidents occurred in 2020, resulting in 215 fatalities, and according to the annual report on road traffic accident statistics for 2020, 80 to 90% of accidents were caused by human factors involving bus drivers [4]. To identify bus drivers with a high propensity for accidents and decrease their frequency, transportation researchers and government authorities have turned their attention to studying the influencing factors related to bus driver accident risk. Accident propensity refers to the fact that some drivers are more likely than others to be involved in accidents due to their physiology, psychology, driving skills, or driving habits.

Previous studies that focused on analyzing risk factors associated with the probability of bus accidents and determining which elements would be related to driver accident

propensity have consistently shown that driver demographic factors significantly affect accident probability. Among these demographic factors, age and driving age have been proven by numerous scholars to be important factors influencing drivers' accident propensity. J. G. Strathman et al. analyzed a large amount of operator data recovered using ITS technology and concluded that both driver age and increasing driving age have a negative impact on the likelihood of an accident [5]. S. Das established a driver accident prediction model to analyze driver human factors and proposed that male drivers aged 15 to 34 are the most likely to be involved in accidents [6]. F. Khanehshenas discovered a u-shaped relationship between driver fatigue and driver age after conducting 14 in-depth interviews with bus drivers and using qualitative content analysis and thematic analysis [7]. Younger and older drivers were more likely to be fatigued and, consequently, more likely to be in an accident. L. Dorn coded the accident data and used direct and induced indirect analyses to conclude that bus drivers are significantly more likely to have accidents when younger [8]. When a longer driving age is considered, age significantly impacts a driver's propensity for accidents, with younger and older drivers having a higher accident risk. A mixed logit model built by K. Goh revealed that drivers 60 years of age or older and those with less than two years of experience behind the wheel had a higher accident probability [9]. J. Huting and J. Reid used a random forest model to determine the factors influencing an accident's likelihood and concluded that older and younger drivers were more likely to be involved in accidents [10]. Most scholars conclude that the longer the driving age, the less likely the driver is to be involved in an accident, but some scholars suggest the opposite. Li et al. analyzed accident data from several countries around the world and concluded that drivers with less than three years of driving age are more likely to be involved in accidents [11], which is consistent with other studies, but that drivers with more than ten years of driving experience are equally likely, if not more likely, to be involved in accidents. According to this researcher, drivers with more than ten years of driving experience have a higher accident rate due to overconfidence in their driving abilities and overestimating their ability to handle emergencies. P. P. Jovanis et al. examined data from over 1800 accidents in the Chicago metropolitan area [12]. They proposed that increasing the driving age reduces the likelihood of accidents, but in the overall data, drivers with 3 to 6 years of driving age are the most likely group to be involved in accidents.

Whether and how gender affects a driver's accident risk has been a topic of study, and some researchers believe that male drivers are more prone to collisions. A. I. Glendon et al. recruited 102 drivers to drive in various scenarios and concluded [13], after analyzing data obtained from driving simulators, that male drivers are more likely to exceed the posted speed limit. Chen and Huei-Yen Winnie surveyed 578 drivers using the Susceptible Driver Distraction Questionnaire (SDDQ) and the Manchester Driver Behavior Questionnaire (DBQ) [14]. Male drivers are more likely than female drivers to be distracted while driving, resulting in accidents, according to the study. However, male drivers have greater perceptual control and superior driving ability than female drivers. According to some researchers, female drivers are more likely to be involved in an accident. D. Tao and R. Zhang conducted a questionnaire survey of 200 drivers and found that in terms of driving behavior and risk of causing accidents, there was little difference between men and women, but female drivers were more likely to commit traffic violations [15].

Despite the difficulty of obtaining drivers' poor driving records, a certain number of studies have examined the effect of poor driving records on the incidence of driver accidents. J. X. Zhang et al. used the Driving Skills Inventory (DSI) and the Driver Behavior Questionnaire (DBQ) to conduct a random survey of drivers and [16], after obtaining a valid sample of 100 drivers, concluded that drivers' violations and faults significantly affect driving performance, and the more violations and faults, the worse the driving performance. M. M. Hamed developed a convex-concave model to analyze the personal information [17], traffic violations, and traffic accident data of 438 minibus drivers. Results indicated that speeding and driving violations could contribute to severe accidents, and the greater the number of violations, the greater the likelihood of severe accidents.

Decision tree models and logistic regression models have been introduced to analyze the relationship between driver factors and the occurrence of accidents. Azhar et al. analyzed the factors influencing the occurrence of accidents among heavy vehicle drivers in Malaysia by building a CART decision tree model and concluded that the age of the driver is an important factor influencing the occurrence of accidents [18]. Taamneh et al. collected a total of 400 questionnaires filled by drivers with different socioeconomic backgrounds and analyzed them using the ID3 decision tree model [19]. The results revealed that driving experience, marital status, age, and educational background can influence a driver's comprehension of traffic signs and are therefore important factors in the occurrence of accidents. Rejali developed a logistic regression model to examine the connection between driving behavior and the degree of correlation between the frequency of accidents and traffic fines [20], and discovered that aggressive and frequent violations were significant predictors of taxi drivers falling into the high accident risk category, while negligent errors and aggressive violations were significant predictors of falling into the medium/high-risk category. Lukongo applied a disorderly multivariate logistic model to investigate the causes of serious crashes in Louisiana, and the study discovered that the driver's gender and age had an impact on traffic safety [21].

However, as the relevant studies cited above, while the relationship between poor driving records and accident occurrence has been studied, the data sets used are primarily single violation data or questionnaire data with low veracity due to excessive subjectivity. These data do not allow an accurate analysis of other critical poor driving records, such as driver status irregularities. Furthermore, few studies have combined bus drivers' poor driving records with demographic factors to analyze the relative importance of each factor in influencing drivers' accident propensity and the relationship between the values assigned to each factor and accident propensity. Due to the relatively limited previous study, bus companies and traffic management departments could not accurately identify drivers' accident propensity and therefore did not achieve the desired safety management outcomes. Suppose the impact of driver violations, alarms, and demographic factors on accidents was better analyzed and comprehended. In that case, bus companies' training and education programs could be more targeted and implemented on the most important factors rather than expending effort considering each driver's factor. The identification and association analysis of the most significant elements influencing the accident probability of bus drivers might assist bus company managers in enhancing their safety management policies and driver training. In this study, we used the accurate lousy driving record data (number of violations, number of alarms) obtained from the real-time monitoring of a bus company in Chongqing, China, combined with data on driver demographic factors (age, driving age, gender, nationality, political background, and education), and established a C5.0 decision tree model to determine the relative importance between a bad driving record and demographic factors on the accident propensity of bus drivers and to rank the importance of these factors. A binary logistic regression model was established to determine the relationship between accident propensity and the value of each influencing factor.

2. Data

2.1. Data Source

The data in this study comes from a bus company in Chongqing, China, and consists of four sub-datasets: the bus driver demographic dataset, bus accident dataset, bus driver alarm dataset, and bus driver violation dataset. The bus driver demographic dataset records the personal information of each driver in the bus company, including age, gender, marital status, education, nationality, political background, initial license date, license expiration date, and other demographic information. The bus accident dataset records all accident data for vehicles operated by this bus company from 1 January 2022 to 30 June 2022, including the accident date, accident driver's name, accident location, accident cause, accident pattern, and other accident-related characteristics. The bus driver alarm dataset records information on each abnormal status alarm of the driver of the bus company from 1 January 2022 to

30 June 2022, including the driver's name, alarm type, alarm level, alarm date, and the occurrence time of each alarm. The bus driver violation data set records the information of each violation that occurred by the driver of this bus company, including the name of the violating driver, violation date, violation time, violation item, and violation behavior.

2.2. Data Processing

Data not related to driver accidents, such as the expiration date of a driver's license, and data that could not be interpolated because of a high number of missing (98% missing) data, such as marital status, were deleted from the four sub-datasets of the bus driver demographic dataset, bus accident dataset, bus driver alarm dataset, and bus driver safety violation dataset. After removing unnecessary information and a significant amount of missing data, the overall data integrity is high, and no further interpolation is needed. As shown in Section 2.3, nine variables were ultimately analyzed by sorting and filtering the data on bus driver demographics, bus accidents, bus driver alarms, and bus driver violations.

2.3. Data Description

As shown in Table 1, the data set included 4925 bus drivers, broken down by age, driving age, gender, political background, nationality, education, number of accidents, number of alarms, and number of violations. Values were assigned to each variable.

Table 1. Descriptive statistics.

Variables	Value	Categories	Count	Percentage
Age	1	20–25	76	1.54%
	2	26–35	1085	22.03%
	3	36–45	1607	32.63%
	4	46–55	1995	40.51%
	5	56–60	162	3.29%
Driving age	1	<2	161	3.37%
	2	2–5	666	13.52%
	3	6–9	1097	22.27%
	4	10–14	1909	38.76%
	5	15–20	728	14.78%
	6	>20	364	7.39%
Gender	1	Male	4524	91.86%
	2	Female	401	8.14%
Political background	1	The masses	4369	81.71%
	2	League member	246	4.99%
	3	Party member	310	6.30%
Nationality	1	Han nationality	4912	99.74%
	2	Minority nationality	13	0.26%
Education	1	Junior high school and below	1405	28.53%
	2	Technical school	247	5.02%
	3	Secondary school	981	19.92%
	4	Senior middle school	1814	36.83%
	5	Vocational high school	54	1.10%
	6	Undergraduate	99	2.01%
	7	Junior college	325	6.60%
Accident occurrence	1	Yes	3407	69.18%
	2	No	1518	30.82%

Table 1. *Cont.*

Variables	Value	Categories	Count	Percentage
Alarm	1	0	2202	44.71%
	2	1–15	2466	50.07%
	3	16–30	218	4.43%
	4	>30	39	0.79%
Violation	1	0	3419	69.42%
	2	1	976	19.82%
	3	>1	530	10.76%

2.3.1. Number of Violations Analysis

It is considered a violation when a bus driver disobeys traffic laws and bus driving safety regulations. The violation record can indicate whether or not the driver is prone to accidents and the strength of their propensity.

The distribution of violation is shown in Table 2. During the data collection period, 1536 drivers had violations, accounting for 30.58%; of which, 976 had only one violation, accounting for 19.82% and 530 had multiple safety violations, accounting for 10.76%. The drivers who violated safety the most had an overall total of eight violations.

Table 2. Violation distribution.

Violation	Count	Percentage	Number of Accident Drivers	Percentage of Accident Drivers
0	3419	69.42%	616	18.02%
1	976	19.82%	328	33.61%
>1	530	10.76%	180	33.96%

The proportion of accident drivers who have violated the rules is nearly twice that of those who have never violated the rules. Bus drivers who have never broken any rules or regulations are less likely to be involved in an accident because they drive more cautiously and maintain good driving habits; this proportion of accident drivers is only 18.02%.

2.3.2. Number of Alarms Analysis

The alarm is a device installed by the bus company to detect the driver's state. When the driver is in an abnormal state, such as fatigue driving, distracted driving, or emotional driving, the system will sound an alarm to remind the driver and record it in the database. The number of alarms can, to some extent, reflect the driver's daily work, driving habits, and driving status, as well as the driver's proclivity for accidents.

The distribution of alarm is shown in Table 3. During the data collection period, 2723 drivers received alarms, accounting for 55.29% of all drivers. The majority (50.07%) received alarms 1 to 15 times; a small number of drivers received more alarms, 4.43% of the total number of drivers received alarms 16 to 30 times, and only 0.79% of the total number of drivers received alarms more than 30 times.

Table 3. Alarm distribution.

Alarm	Count	Percentage	Number of Accident Drivers	Percentage of Accident Drivers
0	2202	44.71%	548	24.89%
1–15	2466	50.07%	508	20.60%
16–30	218	4.43%	57	26.15%
>30	39	0.79%	11	28.21%

The driver with the highest percentage of accidents is the one with more than 30 alarms. The high number of alarms suggests that the driver's mental state is typically unstable while driving and is susceptible to changes brought on by outside events. This abnormal state can easily result in accidents.

2.3.3. Age Analysis

The distribution of age is shown in Table 4. Bus drivers range in age from 22 to 60, with the majority in their middle years. Only 76 are under 25, 1.5% of the total, and 162 are over 56, making up only 3.3%. Bus vehicle driving work is tedious for young people, and the pay is low, making it less appealing to them; on the other hand, because a bus driver must have some driving experience, young people often have less experience and cannot meet the recruitment requirements of public transportation enterprises. A bus driver's long hours and heavy workload may not be suitable for older individuals, resulting in a decrease in the number of older individuals who hold the position.

Table 4. Age distribution.

Age	Count	Percentage	Number of Accident Drivers	Percentage of Accident Drivers
20–25	76	1.54%	25	32.89%
26–35	1085	22.03%	323	29.77%
36–45	1607	32.63%	367	22.84%
46–55	1995	40.51%	379	19.00%
56–60	162	3.29%	20	12.35%

The greatest proportion of accident drivers is aged 20 to 25, and the percentage of accident drivers gradually decreases as they get older. Drivers aged 20 to 25 are more impulsive and inexperienced, and they have not developed good driving habits, which leads to accidents.

2.3.4. Driving Age Analysis

The distribution of driving age is shown in Table 5. The driving ages of the drivers are widely dispersed, ranging from novice drivers who have been licensed for less than a year to senior drivers who have been behind the wheel for up to 36 years. The most significant number of drivers have been driving for 6–10 years and 10–15 years, accounting for 22.3% and 38.8%, respectively, and accounting for more than half of all drivers. Drivers with less than two years of driving experience and those with more than twenty years of driving experience account for 3.3% and 7.4%, respectively.

Table 5. Driving age distribution.

Driving Age	Count	Percentage	Number of Accident Drivers	Percentage of Accident Drivers
<2	161	3.27%	57	35.40%
2–5	666	13.52%	208	31.23%
6–9	1097	22.27%	258	23.52%
10–14	1909	38.76%	407	21.32%
15–20	728	14.78%	136	18.68%
>20	364	7.39%	58	15.93%

Accident drivers with less than two years of driving age accounted for the highest proportion of accidents, 35.40%. As the driving age gradually rises, drivers become more cognizant of driving skills and road conditions, gain more driving experience, and are involved in fewer accidents.

2.3.5. Gender Analysis

The distribution of gender is shown in Table 6. The difference between the number of male and female drivers is significant, with only 8.1% being female. The reason is the same as why the number of older drivers is less: the bus driving job requires long hours and a demanding workload; women are physically less qualified than men for this position [22].

Table 6. Gender distribution.

Gender	Count	Percentage	Number of Accident Drivers	Percentage of Accident Drivers
Male	4524	91.86%	1051	23.23%
Female	401	8.14%	73	18.20%

Men have a greater proportion of accidents than women; male drivers are more likely to be emotional and experience road rage more frequently, whereas women are more meticulous in their behavior and more cautious when driving [23].

2.3.6. Political Background Analysis

The distribution of political background is shown in Table 7. The masses comprised 88.71% of the drivers, the largest proportion, followed by 6.30% of party members and 4.99% of league members. The proportion of accident drivers whose political background is party member is the lowest at 16.45%.

Table 7. Political background distribution.

Political Background	Count	Percentage	Number of Accident Drivers	Percentage of Accident Drivers
The masses	4369	88.71%	985	22.55%
League member	246	4.99%	88	21.95%
Party member	310	6.30%	51	16.45%

2.3.7. Education Analysis

The distribution of education is shown in Table 8. Drivers had the highest percentages of secondary school, junior high school and below, and senior middle school education, accounting for 19.92%, 28.53%, and 36.83%, respectively. The low salary and heavy workload of bus drivers make it difficult to attract workers with a bachelor's degree or higher.

Table 8. Education distribution.

Education	Count	Percentage	Number of Accident Drivers	Percentage of Accident Drivers
Junior high school and below	1405	28.53%	353	25.12%
Technical school	247	0.05%	75	30.36%
Secondary school	981	19.92%	385	39.25%
Senior middle school	1814	36.83%	548	30.21%
Vocational high school	54	1.10%	19	35.19%
Undergraduate	99	2.01%	18	18.18%
Junior college	325	6.60%	94	28.92%

Except for drivers with a bachelor's degree, the percentage of accident drivers with other degrees is greater than 20%. Generally speaking, drivers with a bachelor's degree have a more extensive education and knowledge base than other drivers with less education, are more strict about following traffic laws and regulations, have a better understanding of traffic laws and regulations, signs, and markings, and are less likely to be in accidents.

2.3.8. Number of Accidents Analysis

The distribution of accident is shown in Table 9. There were 1124 accidents during the data recording time, accounting for 22.82% of the 4925 drivers, and 266 had multiple accidents during the data recording time, accounting for 5.40%, with the drivers who had the most accidents having a total of 6 accidents.

Table 9. Accident distribution.

Accident Occurrence	Count	Percentage
0	3801	77.18%
1	858	17.42%
>1	266	5.40%

2.4. Factors

According to previous related studies [24,25], the risk of a driver having an accident is generally classified into two categories, one in which the driver will not have an accident and the other in which the driver will have an accident. Combining the above theory and analyzing the currently available data, this paper divides drivers' accident propensity into two categories: no accident propensity (indicating that they will not have accidents) and accident propensity (indicating that they will have accidents or even multiple accidents) to analyze the factors influencing drivers' accident propensity.

A Pearson correlation analysis was carried out for each independent variable to prevent the issue of covariance among the independent variables, as shown in Table 10. The correlation between age and driving age (0.608) had the highest absolute value but was still less than 0.7, indicating that there is no covariance issue with the independent variables, so age, driving age, nationality, political background, education, number of alarms, number of violations, and gender were chosen as independent variables.

Table 10. Correlations of independent variables.

Independent Variables	Age	Driving Age	Nationality	Political Background	Education	Alarm	Violation	Gender
Age	1							
Driving age	0.608 ** (0.000)	1						
Nationality	−0.013 (0.366)	−0.012 (0.386)	1					
Political background	−0.020 (0.157)	−0.045 ** (0.002)	0.031 * (0.027)	1				
Education	−0.329 ** (0.000)	−0.217 ** (0.000)	−0.008 (0.598)	0.135 ** (0.000)	1			
Alarm	−0.022 (0.115)	−0.028 (0.052)	0.013 (0.356)	0.022 (0.121)	0.026 (0.070)	1		
Violation	−0.075 ** (0.000)	−0.065 ** (0.000)	−0.026 (0.073)	−0.065 ** (0.000)	0.031 * (0.030)	0.123 ** (0.000)	1	
Gender	−0.164 ** (0.000)	−0.090 ** (0.000)	−0.015 (0.083)	0.000 (0.990)	0.125 ** (0.000)	−0.023 (0.109)	−0.074 ** (0.000)	1

**: Correlation is significant at the 0.01 level (2-tailed). *: Correlation is significant at the 0.05 level (2-tailed).

3. Methods

In this study, we developed a C5.0 decision tree model to rank the relative significance of poor driving records and demographic factors on the accident propensity of bus drivers. We then developed a binary logistic regression model to analyze the relationship between the values of each significant factor and the occurrence of accidents to determine the likelihood of bus drivers' accidents under the values of each factor.

3.1. C5.0 Decision Tree

3.1.1. C5.0 Decision Tree Model

The decision tree model, named for its tree-like structure diagram, employs visual graphs and decision models in the form of tree branches to aid decision-making and, ultimately, to determine the most likely strategy to achieve a goal. The decision tree model does not rely on knowledge from other domains or parameter assumptions. It can clearly demonstrate the importance of features by generating different, simple feature classification rules. Depending on the level, the decision tree's structure is divided into three categories: root nodes, internal nodes, and leaf nodes [26]. The decision tree's root node is a collection of data samples, and the highest feature attribute is chosen by calculating the information gain ratio for each element in the collection. Then it becomes a child node of the current node, and the node is used as the root node to continue splitting until the classification ends when all attributes are grouped into the same category. Contrary to the classical CART decision tree, a method for creating binary trees that can only produce two branches per node, the C5.0 decision tree can generate a different number of branches per node. The C5.0 decision tree model was introduced into the bus driver accident propensity analysis based on the types of independent variables used in this study to determine the elements in the driver data that have a greater bearing on accident propensity.

The independent variable X has m values in the data set, denoted X_i, $i = 1, 2, \ldots, m$. The dependent variable Y has n values, denoted Y_j, $j = 1, 2, \ldots, n$; it divides the dataset into n subsets, denoted V_j, $j = 1, 2, \ldots, n$.

The total number of samples in the data set is T, the number of independent variables $X = X_i$ is C_i, the number of dependent variables $Y = Y_j$ is D_j, and the number of independent variables $X = X_i$ under the dependent variable $Y = Y_j$ is $X_i Y_j$.

The independent variable $X = X_i$ probability of occurrence is:

$$P(X_i) = \frac{C_i}{T} \tag{1}$$

The dependent variable $Y = Y_j$ probability of occurrence is:

$$P(Y_j) = \frac{D_j}{T} \tag{2}$$

Moreover, the independent and dependent variable probabilities satisfy the following:

$$\sum_{i=1}^{m} P(X_i) = \sum_{j=1}^{n} P(Y_j) = 1 \tag{3}$$

The conditional probability of occurrence of the independent variable X_i under the condition that the dependent variable $Y = Y_j$ occurs is:

$$P(X|Y) = \frac{X_i Y_j}{D_j} \tag{4}$$

The information entropy reflects the magnitude of the random uncertainty of the data within the decision tree nodes [27]. C5.0 decision trees use the rate of decline of information entropy to determine the optimal branching variables and segmentation thresholds [28].

The information entropy of the independent variable X is expressed as:

$$H(X) = -\sum_{i=1}^{m} P(X_i) \log_2 P(X_i) = -\sum_{i=1}^{m} \frac{C_i}{T} \log_2 \frac{C_i}{T} \tag{5}$$

The information entropy of the dependent variable Y is expressed as:

$$H(Y) = -\sum_{j=1}^{n} P(Y_j) \log_2 P(Y_j) = -\sum_{j=1}^{n} \frac{D_j}{T} \log_2 \frac{D_j}{T} \tag{6}$$

Based on the dependent variable Y, the data set is partitioned, and the conditional information entropy of each variable X is expressed as:

$$H(X|Y) = -\sum_{i=1}^{m} P(X|Y) \log_2 P(X|Y) = -\sum_{i=1}^{m} \frac{X_i Y_j}{D_j} \log_2 \frac{X_i Y_j}{D_j}, \ j = 1, 2, \ldots, n \tag{7}$$

The information gain reflects the degree of information entropy reduction in the information transfer within the decision tree, which is expressed as:

$$Gain(X, Y) = H(X) - H(X|Y) = -\sum_{i=1}^{m} \frac{C_i}{T} \log_2 \frac{C_i}{T} + \sum_{i=1}^{m} \frac{X_i Y_j}{D_j} \log_2 \frac{X_i Y_j}{D_j}, \ j = 1, 2, \ldots, n \tag{8}$$

The information gain rate is the criterion for evaluating the selection of independent variables and the division of dependent variables when the structure of a decision tree is diffused from its root node to its leaf nodes. The direction of data selection with the largest information gain rate is the direction of the selection of independent variables and the division of dependent variables. The information gain rate is expressed as:

$$Gain_{\text{ratio}}(X, Y) = \frac{Gain(X, Y)}{H(Y)} = \frac{\sum\limits_{i=1}^{m} \frac{C_i}{T} \log_2 \frac{C_i}{T} - \sum\limits_{i=1}^{m} \frac{X_i Y_j}{D_j} \log_2 \frac{X_i Y_j}{D_j}}{\sum\limits_{j=1}^{n} \frac{D_j}{T} \log_2 \frac{D_j}{T}}, \ j = 1, 2, \ldots, n \tag{9}$$

3.1.2. Parameter Tuning

Pruning severity

Pruning severity indicates the degree of pruning of the decision tree, which is expressed as:

$$V = 100(1 - C_F) \tag{10}$$

where C_F denotes the confidence level of the error, and the optimal value is chosen by adjusting the pruning severity and comparing the resulting models. To ensure the model's overall accuracy, the minimum classification error rate value is chosen as the selection criterion. In this study, a pruning severity index of 45 was selected to prune the decision tree after several experimental comparisons to ensure the model's accuracy and avoid the phenomenon of overfitting.

Misjudgment costs

The value of the misjudgment cost reflects the severity of the consequences of a misjudgment; the higher the value, the more severe the consequences. The dependent variables specified in this paper are no accident propensity (indicating that no accident will occur) and accident propensity (indicating that multiple accidents will happen). The difference between accident propensity and no accident propensity is substantial, and the consequences of misjudging a driver's accident propensity as no accident propensity are more severe, thereby increasing the cost of miscalculation. In this study, the misclassification cost is set, as shown in Table 11, following multiple experimental comparisons to ensure the model's accuracy.

Table 11. Misjudgment Costs.

	No Accidents Will Occur	Accidents Will Happen or Even Happen Multiple Times
No accidents will occur	0.0	1.3
Accidents will occur or even occur multiple times	1.2	0.0

3.2. Binary Logistic Regression Model

The logistic regression model is a classification model that examines the relationship between classification outcomes and influencing factors and can be expressed as the probability of an outcome given a particular influencing factor. The logistic regression model, which can analyze the influence of one or more factors on the outcome and describe the decision-making behavior of individuals or groups more precisely, carefully, and comprehensively, is essential tool for analyzing individual traffic behavior in the field of road traffic. This helps researchers achieve relatively complex research objectives and produce relatively rich research findings. The binary logistic regression model is utilized in this study to further analyze the link between each factor and bus driver's accident proneness after the C5.0 decision tree model has been used to identify the relative importance of each component on the bus driver's accident proneness. In this paper, X_i (i = 1, 2, ..., n) represents different independent variables, such as age, gender, and other factors to be analyzed. Y is a binary dependent variable with a value of 1 or 0, indicating whether the bus driver has a propensity for accidents.

$$Y = \begin{cases} 0, & \text{If the driver does not have accident propensity} \\ 1, & \text{If the driver has accident propensity} \end{cases}$$

Constructing a binary logistic regression model [29]:

$$P_i = P(Y = 1|X_i) = \frac{\exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)}{1 + \exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)} \tag{11}$$

where P_i is the conditional probability when Y = 1(the driver has accident propensity) and when the independent variable is X_i, where $0 \leq P \leq 1$, i = 1, 2, ..., n.

The odds ratio is used to express the model findings in the logistic regression model, and the odds ratio refers to the change in the probability of occurrence of the dependent variable for each unit change in the independent variable while the remaining variables remain constant [30]. To assist comprehension of the model results, the odds ratio is translated into a standard ratio coefficient given as the ratio of P_i to $1-P_i$, with $1-P_i$ referring to the likelihood that the driver has no accident propensity, and the probability formula is:

$$1 - P_i = \frac{1}{1 + \exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)} \tag{12}$$

The odds ratio can be written as:

$$\frac{P_i}{1 - P_i} = \exp(\beta_0 + \sum_{i=1}^{n} \beta_i X_i) \tag{13}$$

An odds ratio greater than 1 describes a positive connection between the independent variable and the dependent variable; as the value of the independent variable increases, so does the driver's accident likelihood; an odds ratio less than 1 suggests a described association.

4. Results

4.1. C5.0 Decision Tree Model Results

Figure 1 depicts the tree diagram generated by the C5.0 decision tree method. The prediction accuracy of the model achieved 82.61%. Although there are multiple metrics for decision tree evaluation, such as ROC, AUC, recall rate, and precision rate, prediction accuracy in the unbalanced decision tree classification problem has representativeness and increased information [31], so prediction accuracy is used as the evaluation metric.

Figure 1. Schematic diagram of the C5.0 decision tree.

Each rule set of bus driver accident propensity represents each non-closed directed path in the decision tree split from the top-down nodes. For each rule, each independent variable that can be seen to affect the outcome was selected with an observed sample size of at least 5, a proportion (*S*) of at least 0.1%, and a probability (*Pr*) of at least 10% for the occurrence of the accident dependent variable. The final two columns of the table display the proportion of the rule sample (*V*) and the accident incidence probability (*P*) under the rule. As stated in Table 12, the rules are ordered according to their proportion size.

The decision tree rules show that the number of violations exists in all 12 rules and branch out with the number of violations as the first node in the tree diagram, having the most significant impact on the likelihood of bus driver accidents. The second, third, and fourth most relevant factors are age, driving age, and the number of alarms, which are listed in 11 rules, 9 rules, and 8 rules, respectively. The remaining factors appear less frequently in the rules, and at the tree diagram's more terminal nodes, so they have less influence on bus driver accident probability.

According to the C5.0 decision tree model, the number of violations is the most crucial element influencing the accident propensity of bus drivers, followed by age, driving age, and the number of alarms. The impact of education, nationality, gender, and political background on the accident propensity of bus drivers is insignificant, as shown in Figure 2.

Table 12. Accident occurrence rule set.

Num	Age	Driving Age	Nationality	Political Background	Education	Alarm	Violation	Gender	V	Pr
1	2	1				2	2		0.12%	100.00%
2	2	2			5, 7	2	2		0.28%	71.43%
3	1						2		0.47%	69.57%
4	2	2		1		1	2		0.59%	58.62%
5	2	2		1	3	2	2		0.39%	57.90%
6	2	3, 4, 5					2		2.46%	34.71%
7	2	2		2		1	2		0.31%	33.33%
8	2	1			4, 5	1	2		0.12%	33.33%
9	3, 4, 5						2		14.34%	29.04%
10	2	2		2	3	2	2		0.16%	25.00%
11							1, 3		80.18%	20.16%
12	2	2		1	4	3	2		0.12%	16.68%

Column "Num" is the number of the rule.
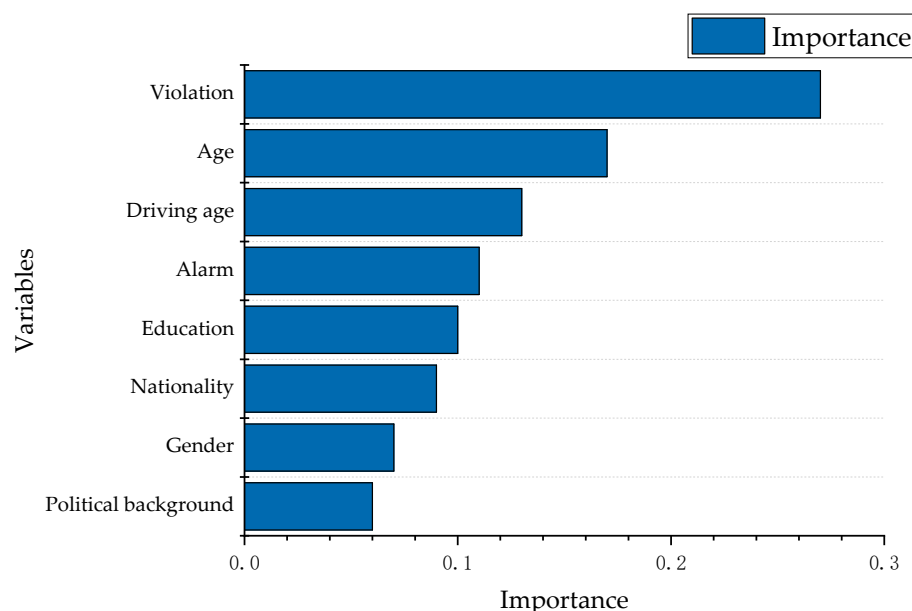
Figure 2. Importance of factors.

4.2. Binary Logistic Model Results

The independent variables were subjected to correlation regression analysis and those that satisfied significance $p < 0.05$ were screened out, as shown in Table 13. Nationality ($p = 0.279$), political background ($p = 0.274$), and education ($p = 0.224$) had a weak significance relationship with the dependent variable, which is also consistent with the results of the aforementioned decision tree model, in which nationality, political background, and education had a smaller impact on drivers' accident propensity.

Table 13. Score test results.

Variables	B	Degrees of Freedom	*p*-Value
Age	−0.228	1	0.000 *
Driving age	−0.099	1	0.012 *
Nationality	1.245	1	0.279
Political background	−0.056	1	0.274
Education	0.027	1	0.224
Gender	−0.383	1	0.006 *
Alarm	−0.204	1	0.000 *
Violation	0.497	1	0.000 *

*: Correlation is significant at the 0.05 level.

A binary logistic regression model was developed for age, driving age, number of alarms, number of violations, and gender, and the relationship between the various values taken within each independent variable and driver accident propensity was analyzed; the results are shown in Table 14. The independent variables with *p*-values less than 0.05 were correlated. Among them, the number of alarms was 1–15 ($p = 0.315$), which was not significantly correlated with accidents.

Table 14. Binary logistic regression parameter estimates and odds ratios.

Variables	Categories	Reference Category	B	Odds Ratio	*p*-Value
Age	26–35 36–45 46–55 56–60	20–25	−0.484 −0.687 −0.870 −1.443	0.616 0.503 0.419 0.236	0.002 * 0.010 * 0.001 * 0.000 *
Driving age	2–5 6–9 10–14 15–20 >20	<2	−0.034 −0.228 −0.276 −0.400 −0.529	0.967 0.796 0.759 0.670 0.589	0.009 * 0.003 * 0.017 * 0.006 * 0.013 *
Violation	1 >1	0	0.641 0.821	1.635 2.272	0.003 * 0.000 *
Alarm	1–15 16–30 >30	0	0.347 0.593 0.805	1.195 1.243 1.462	0.315 0.028 * 0.009 *
Gender	Female	Male	−0.360	0.698	0.006 *

*: Correlation is significant at the 0.05 level.

The results indicate that the accident propensity of bus drivers is inversely proportional to their age and driving age and proportionate to their number of violations and alarms. Bus drivers aged 20–25 years (OR = 1) were the most accident prone, while those aged 26–35 years (OR = 0.616), 36–45 years (OR = 0.503), 46–55 years (OR = 0.419), and 56–60 years (OR = 0.236) had a 38.4%, 49.7%, 58.1%, and 76.4% lower risk of accidents than bus drivers aged 20–25 years, respectively. In terms of driving age, bus drivers with less than two years (OR = 1) and 2–5 years (OR = 0.967) had significantly higher accident probability, while those with 6–9 years (OR = 0.796), 10–14 years (OR = 0.759), 15–20 years (OR = 0.670), and 20 years or more (OR = 0.589) had a 20.4%, 24.1%, 33.0%, and 41.1% reduction in accident risk, respectively, compared to drivers with less than two years.

Bus drivers who had a violation were more likely to have an accident compared with those who had one violation (OR = 1.635), being 1.635 times more likely than those who had never had a violation (OR = 1); those who had multiple violations (OR = 2.272) were 2.272 times more likely to have an accident compared with those who had never had a violation. A larger number of alarms was linked to a greater probability of bus drivers being involved in an accident. Bus drivers with 16–30 alarms (OR = 1.243) had a 24.3% higher risk of an accident than bus drivers who had never had an alarm (OR = 1), while bus drivers with more than 30 alarms (OR = 1.462) had a 46.2% higher risk of an accident than bus drivers who had never had an alarm. Regarding gender, men bus drivers (OR = 1) were 30.2% more likely to be involved in an accident than female drivers (OR = 0.698).

In this study, the Hosmer–Lemeshow tests are used to test the goodness of fit of the model. The *p*-value of the model was 0.464, greater than 0.05 at a 95% confidence level, indicating that the model had a good goodness of fit. As shown in Figure 3, the sensitivity and specificity of the model were further evaluated by constructing a ROC curve to calculate the area under the curve (AUC). The AUC ranges from 0 to 1, with values closer to 1 indicating better model fit. The AUC value of this model is 0.842, which is regarded as satisfactory, and the ROC curve is above the diagonal, indicating that the model has a reasonable discriminating ability.
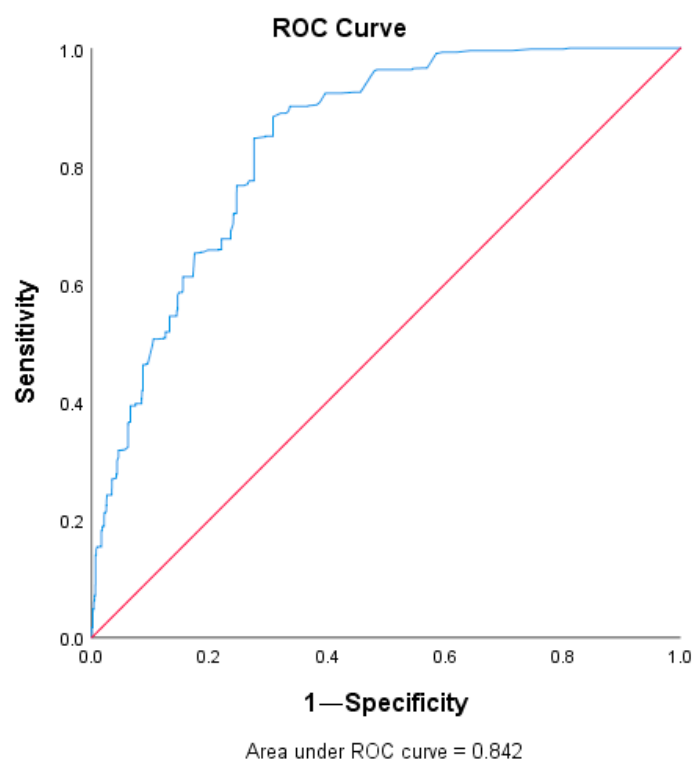
Figure 3. ROC curve.

5. Discussion

This study examined the importance of poor driving records and demographic factors in influencing bus drivers' degree of accident propensity and the relationship between the values taken for each factor and accident propensity. In order of importance, violations, age, driving age, and alarms were the main characteristics that significantly influenced the accident propensity of bus drivers. Other factors, such as nationality and political background, had no discernible influence on the likelihood of accidents. Four kinds of bus drivers are more likely to be involved in collisions: those with more violations, a younger age, a shorter driving age, and more alarms.

The C5.0 Decision Tree Model captured the critical factors that affect whether a bus driver has a high or low propensity for accidents. The results indicate that the bus driver's number of violations is the most important factor influencing accident propensity; those bus drivers who have committed safety violations are likelier to have accidents. The violations covered by this study range from infractions of traffic laws, such as speeding, illegal lane changes, driving with the door open, disobeying traffic signals, talking on the phone while driving, eating, smoking, and using cell phones, to bus service safety violations as determined by the bus company. The more safety violations a driver commits, the less safety conscientious or ignorant of traffic laws they are. As bus drivers, they spend the majority of their time on the road operating buses, which is an active component of road traffic. Bus drivers will be far more likely to be in accidents due to their poor safety awareness and ignorance of traffic laws and regulations, which their lengthy commutes and high mileage will further exacerbate. This study concluded that bus drivers who have committed safety violations are more likely to have accidents. The accident probability of bus drivers who have committed one violation and multiple violations is 1.635 times and 2.272 times higher, respectively, than bus drivers who have never committed violations, consistent with previous research. P. Xie proposed that the more regular violations bus drivers engage in, the more likely they are to make driving errors that result in accidents [32]. L. Mallia et al. evaluated the gathered bus driver driving scale and discovered that self-reported violations predicted the number of accidents among bus drivers [33]; the greater

the number of these violations, the greater the number of accidents. Violations break the normative expectations of all drivers while driving, making them prone to driving errors and exposing other vehicles to serious traffic risks by creating unexpected traffic situations. Managers of public transportation organizations should be on the lookout for drivers who have repeatedly made traffic violations and implement training management strategies to lower their accident propensity.

Generally, it is believed that age is a significant factor in determining a driver's accident propensity, and younger drivers tend to be more accident-prone [9,34]. According to the conclusions of this study, the accident propensity of bus drivers decreases with age from 20 to 60 years old, and bus drivers aged 20 to 25 have the highest accident propensity; their accident probability is 76.4% higher than that of bus drivers aged 56 to 60 years old, and 38.4%, 49.7%, and 58.1% higher than that of bus drivers aged 26 to 35, 36 to 45, and 46 to 55 years old, respectively. Young bus drivers frequently lack sufficient driving experience and skillful driving skills; the longer bus routes and more complex traffic environments also made young bus drivers more prone to accidents than older drivers. In addition, young people tend to have a more confident driving attitude and more aggressive driving behavior, making them more prone to driving negligence or road rage behaviors, which is more likely to lead to accidents [35,36]. Similarly, other researchers have discovered that younger drivers are more active and hence more distracted than middle-aged and older drivers, resulting in accidents with insufficient time to conduct driving maneuvers. According to S. Kaplan et al., the chance of distraction while driving for drivers aged 18 to 24 years is 1.63 times that of drivers aged 40 to 59 years [2], and the probability of distraction for drivers aged 25 to 39 years is similarly 17% greater than that of drivers aged 40 to 59 years, which is consistent with A. Sheykhfard's conclusion that young drivers are prone to distraction [37]. The likelihood of distraction may be raised by underestimating the amount to which distraction influences the occurrence of driving errors. It is proposed in this study that, in regard to bus drivers younger than 60 years old, the older they are, the less likely they are to be involved in accidents. Some scholars argue the opposite, that as drivers become older than 50 years old, their physical fitness gradually deteriorates, and their ability to react and perceive the traffic environment gradually decreases, leading to an increase in the probability of accidents [10,38]. In the case of bus drivers, however, bus firms periodically evaluate the health and fitness of their affiliated drivers; if a driver's physical condition is insufficient for driving a bus, they will not be allowed to continue working as a bus driver. Furthermore, after using driving simulators to collect driving data from older drivers, M. P. Isabelle et al. discovered that [39], while older drivers may not perform as well as younger drivers, they will actively adopt compensatory strategies to reduce their risk of traffic accidents by paying more attention while driving, ensuring that bus drivers' physical abilities, though declining with age, can still safely and adequately complete the drive.

According to existing studies, driving age is an important factor influencing drivers' risk of accidents. C. M. Tseng's study discovered that rookie drivers with fewer than three years of driving age had the highest likelihood of being at fault, 12.4% [40]. The conclusions in this study are consistent with the findings of most scholars that bus drivers with shorter driving ages are more likely to be involved in accidents. Accidents are 41.1% and 37.8% more likely for bus drivers with fewer than two years of driving age and 2 to 5 years of driving age, respectively, than for bus drivers with more than 20 years of driving age. The likelihood of accidents for bus drivers with 6 to 10 and 10 to 15 years of driving age is 17.0% and 8.1% higher, respectively, than the likelihood of accidents for bus drivers with more than 20 years of driving age: 17.0% and 8.1%. Drivers with less driving age are inexperienced enough to react inappropriately and quickly in the face of various unforeseen driving scenarios, resulting in accidents [41]. In addition, drivers with a lower driving age are less adept at vehicle operation, and their psychophysiological condition swings more when driving. Comparatively, drivers with more years behind the wheel tend to be calmer and have a more steady driving condition. Similar to age, driving age

is a significant factor in distraction. Drivers with a low driving age are less able to resist interference and attention distribution than drivers with a higher driving age, and drivers with less driving experience are frequently more prone to fatigue due to fluctuations in their psychophysiological state and the occurrence of distracting behavior.

There are two contrasting perspectives regarding whether gender influences the incidence of driver accidents. According to some scholars, men drivers are less likely to be involved in accidents and are less accident-prone than female drivers [42]. Male drivers are less likely than female drivers to have traffic offenses for the same amount of driving, according to the Internationality Road Traffic and Accident Database (IRTED). Other scholars have suggested that male drivers are more likely to be involved in accidents. They indicated that male drivers are more likely to commit traffic violations and are more likely to incur speeding and safety distance violations on the road and that a higher number of traffic violations means a higher probability of traffic accidents [43–45]. By studying traffic accident data in China's Guangdong Province, G. N. Zhang discovered that male drivers are among the high-risk group for causing fatigue-related accidents [46]. This study concludes that although gender has a lower impact on the propensity of bus drivers to be involved in accidents than factors such as driving age, age, violations, and alarms, it still had an effect. Male bus drivers are 30.2% more likely to be involved in an accident than female bus drivers. Compared to women's softer emotions and cautious driving styles, men are more daring and more likely to undergo mood swings while driving, leading to road rage and other aberrant states that can lead to traffic accidents [23]. In addition to emotional driving, male drivers are at a higher risk of exhaustion and distracted driving [47], which are common causes of traffic accidents. To prevent accidents, the bus firm should accurately monitor drivers and issue timely warning tips to drivers or suspend their driving duties when emotional, weary, or distracted driving conditions manifest.

Alarms have a comparable impact on bus drivers' propensity to have accidents as violations do. The alarms in this study's data include abnormal driver states such as fatigued driving, distracted driving, emotional driving, and so on. The alarms are triggered when the bus driver's abnormal state performance reaches a certain threshold, reflecting the degree of stability of the driver's physiological and psychological state during the daily driving process. The more alarms received, the more times the bus driver is in an aberrant state while driving, and the more unstable the driving becomes. According to this study, bus drivers who were notified by alarms showed a gradually growing propensity for accidents as the number of alarms increased. Accidents were 24.3% and 46.2% more likely for bus drivers with 15–30 and more than 30 alarms, respectively, than bus drivers who had never been alarmed. Bus drivers' accident propensity will be significantly reduced, and accidents will be avoided if they are strictly managed and the corresponding traffic regulations are popularized. Their dangerous behaviors and driving status should be closely examined to reduce their safety violations and alarms. The findings of A. T. Kashani et al. support our conclusion that a one-unit increase in the monitoring score for risky driver behavior is associated with a 39% decrease in the likelihood of an accident [48].

6. Conclusions and Recommendations

In this study, a C5.0 decision tree model and a binary logistic model were used to analyze the data of bad driving records and demographic factors of Chinese bus company drivers to determine the relative importance of each factor in influencing accident propensity and the relationship between the value of each factor and accident propensity. According to the study results, the number of received violations is the most influential factor in bus drivers' accident propensity. The more violations they had, the more likely they were to have traffic accidents. The following factors in order of importance were age, driving age, and number of alarms, respectively; gender had an effect on accident propensity but was not significant. Violations and alarms are positively related to bus driver accident propensity, and drivers with more violations and alarms are more likely to be involved in accidents; age and driving age are inversely associated with bus driver

accident propensity, and younger drivers with a shorter driving age are more likely to be involved in accidents; men have a higher accident risk probability than women.

On the finding that poor driving records such as violations and alarms increase the accident propensity of the bus driver, we recommend that public transportation companies reduce the occurrence of bus driver violations and alarms by promoting awareness of traffic laws and regulations and changing people's attitudes toward traffic safety, in order to lower each driver's propensity for accidents.

On the other hand, bus companies can also screen their drivers appropriately, hiring fewer or no bus drivers with a high propensity for accidents, such as young people under the age of 25, inexperienced drivers with less than two years of driving experience, and drivers with a high number of violations in their driving history. Second, public transportation companies can implement a corresponding penalty system to lessen the appearance of the bus driver's poor behavior and condition. For example, if the driver has a poor driving record, they may be subject to education and fines; if the bus driver commits repeated violations and alarms, depending on the situation, he may be suspended from driving altogether. Additionally, it is crucial for bus operators to regularly examine bus drivers medically and psychologically in order to prevent physical and mental disorders that could impair their ability to drive and result in traffic accidents [1].

This study also has limitations. Only eight variables are used to analyze the factors influencing bus drivers' accident propensity, and the study is insufficient. Previous research has demonstrated that the marital status [17] and income [33] of bus drivers also influence the likelihood of accidents. In the subsequent study, a more comprehensive analysis of the accident propensity of bus drivers will be conducted using a dataset with more variables and specific information.

Author Contributions: Conceptualization, L.Z. and X.H.; methodology, X.H.; software, X.H.; validation, L.Z., X.H., Z.X. and T.D.; resources, L.Z.; data curation, X.H.; writing—original draft preparation, X.H.; writing—review and editing, X.H.; visualization, X.H. and Y.L.; supervision, L.Z.; project administration, L.Z. and T.D.; funding acquisition, L.Z. and X.H. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chimba, D.; Sando, T.; Kwigizile, V. Effect of bus size and operation to crash occurrences. *Accid. Anal. Prev.* **2010**, *42*, 2063–2067. [CrossRef] [PubMed]
2. Kaplan, S.; Prato, C.G. Risk factors associated with bus accident severity in the United States: A generalized ordered logit model. *J. Saf. Res.* **2012**, *43*, 171–180. [CrossRef] [PubMed]
3. Barua, U.; Tay, R. Severity of urban transit bus crashes in Bangladesh. *J. Adv. Transp.* **2010**, *44*, 34–41. [CrossRef]
4. Traffic Administration Bureau of the Ministry of Public Security of the People's Republic of China. In *Annual Report on Road Traffic Accidents 2020[R]*; Research Institute of Traffic Management, Ministry of Public Security: Beijing, China, 2021.
5. Strathman, J.G.; Wachana, P.; Callas, S. Analysis of bus collision and non-collision incidents using transit ITS and other archived operations data. *J. Saf. Res.* **2010**, *41*, 137–144. [CrossRef] [PubMed]
6. Das, S.; Sun, X.; Wang, F.; Leboeuf, C. Estimating likelihood of future crashes for crash-prone drivers. *J. Traffic Transp. Eng. (Engl. Ed.)* **2015**, *2*, 145–157. [CrossRef]
7. Khanehshenas, F.; Mazloumi, A.; Jalaldehi, P.A.; Kaveh, M. Drivers' subjective perceptions of the contextual factors influencing fatigue: A qualitative study of suburban bus drivers in Iran. *Work* **2022**, *72*, 1481–1491. [CrossRef]
8. Dorn, L.; Af Wåhlberg, A. Work-related road safety: An analysis based on U.K. bus driver performance. *Risk Anal. Int. J.* **2008**, *28*, 25–35. [CrossRef]
9. Goh, K.; Currie, G.; Sarvi, M.; Logan, D. Factors affecting the probability of bus drivers being at-fault in bus-involved accidents. *Accid. Anal. Prev.* **2014**, *66*, 20–26. [CrossRef] [PubMed]

10. Huting, J.; Reid, J.; Nwoke, U.; Bacarella, E.; Ky, K.E. Identifying Factors That Increase Bus Accident Risk by Using Random Forests and Trip-Level Data. *Transp. Res. Rec. J. Transp. Res. Board* **2016**, *2539*, 149–158. [CrossRef]
11. Li, D.-H.; Liu, Q.; Yuan, W.; Liu, H.-X. Relationship between fatigue driving and traffic accident. *J. Traffic Transp. Eng.* **2010**, *10*, 104–109.
12. Jovanis, P.P.; Schofer, J.L.; Prevedouros, P.; Tsunokawa, K. *Analysis of Bus Transit Accidents: Empirical, Methodological and Policy Issues*; Transportation Research Board: Springfield, IL, USA, 1991.
13. Glendon, A.I.; McNally, B.; Jarvis, A.; Chalmers, S.L.; Salisbury, R.L. Evaluating a novice driver and pre-driver road safety intervention. *Accid. Anal. Prev.* **2014**, *64*, 100–110. [CrossRef] [PubMed]
14. Chen, H.-Y.W.; Donmez, B.; Liberty Hoekstra-Atwood, L.; Marulanda, S. Self-reported engagement in driver distraction: An application of the Theory of Planned Behaviour. *Transp. Res. Part F Traffic Psychol. Behav.* **2016**, *38*, 151–163. [CrossRef]
15. Tao, D.; Zhang, R.; Qu, X. The effects of gender, age and personality traits on risky driving behaviors. *J. Shenzhen Univ. Sci. Eng.* **2016**, *33*, 551–660. [CrossRef]
16. Zhang, J.; Wang, L.; Yuan, H.; Duan, M.; Xie, L. Relationship between Driving Skills and Driving Behaviors Based on a Structural Equation Model. In Proceedings of the 21st COTA International Conference of Transportation, Xi'an, China, 17–20 December 2021; pp. 1704–1713.
17. Hamed, M.; Jaradat, A.; Easa, S.M. Analysis of commercial mini-bus accidents. *Accid. Anal. Prev.* **1998**, *30*, 555–567. [CrossRef]
18. Azhar, A.; Ariff, N.M.; Abu Bakar, M.A.; Roslan, A. Classification of driver injury severity for accidents involving heavy vehicles with decision tree and random forest. *Sustainability* **2022**, *14*, 4101. [CrossRef]
19. Taamneh, M. Investigating the role of socio-economic factors in comprehension of traffic signs using decision tree algorithm. *J. Saf. Res.* **2018**, *66*, 121–129. [CrossRef] [PubMed]
20. Rejali, S.; Aghabayk, K.; Shiwakoti, N. A clustering approach to identify high-risk taxi drivers based on self-reported driving behavior. *J. Adv. Transp.* **2022**, *2022*, 6511225. [CrossRef]
21. Lukongo, O.E.B. Examining prominent causes of traffic injury severity in Louisiana with multinomial logistic models. *Transp. Res. Rec. J. Transp. Res. Board* **2020**, *2675*, 245–257. [CrossRef]
22. Xiao, Y.; Liu, H.; Liang, Z. An analysis of the influential factors of violations in urban-rural passenger transport drivers. *J. Adv. Transp.* **2022**, *2022*, 1652923. [CrossRef]
23. Smart, R.G.; Stoduto, G.; Mann, R.E.; Adlaf, E. Road Rage Experience and Behavior: Vehicle, Exposure, and Driver Factors. *Traffic Inj. Prev.* **2004**, *5*, 343–348. [CrossRef]
24. Fang, A.; Qiu, C.; Zhao, L.; Jin, Y. Driver Risk Assessment Using Traffic Violation and Accident Data by Machine Learning Approaches. In Proceedings of the 2018 3rd IEEE International Conference on Intelligent Transportation Engineering (Icite), Singapore, 3–5 September 2018; pp. 291–295.
25. Chu, W.; Wu, C.; Atombo, C.; Zhang, H.; Özkan, T. Traffic climate, driver behaviour, and accidents involvement in China. *Accid. Anal. Prev.* **2018**, *122*, 119–126. [CrossRef] [PubMed]
26. Segura, M.; Mello, J.; Hernández, A. Machine Learning Prediction of University Student Dropout: Does Preference Play a Key Role? *Mathematics* **2022**, *10*, 3359. [CrossRef]
27. Chang, W.-C.; Lan, T.-H.; Ho, W.-C.; Lan, T.-Y. Factors affecting the use of health examinations by the elderly in Taiwan. *Arch. Gerontol. Geriatr.* **2010**, *50*, S11–S16. [CrossRef]
28. Fadlalla, A. An experimental investigation of the impact of aggregation on the performance of data mining with logistic regression. *Inf. Manag.* **2005**, *42*, 695–707. [CrossRef]
29. Cobas, J.A. Regression models for categorical and limited dependent variables. *Sociol. Inq.* **2000**, *70*, 109.
30. Gómez-Fernández, N.; Mediavilla, M. Factors Influencing Teachers' Use of ICT in Class: Evidence from a Multilevel Logistic Model. *Mathematics* **2022**, *10*, 799. [CrossRef]
31. Yan, X.; He, J.; Zhang, C.; Liu, Z.; Qiao, B.; Zhang, H. Single-vehicle crash severity outcome prediction and determinant extraction using tree-based and other non-parametric models. *Accid. Anal. Prev.* **2021**, *153*, 106034. [CrossRef]
32. Xie, P.; Qin, D.; Zhu, T. Impact of rule-violating behaviors on the risk of bus drivers being at-fault in crashes. *Traffic Inj. Prev.* **2022**, *23*, 364–368. [CrossRef]
33. Mallia, L.; Lazuras, L.; Violani, C.; Lucidi, F. Crash risk and aberrant driving behaviors among bus drivers: The role of personality and attitudes towards traffic safety. *Accid. Anal. Prev.* **2015**, *79*, 145–151. [CrossRef] [PubMed]
34. Factor, R. An empirical analysis of the characteristics of drivers who are ticketed for traffic offences. *Transp. Res. Part F Traffic Psychol. Behav.* **2018**, *53*, 1–13. [CrossRef]
35. Taubman-Ben-Ari, O.; Mikulincer, M.; Gillath, O. The multidimensional driving style inventory-scale construct and validation. *Accid. Anal. Prev.* **2004**, *36*, 323–332. [CrossRef]
36. Maxwell, J.; Grant, S.; Lipkin, S. Further validation of the propensity for angry driving scale in British drivers. *Pers. Individ. Differ.* **2005**, *38*, 213–224. [CrossRef]
37. Sheykhfard, A.; Haghighi, F. Driver distraction by digital billboards? Structural equation modeling based on naturalistic driving study data: A case study of Iran. *J. Saf. Res.* **2019**, *72*, 1–8. [CrossRef]
38. Zwerling, C.; Peek-Asa, C.; Whitten, P.S.; Choi, S.-W.; Sprince, N.L.; Jones, M.P. Fatal motor vehicle crashes in rural and urban areas: Decomposing rates into contributing factors. *Inj. Prev.* **2005**, *11*, 24–28. [CrossRef]

39. Isabelle, M.P.; Simon, M. Comparison between elderly and young drivers' performances on a driving simulator and self-assessment of their driving attitudes and mastery. *Accid. Anal. Prev.* **2020**, *135*, 105317.

40. Tseng, C.M. Social-demographics, driving experience and yearly driving distance in relation to a tour bus driver's at-fault accident risk. *Tour. Manag.* **2012**, *33*, 910–915. [CrossRef]

41. Hutchens, L.; Senserrick, T.M.; Jamieson, P.E.; Romer, D.; Winston, F.K. Teen driver crash risk and associations with smoking and drowsy driving. *Accid. Anal. Prev.* **2008**, *40*, 869–876. [CrossRef] [PubMed]

42. Zhang, F.; Mehrotra, S.; Roberts, S.C. Driving distracted with friends: Effect of passengers and driver distraction on young drivers' behavior. *Accid. Anal. Prev.* **2019**, *132*, 105246. [CrossRef]

43. Shinar, D.; Compton, R. Aggressive driving: An observational study of driver, vehicle, and situational variables. *Accid. Anal. Prev.* **2004**, *36*, 429–437. [CrossRef]

44. Masuri, M.G.; Mustaffa, D.N.A.; Dahlan, A.; Isa, K.A. The Intention in Speeding Behavior between Low and High Intended Young Driver in Urban University. *Environ. Proc. J.* **2016**, *1*, 330–335. [CrossRef]

45. Claret, P.L.; del Castillo, J.D.D.L.; Moleón, J.J.J.; Cavanillas, A.B.; Martín, M.G.; Vargas, R.G. Age and sex differences in the risk of causing vehicle collisions in Spain, 1990 to 1999. *Accid. Anal. Prev.* **2003**, *35*, 261–272. [CrossRef]

46. Zhang, G.; Yau, K.K.; Zhang, X.; Li, Y. Traffic accidents involving fatigue driving and their extent of casualties. *Accid. Anal. Prev.* **2016**, *87*, 34–42. [CrossRef] [PubMed]

47. D'Souza, K.A.; Maheshwari, S.K. Multivariate Statistical Analysis of Public Transit Bus Driver Distraction. *J. Public Transp.* **2012**, *15*, 1–23. [CrossRef]

48. Kashani, A.T.; Besharati, M.M. An investigation of the relationship between demographic variables, driving behaviour and crash involvement risk of bus drivers: A case study from Iran. *Int. J. Occup. Saf. Ergon.* **2021**, *27*, 535–543. [CrossRef] [PubMed]