


Article

Bus Fleet Accident Prediction Based on Violation Data: Considering the Binding Nature of Safety Violations and Service Violations

Tongqiang Ding¹, Lianxin Zhang¹, Jianfeng Xi^{1,*} , Yingjuan Li², Lili Zheng¹ and Kexin Zhang¹¹ Transportation College, Jilin University, Changchun 130022, China² Department of Traffic Management, Jilin Police College, Changchun 130022, China

* Correspondence: xijf@jlu.edu.cn

Abstract: The number and severity of bus traffic accidents are increasing annually. Therefore, this paper uses the historical data of Chongqing Liangjiang Public Transportation Co., Ltd. bus driver safety violations, service violations, and road traffic accidents from January to June 2022 and constructs road traffic accident prediction models using Extra Trees, BP Neural Network, Support Vector Machine, Gradient Boosting Tree, and XGBoost. The effects of safety and service violations on vehicular accidents are investigated. The quality of the prediction models is measured by five indicators: goodness of fit, mean square error, root mean square error, mean absolute error, and mean absolute percentage error. The results indicate that the XGBoost model provides the most accurate predictions. Additionally, simultaneously considering safety and service violations can improve the accuracy of the model's predictions compared to a model that only considers safety violations. Bus safety violations, bus service violations, and bus safety operation violations significantly influence traffic accidents, which account for 27.9%, 20%, and 16.5%, respectively. In addition to safety violations, the service violation systems established by bus companies, such as bus service codes, can be an effective method of regulating the behavior of bus drivers and reducing accidents. They are improving both the safety and quality of public transportation.

Keywords: public transportation; safety violations; service violations; accident number prediction; safety operation management



Citation: Ding, T.; Zhang, L.; Xi, J.; Li, Y.; Zheng, L.; Zhang, K. Bus Fleet Accident Prediction Based on Violation Data: Considering the Binding Nature of Safety Violations and Service Violations. *Sustainability* **2023**, *15*, 3520. <https://doi.org/10.3390/su15043520>

Academic Editor: Biyu Chen

Received: 9 December 2022

Revised: 9 February 2023

Accepted: 11 February 2023

Published: 14 February 2023



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

1. Introduction

Due to frequent traffic congestion and the promotion of green travel, public transportation is becoming the preferred mode of transportation for some individuals [1]. According to the Central City Passenger Volume Report published in January 2022 by China's Ministry of National Transportation, the cumulative passenger volume of public trams in China's central cities reached 166,654,000 in January. By the beginning of 2021, there were 704,400 public buses and trams in China, an increase of 1.6% over 2020.

The rapid increase in bus passenger volume and ownership has been accompanied by an increase in bus traffic accidents, with approximately 14,000 bus-road traffic accidents resulting in 3500 deaths and 16,000 injuries in China over the past five years. As a result, many scholars have studied models to investigate the factors influencing bus traffic accidents and to forecast the number of accidents [2–5].

Driver violations are frequently considered in the prediction of traffic accidents [6–9]. In the existing literature, violations are primarily associated with unsafe driving behaviors such as speeding, red light running, and DUI (driving under the influence) [10–12]. However, because public transportation is a service, irregular service can also result in violations. Sufficient research has not been conducted to determine whether service violations influence the occurrence of traffic accidents and, consequently, traffic accident prediction models.

In this paper, the violations are categorized as safety violations and service violations, and the data of safety violations, service violations, and accident statistics for a bus company in Chongqing, China, from January to June 2022 are analyzed. Extra Trees (ETs), BP Neural Network, Support Vector Machine, Gradient Boosting Tree, and XGBoost are used to construct five different models for predicting the number of traffic accidents and the degree of influence of various violation types on the occurrence of bus traffic accidents, respectively.

The rest of the paper is organized as follows. In Section 2, a summary and literature review are presented. Section 3 describes data processing. Five prediction models are developed in Section 4. In Section 5, the modeling results are discussed. Finally, the article is summarized in Section 6. Figure 1 shows the conceptual model depicting the process of study.

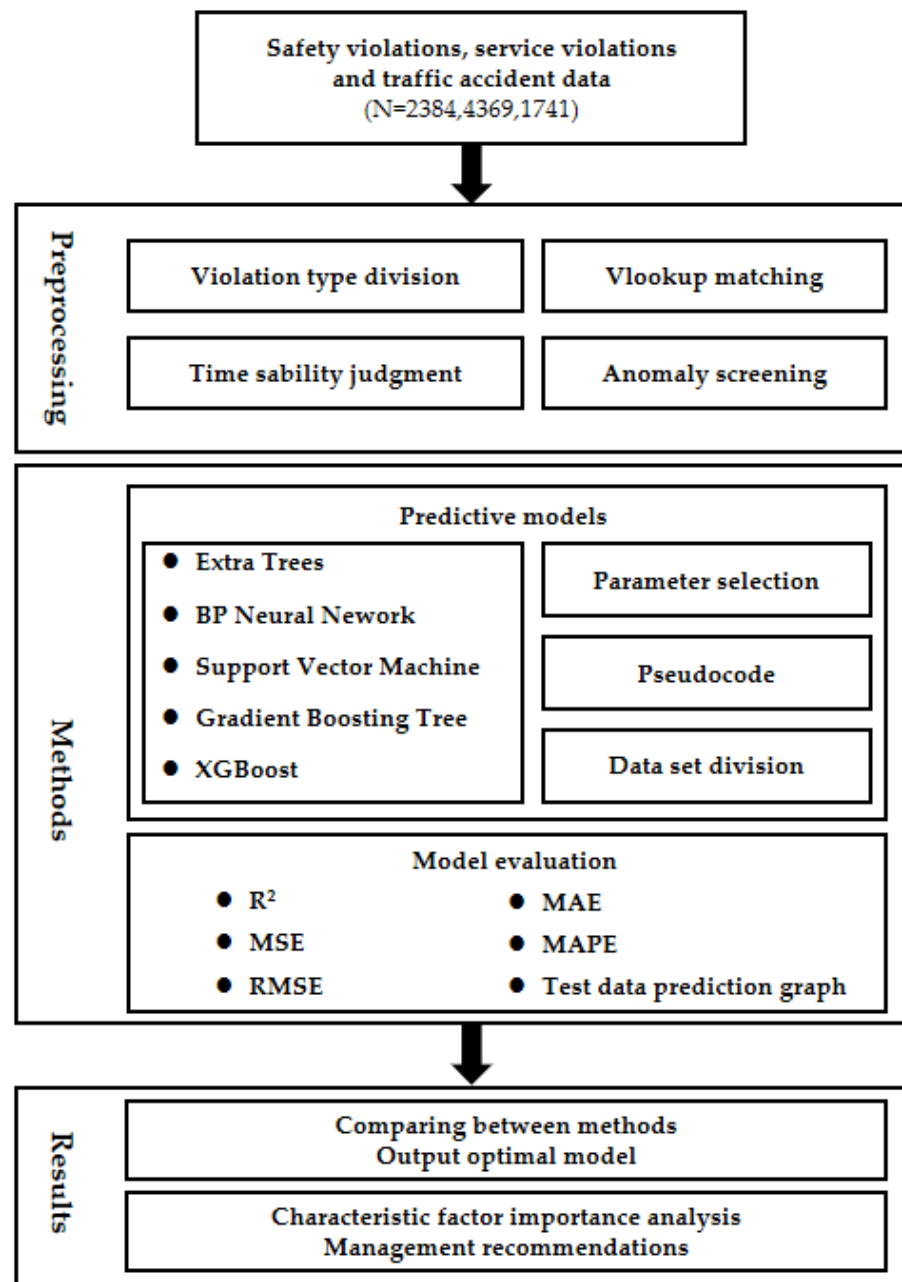


Figure 1. The conceptual model depicting the process of study.

2. Literature Review

2.1. Traffic Violations and Traffic Accidents

Numerous studies have revealed a significant correlation between traffic violations and accidents. In most cases, accidents are typically caused by violations of one or more traffic laws [13]. The findings of Guangnan Zhang identified traffic violations as one of the major threats to road safety [14]. According to the research conducted by Ayuso and Ebrahemzadih in Spain and Iran, drivers who violate traffic laws increase the accident rate [15,16]. Alver modeled the scenario data using an ordered probit model and discovered that drivers with at least one ticket had a 50.4% accident rate in the past three years [17].

On this basis, scholars have investigated the connection between particular types of violations and traffic accidents. Mao used logistic regression to discover that fatigued driving was strongly associated with traffic accidents [18]. Mansour Hadji Hosseinlou utilized a zero-truncated Poisson model to confirm that speeding violations and collisions were positively correlated [19]. Sigal Kaplan utilized an ordered generalized logit model [20]. It was discovered that the accident rate was raised when a vehicle exceeded or fell below the minimum speed requirement for road travel. Anebonam discovered through data analysis that the primary human factors in traffic accidents were speeding, loss of vehicle control, and dangerous driving [21]. Terje Assum discovered that when drunk driving laws became stricter, both the number of DUIs and the number of traffic accidents decreased [22]. David Shinar discovered that illegal overtaking, lane changes, and traffic sign violations could result in traffic collisions on urban roads [23]. Iversen's study revealed a significant correlation between driving offenses like DUI and seatbelt violations and traffic accidents [24]. The G. Maycock survey revealed that people with a DUI were more likely to be involved in a more severe DUI accident in the future [25]. Feraud noted that violations of the road safety code were one of the major causes of traffic accidents [26]. Maowei Chen examined truck traffic safety and concluded that "lane violations" and "signal violations" had a significant impact on the severity of truck accidents [27]. Table 1 provides a summary of the specific violations that were investigated.

Table 1. Summarizes the literature related to the specific traffic violations.

Violations	References
Fatigue driving	Mao et al. [18]
Speeding	Hadji et al. [19]; Kaplan et al. [20] and Anebonam et al. [21]
Dangerous driving	Anebonam et al. [21]
DUI	Assum [22]; Iversen [24] and Maycock [25]
Not wearing seat belt	Maycock [25]
Illegal lane change	Shinar et al. [23] and Feraud et al. [26]
Signal violation	Feraud et al. [26]
Violation of traffic signs and markings	Shinar et al. [23]

2.2. Traffic Accident Prediction Model Based on Machine Learning

Traditional statistical methods require assumptions about the data distribution when modeling and often require a linear functional form between the dependent and explanatory variables. However, when assumptions are violated, incorrect estimates and incorrect inferences may be generated [28]. Machine learning-based methods can avoid this limitation and more accurately predict traffic accidents. Thus, they have been widely used in traffic prediction problems in recent years [29–31].

Farhangi used bagged decision trees, ETs, and random forest (RF) algorithms for accident risk prediction on a geographic information system (GIS) platform [32]. Wang used violation and accident records and performed predictions in a connected vehicle environment through LSTM-RNN [33]. Mohammad predicted traffic accidents through histogram-based gradient augmentation (HistGBDT) [34]. He used the prediction results as a reference to decide whether an ambulance should be dispatched to the accident scene. Ju Yang found that the BP neural network prediction model has the best generalization when

the activation functions are ReLU and sigmoid [35]. B. Yu found that both artificial neural network models (ANN) and SVM models can predict traffic accidents within an acceptable range [36]. ANN is better for long-time events, and SVM has better overall performance than ANN in prediction. Li and Chen compared SVM with ordered probability (OP) and Gaussian radial basis function (RBF) and demonstrated that the SVM has a more reasonable prediction performance [37,38]. Katha Mehta compared RF, SVM, Stochastic Gradient Descent, ANN, and XGBoost model performance and found that XGBoost provides the best results [39]. Moreover, Yookyung Boo compared the prediction models of RF, ETs, and XGBoost and demonstrated that the XGBoost model combined with SMOTE samples has the best prediction performance [40].

The above-mentioned models were confirmed for their reliability in prediction. In this paper, five of them are selected for accident prediction modeling: Extra Trees, BP Neural Networks, Support Vector Machines, Gradient Boosting Trees (GBDT), and XGBoost.

2.3. Summary

Section 2.1 demonstrates that there is an undeniable correlation between the occurrence of safety violations such as speeding, drunk driving, and illegal lane changes and the number of traffic accidents. Thus, safety violations can be used to predict the number of traffic accidents. However, there are some service violations for bus drivers that may be worth observing, such as answering the phone during non-driving time while in the service process, allowing a missing reflector in the bus, and not running in the bus priority lane. Service violations reflect the driver's work ethic. Hussain demonstrated that a poor work ethic may increase the risk of traffic accidents [41]. Thus, a scientifically feasible qualitative or quantitative analysis is worth conducting to explore whether these service violations also affect the occurrence of traffic accidents, similar to safety violations, and whether considering service violations can improve the accuracy of the accident prediction model.

Consequently, this paper proposes a model for predicting the frequency of bus accidents based on safety violations and service violations, and reveals the degree of impact of each type of violation. The model with the best prediction effect is selected by comparing five machine learning prediction models with five indicators: goodness of fit, root mean square error, root means square error, average absolute error, and average absolute percentage error. Then, a prediction model is constructed using safety violation data. The new prediction model is compared to the model constructed with service violation data and safety violation data to determine whether the addition of service violation data significantly improves the prediction accuracy of the model.

3. Data

3.1. Data Processing

The data were obtained from the Chongqing Liangjiang Public Transportation Co., Ltd. data of safety violations, service violations, and traffic accidents from January to June 2022. The data contained 2384 safety violations, 4369 service violations, and 1741 traffic accidents and involved 227 bus fleets. The safety and service violation types were set according to Wang's article and the *Passenger Transport Services Specifications for Urban Bus/Trolleybus* [42,43]. The violation type and traffic accident occurrence were matched and counted for each bus fleet using the Vlookup function for subsequent analysis. Table 2 shows some of the processed data.

Safety violations consist of five categories including bus GPS speeding violations (S_a), dangerous driving behavior of bus (S_b), violation of bus safety operation regulations (S_c), violation of bus safety regulations (S_d), and violation of general bus road traffic regulations (S_e). Furthermore, service violations consist of seven categories including bus signs and markings (S_1), bus service specifications (S_2), bus cleaning and sanitation (S_3), bus facilities and equipment (S_4), bus station transport order (S_5), complaints (S_6), and other violations (S_7). Table 3 shows the specific violations for each violation type.

Table 2. Statistics on the types of violations and the number of accidents in some fleets.

Fleet	Safety Violation Type					Service Violation Type							Number of Accidents
	S _a	S _b	S _c	S _d	S _e	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	
1	1	0	3	1	1	3	9	2	3	3	0	2	7
2	1	4	4	1	2	4	18	13	0	10	1	2	18
3	0	0	2	3	1	0	4	2	3	1	0	0	6
4	2	0	2	0	0	0	2	1	1	0	0	1	5
5	4	3	16	12	12	5	20	1	12	9	1	5	49
6	2	0	1	1	0	0	1	11	0	3	1	4	21
7	2	3	7	19	10	2	19	8	14	5	3	1	46
8	4	2	5	20	8	6	9	11	16	7	0	4	37
9	1	1	5	14	3	2	10	0	1	7	2	2	14
.....
227	6	0	1	0	0	0	3	0	0	0	0	0	0

Table 3. Classification and interpretation of violations.

Type of Violation	Violation Item	Violations
Safety Violations	Bus GPS speeding violations	The speed limit specified by the driving section as the standard, divided into speeding 10 km/h (excluding) below, speeding 10 km/h (including) above—to 20 km/h (excluding) below, speeding 20 km/h (including) above—to 50 km/h (excluding) below and speeding 50 km/h (including) above.
	Dangerous driving behavior of bus	<ul style="list-style-type: none"> • Running red lights or yellow lights. • Not stopping to yield to pedestrians through crosswalks or intersections. • Driving against traffic. • Fatigue driving, driving with illness, or lousy state driving. • Driving without maintaining a safe distance. • Chase driving or speed driving. • Other public transport dangerous driving behaviors.
	Violation of bus safety operation regulations	<ul style="list-style-type: none"> • Not correctly using the seat belt, the lights, the wipers, or the cab safety guard. • Violation of the provisions of the opening and closing of the door and driving while opening the door. • Driving while gossiping, eating, drinking, or engaged in activities unrelated to driving work. • Driving while using cell phones or other electronic devices. • One-handed driving or driving with both hands off the steering wheel. • Other violations of bus safety operating regulations of driving behavior.

Table 3. Cont.

Type of Violation	Violation Item	Violations
	Violation of bus safety regulations	<ul style="list-style-type: none"> • Not waiting for passengers to pull or sit firmly before starting to drive. • Shut down, interfere with or destroy GPS or 4G video monitoring equipment or system without permission. • Leaving the cab without taking the engine key. • Not using the triangle wood after parking. • Driving a sick car. • Late reporting or hiding the accidents. • Not carrying a driver's license or a professional qualification certificate. • Before the start of the "wine state." • Other driving violations of public transportation safety regulations.
	Violation of general bus road traffic regulations	<ul style="list-style-type: none"> • Competition for lane rush, illegal lane change, and illegal U-turns. • Fast speed into the station. • Not slowing down in advance when passing crosswalks, intersections, visual blindness, or narrow sections. • Failure to follow the prescribed lane and ride the dividing lane. • Overtake two- and three-wheeled vehicles in the same lane. • Other violations of the general road traffic regulations of public transport.
	Bus signs and markings	Bus compartment logos, bus body logos, and other bus signs logos are missing, broken, and not standardized.
Service violation	Bus service specifications	<ul style="list-style-type: none"> • Bus in the waiting time without monitoring the ticket. • Not opening the mobile TV or the LED display. • Answering the phone, wearing headphones, reading text messages, and surfing the Internet at non-driving times in the service process. • Artificially terminate the operation, dumping passengers in the middle or transferring passengers without any reason. • Station annunciator failure when not manually reporting the station. • Not using Mandarin service, cross-use Mandarin or Chongqing dialect, and using the impermissible service language. • Not wearing a work number plate or safety cuffs. • Strange hair color and style. • Other violations of bus service specifications.
	Bus cleaning and sanitation	<ul style="list-style-type: none"> • Air-conditioning returns air mesh cover, window sill, seat head cover, bus wall panel, mobile TV, toolbox, and other parts of the bus cleaning are unqualified. • Window glass, mirrors, and other parts of the bus body cleaning are unqualified. • The hood or table pile of debris. • Not dumping the garbage at the starting station. • Other bus cleaning and hygiene violations.

Table 3. Cont.

Type of Violation	Violation Item	Violations
	Bus facilities and equipment	<ul style="list-style-type: none"> • Windows, door strips, various lamps (lampshades), light belts, reflectors, air conditioning vents, and other bus body facilities are missing, broken, or off. • Body appearance paint color is not uniform. • Seats are missing or broken. • Power lines are messy. • Other bus facilities and equipment violations.
	Bus station transportation order	<ul style="list-style-type: none"> • The first bus does not stop at the front of the station to get on and off. • Do not follow the last bus to stop at the front of the station to get on and off. • Outside the bus stop to get on and off. • Not driving bus priority lanes. • Press traffic signs while getting on and off. • Get on and off at crosswalks. • Other bus operation order violations.
	Complaints	No liability passenger complaints, minor liability passenger complaints, and general liability passenger complaints.
	Other violations	Other violations in addition to the above violations of the bus company regulations.

To reduce the influence of exceptional data points on the model, Rstudio software was applied to discriminate high leverage points, outliers, and strong influence points with hat diag H, student residual, and Cook’s distance (Cook’s D) as measures, respectively. Figure 2 shows the discriminant results.

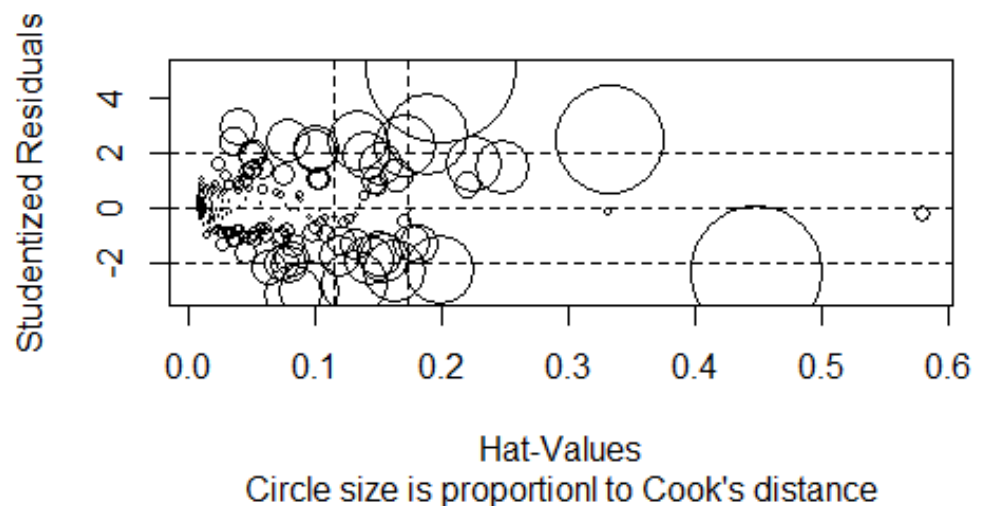


Figure 2. Anomaly screening.

According to the screening results of anomalies, six fleet data of numbers 26, 38, 44, 54, 154, and 161 were deleted. Figures 3 and 4 show the final statistics of the traffic accidents and various violations for each bus fleet, respectively.

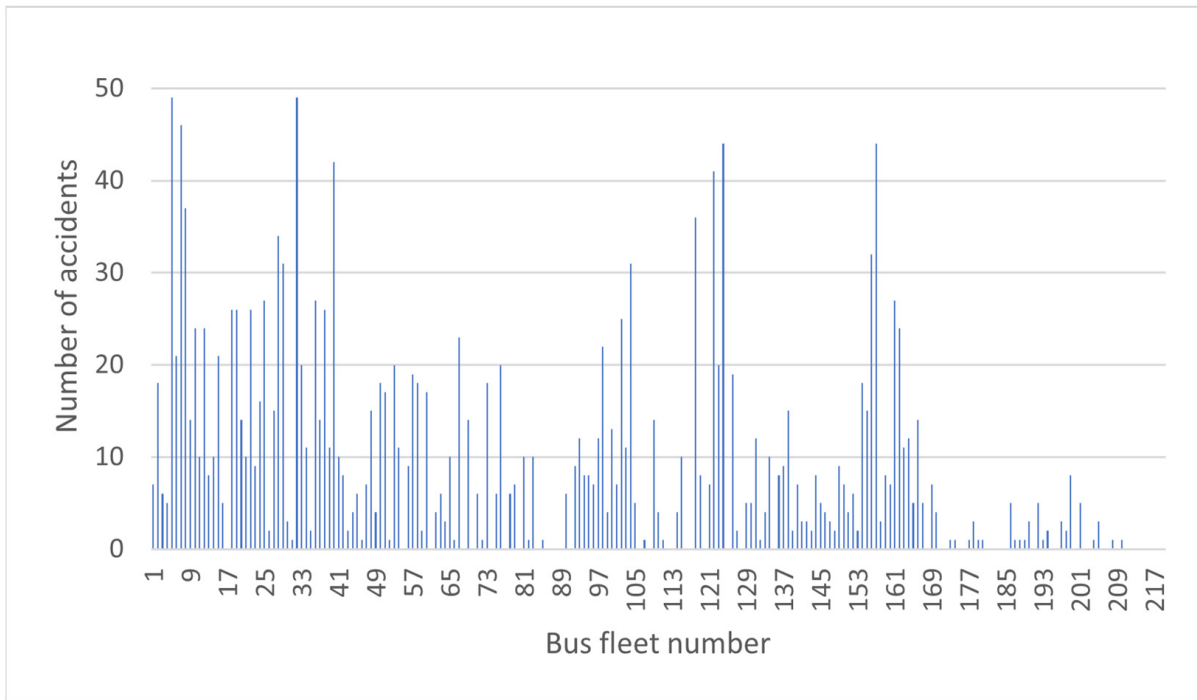


Figure 3. Statistics of the number of traffic accidents in 219 fleets from January to June 2022.

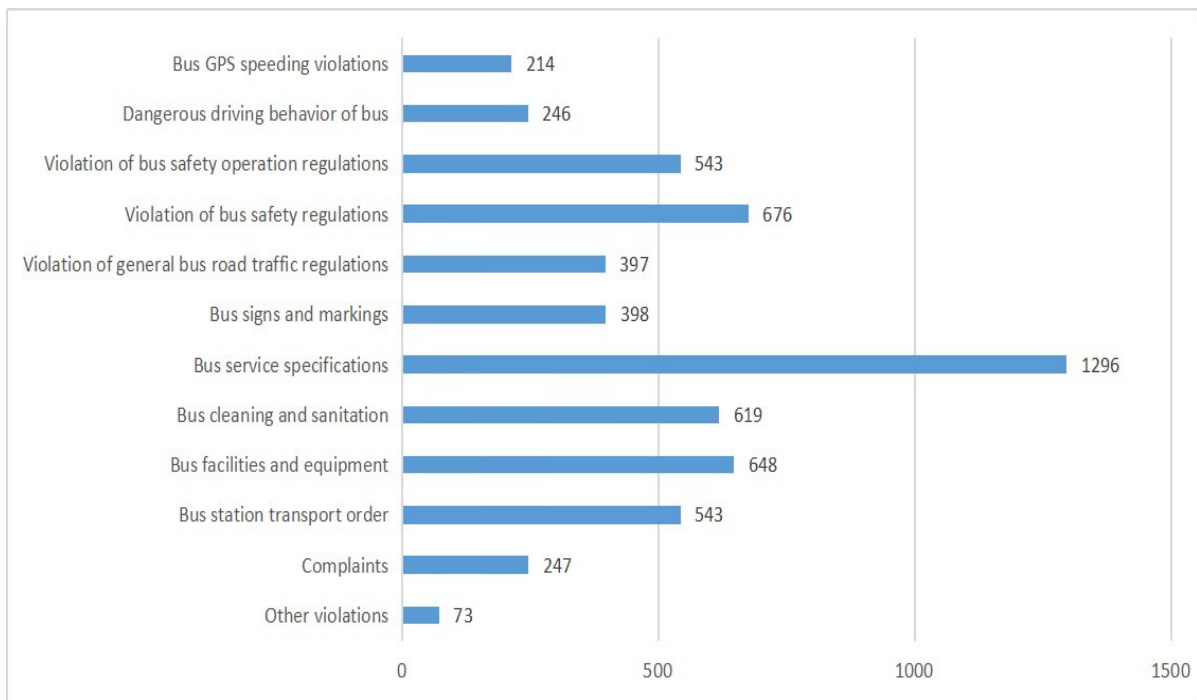


Figure 4. Statistics of the number of violations for 219 fleets from January to June 2022.

It can be seen from Figure 3 that there is a large variability in the number of traffic accidents by bus fleet, with 52 fleets having no accidents from January to June 2022, 64 fleets having up to and including five accidents, 41 fleets having more than five but not more than ten accidents and 34 fleets having more than ten accidents but not more than twenty. There were 28 fleets with more than twenty accidents. The highest number of accidents was 49, which occurred in numbers 6 and 34.

It can be seen from Figure 4 that the occurrence of violations of bus service specifications is significantly higher. Hence, the impact of bus service specifications regarding accidents may be more significant. The impacts of bus station transportation orders, bus facilities and equipment, bus cleaning and sanitation, violations of bus safety operation regulations, and violations of bus safety regulations regarding accidents are slightly less significant. Complaints, dangerous driving behavior, violations related to bus signs and markings, violations of general road traffic regulations, and bus GPS speeding violations are third in terms of the impact on the accidents. Other violations are of the weakest impact. The specific impact needs to be followed by modeling analysis.

3.2. Correlation Analysis

The cardinality test (cross-tabulation analysis) was used to study the different relationships between the types of violations (safety violations and service violations) on the occurrence of accidents [44]. It can be seen from Table 4 that the samples of different violation types showed significance ($p < 0.05$) on the occurrence of accidents, indicating that the samples of different violation types showed differences in the occurrence of accidents.

Table 4. Results of cross-tabulation (chi-square) analysis.

Title	Name	Type of Violation		Total	X^2	p
		Safety Violations	Service Violations			
Accidents happen	Accidents occur	858 (39.20%)	1331(60.80%)	2189	11.828	0.001 **
	Accident-free	1493 (34.85%)	2791(65.15%)	4284		

** $p < 0.01$.

The correlation analysis was used to investigate the correlation between traffic accidents and 12 types of violations. The Spearman correlation coefficient was used to indicate the strength of the correlation. The results are shown in Table 5, all of which showed significance at the 0.01 level [45].

Table 5. Spearman correlation analysis.

Type of Violation	Accidents Occur
Bus GPS speeding violation	0.508 **
Dangerous driving behavior of bus	0.502 **
Violation of bus safety operation regulations	0.696 **
Violation of bus safety regulations	0.692 **
Violation of general bus road traffic regulations	0.593 **
Bus signs and markings	0.662 **
Bus service specification	0.756 **
Bus cleaning and sanitation	0.685 **
Bus facilities and equipment	0.637 **
Bus station transport order	0.695 **
Complaints	0.563 **
Other violations	0.437 **

** $p < 0.01$.

3.3. Time Smoothness Assessment

Due to the global and local temporal dependencies of the traffic data, the temporal stability of accident data needs to be evaluated before formal modeling can be performed [46]. Table 6 shows a sample of the various types of violations of the fleet from January to June. The correlation test between the violation months and the number of accidents for 219 fleets was conducted using Pearson and the results are shown in Table 7. It can be seen from the table that the correlation coefficient value is 0.010, which is close to 0, and the p -value is $0.421 > 0.05$, indicating that there is no direct correlation between accident occurrence and

month of violation. Therefore, the accident prediction model can be constructed without considering the time factor.

Table 6. Number of violations and accidents in a fleet from January to June.

Month	Safety Violation Type					Service Violation Type							Number of Accidents
	S_a	S_b	S_c	S_d	S_e	S_1	S_2	S_3	S_4	S_5	S_6	S_7	
January	0	0	0	1	0	1	2	0	0	1	0	0	2
February	0	0	1	0	0	0	1	1	0	0	0	0	0
March	1	0	0	0	1	0	0	0	0	0	0	1	1
April	0	0	1	0	0	1	3	0	2	1	0	0	2
May	0	0	1	0	0	0	2	1	1	0	0	0	1
June	0	0	0	0	0	1	1	0	0	1	0	1	1
Total	1	0	3	1	1	3	9	2	3	3	0	2	7

Table 7. Time Smoothing Evaluation.

Correlation Coefficient	0.010
p value	0.421

4. Modeling of the Relationship between the Type of Bus Traffic Violation and the Number of Accidents

In this paper, ETs, BP Neural Network, SVM, GBDT, and XGBoost models are used for accident prediction modeling. The dependent variable is the number of bus accidents, and the input variable is the number of occurrences of the 12 violation types. The goodness-of-fit (R^2), mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are used as indicators to compare and measure the fitting effects of the five models.

4.1. Extra Trees

The ETs algorithm is based on the RF method, which uses random features and thresholds for decision tree node partitioning to ensure that each decision tree shape and variance are larger and more random. The pseudo-code for ETs is shown in Appendix A.1.

The MSE was selected as the node-splitting evaluation criterion. The number of decision trees was set to 100, and a put-back sampling approach was adopted. According to the data scale of this study, the minimum number of samples for internal node-splitting was set to 2, the minimum and maximum numbers of leaf nodes were set to 1 and 50, respectively, and the maximum depth of the tree was set to 10. No sample weights were introduced because the deviation of the data distribution was insignificant.

4.2. BP Neural Network

A BP neural network is a typical multilayer forward neural network with an input layer, multiple hidden layers (one or more layers), and an output layer. The layers in the BP Neural Network are fully connected, and the neurons in the same layer are not interconnected. The pseudo-code of the BP Neural Network is shown in Appendix A.2.

According to the findings of [29], ReLU (Rectified Linear Unit) was used as the activation function. For the solver, lbfgs was selected, which is an improvement on the second-order Taylor expansion for local approximate average loss that can accelerate the model convergence. The learning rate, the L2 regular term, and the number of iterations were set to 0.1, 1, and 1000, respectively, while the number of hidden layers of one neuron was set to 100.

4.3. Support Vector Machine

The SVM prediction model maps the data into a high-dimensional data feature space using a nonlinear mapping to ensure that the independent and dependent variables have better linear regression characteristics in the high-dimensional data feature space. After the variables are fitted in that feature space, the model then returns to the original space. The SVM is suitable for the small number of sample data. Its pseudo-code is shown in Appendix A.3.

Since different kernel functions are used for different decision functions, the choice of kernel function is significant for SVM. A linear kernel is suitable for linearly divisible data. Therefore, the linear kernel function was used for classification in this study. The parameters were generally adjusted to penalize the parameters $C = 100$ and $\gamma = 0.01$ or $C = 1000$, and $\gamma = 0.01$, where $C = 1000$ and $\gamma = 0.01$ were chosen.

4.4. Gradient Boosted Decision Tree (GBDT)

The GBDT is an additive model based on boosting integration, which is trained with a forward distribution algorithm for greedy learning. Each iteration learns a CART tree to fit the residuals of the prediction results of the previous $t - 1$ trees with the actual values of the training samples. The pseudo-code of GBDT is shown in Appendix A.4.

Usually, squared-error and Friedman-MSE are used as the loss function and the node-splitting evaluation criterion for GBDT models, respectively. The default number of base learners is 100, the learning rate is 0.1, the minimum number of samples for internal node-splitting is 2, the minimum and maximum numbers of samples for leaf nodes are 1 and 50, respectively, and the maximum depth of the tree is 10. When the number of features is significant, only proportional features need to be considered for splitting in each cut to control the tree generation time, which is unnecessary in this paper.

4.5. eXtreme Gradient Boosting

XGBoost is an algorithm partially based on GBDT. In XGBoost, second-order derivatives make the loss function more accurate, regular terms avoid tree overfitting, and Block storage allows parallel computation. The pseudo-code of XGBoost is shown in Appendix A.5.

The base learner was `gbtree`, while `subsample`, `colsample_bytree`, `min_child_weight`, `lambda`, and `alpha` were set to default values of 1, 1, 1, 1, and 0, respectively. Sample weights were not introduced because the data distribution category deviation was insignificant. The `max_depth` was set to 10 as determined by the cross-validation `cv` function.

The specific parameters of the XGBoost model used in this study are shown in Table 8.

To compare the prediction effects of the five models, all the datasets were divided into training and test sets according to the ratio of 7:3, and 10-fold cross-validation was performed on the training set to prevent the overfitting phenomenon due to the unreasonable division of the datasets. The training set was randomly divided into ten mutually exclusive subsets of similar sizes, the concurrent set of nine subsets was used as the training set each time, and the remaining subsets were used as the test sets for ten training and testing sessions. Finally, the mean of the ten evaluation results was obtained.

Table 8. Model parameters setting.

Model	Parameter Name	Value
Extra Trees	Node split evaluation guidelines	MSE
	Minimum sample size for internal node splitting	2
	Minimum number of samples of leaf nodes	1
	Minimum weights of samples in leaf nodes	0
	Maximum depth of the tree	10
	Maximum number of leaf nodes	50
	Threshold for node division impurity	0
	Number of decision trees	100
BP Neural Network	activation function	ReLU
	Solver	lbfgs
	Learning Rate	0.1
	L2 canonical term	1
	Number of iterations	1000
	Number of hidden layer 1 neurons	100
SVM	Penalty Factor	1
	Kernel functions	linear
	Kernel function coefficients	scale
	Nuclear function constants	0
	Maximum number of terms in the kernel Function	3
	Error convergence condition	0.001
GBDT	Maximum number of iterations	1000
	Loss function	Friedman MSE
	Node split evaluation guidelines	Friedman MSE
	Number of base learners	100
	Learning Rate	0.1
	No put-back sampling ratio	1
	Minimum sample size for internal node splitting	2
	Minimum number of samples of leaf nodes	1
	Minimum weights of samples in leaf nodes	0
	Maximum depth of the tree	10
	Maximum number of leaf nodes	50
Threshold for node division impurity	0	
XGBoost	Base Learners	gbtree
	Number of base learners	100
	Learning Rate	0.1
	L1 canonical term	0
	L2 canonical term	1
	Sample Collection Sampling Rate	1
	Tree feature sampling rate	1
	Node feature sampling rate	1
	Minimum weights of samples in leaf nodes	0
	Maximum depth of the tree	10

5. Results and Discussion

5.1. Results

The prediction effect of each model was measured using R^2 , MSE, RMSE, MAE, and MAPE. Table 9 lists the metrics for each model training set, cross-validation set, and test set. The prediction plots for the test data are shown in Figures 5–9.

Table 9. Model evaluation results.

Methods		MSE	RMSE	MAE	MAPE	R ²
Extra Trees	Training set	4.841	2.200	1.558	34.540	0.961
	Cross-validation set	36.695	5.7144	4.104	52.392	0.726
	Test set	20.078	4.481	2.953	102.801	0.703
BP Neural Network	Training set	18.979	4.356	3.277	62.034	0.847
	Cross-validation set	31.987	5.326	4.050	211.567	0.756
	Test set	17.000	4.123	2.654	129.746	0.748
Support vector machines	Training set	27.561	5.250	3.501	112.135	0.838
	Cross-validation set	36.583	5.742	4.432	107.288	0.719
	Test set	17.074	4.132	2.712	199.637	0.787
GBDT	Training set	0.000	0.003	0.002	15.745	1.000
	Cross-validation set	37.920	6.527	5.232	110.922	0.696
	Test set	18.546	4.307	2.797	554.305	0.725
XGBoost	Training set	0.001	0.030	0.008	15.354	1.000
	Cross-validation set	36.211	5.191	4.007	50.685	0.853
	Test set	10.597	3.255	1.572	72.688	0.868

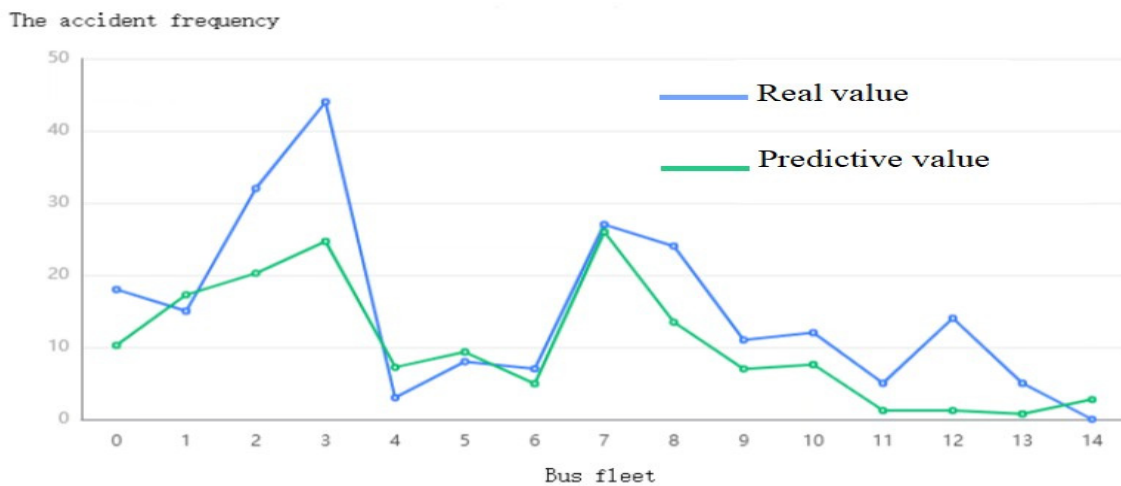


Figure 5. Extra Trees model test data prediction graph.

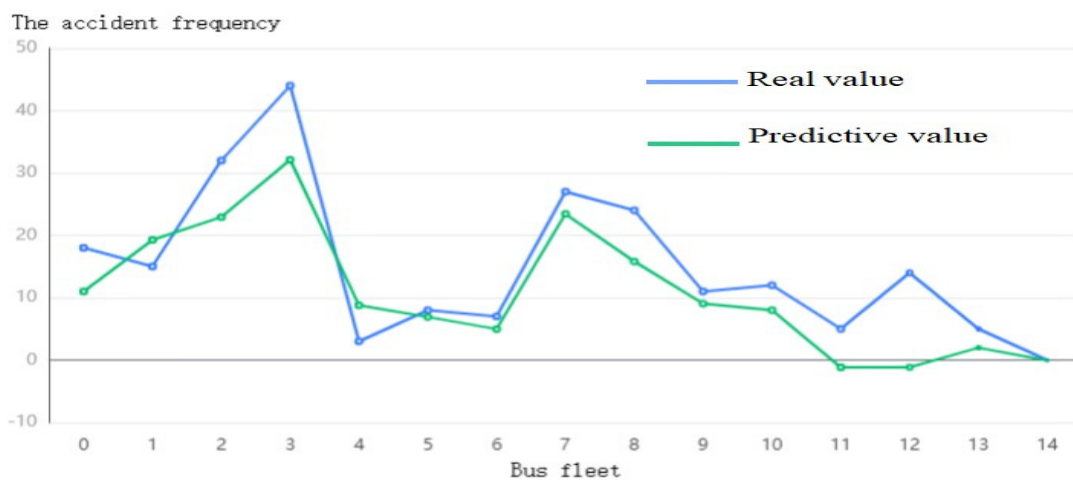


Figure 6. BP Neural Network model test data prediction graph.

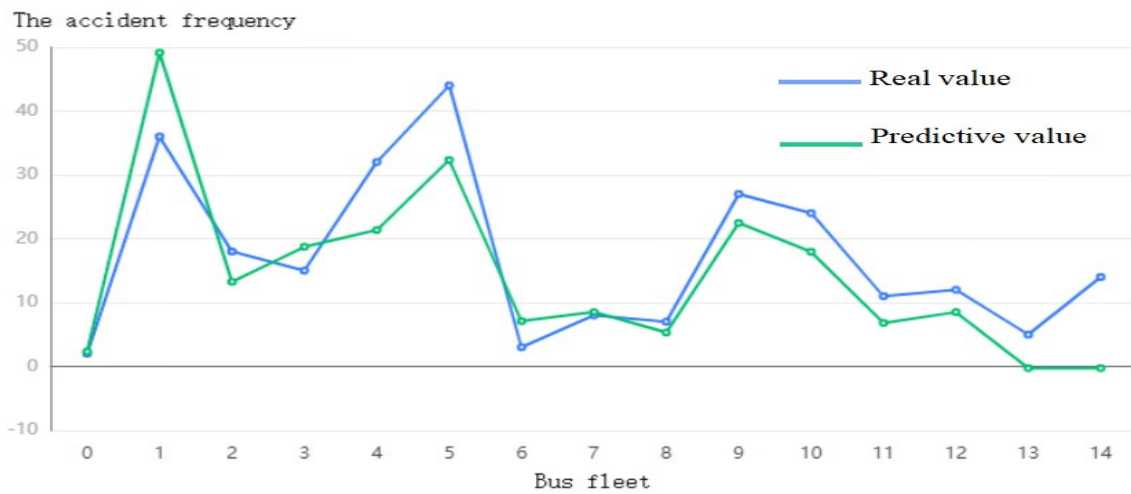


Figure 7. Support Vector Machine model test data prediction graph.

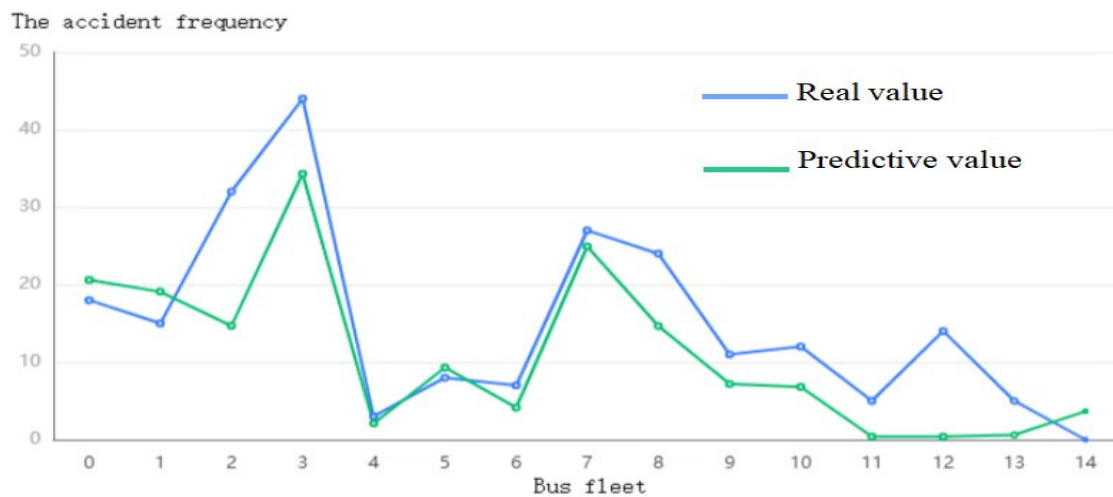


Figure 8. GBDT model test data prediction graph.

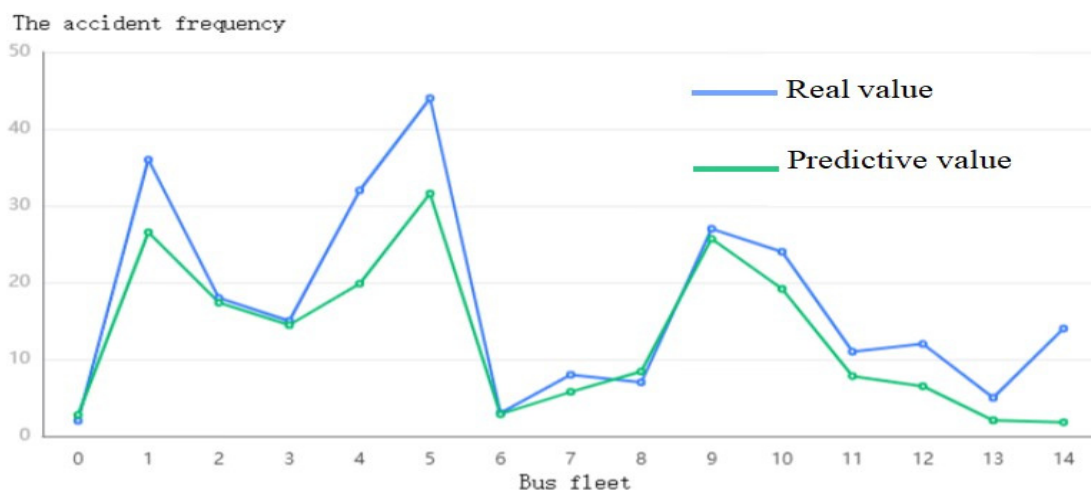


Figure 9. XGBoost model test data prediction graph.

Further, the segmentation effect on the dataset is discussed using 60% of the data as the training set. The results are shown in Table 10. Taking the BP Neural Network and

XGBoost models as examples, when using 70% of the data as the training set, the MSE, RMSE, MAE, MAPE, and R^2 of the BP Neural Network model are 17.000, 4.123, 2.654, 129.746, and 0.748, respectively. The five metrics of the XGBoost model are 10.597, 3.255, 1.572, 72.688 and 0.868. When using 60% of the data as the training set, the MSE, RMSE, MAE, MAPE, and R^2 of the BP Neural Network model are 16.999, 4.1225, 2.652, 129.742, and 0.747, respectively. While the MSE, RMSE, MAE, MAPE, and R^2 of the XGBoost model are 10.596, 3.254, 1.5715, 72.679, and 0.867, respectively. The results demonstrate that the performance of the proposed model is almost independent of the division of the dataset.

Table 10. Model evaluation results (60%).

Model		MSE	RMSE	MAE	MAPE	R^2
Extra Trees	Training set	4.845	2.201	1.5578	34.543	0.9623
	Cross-validation set	36.694	5.7145	4.105	52.395	0.727
	Test set	20.079	4.4812	2.954	102.807	0.704
BP Neural Network	Training set	18.978	4.3558	3.275	62.031	0.848
	Cross-validation set	31.986	5.324	4.048	211.560	0.757
	Test set	16.999	4.1225	2.652	129.742	0.747
SVM	Training set	27.560	5.250	3.501	112.134	0.837
	Cross-validation set	36.582	5.741	4.429	107.283	0.716
	Test set	17.073	4.131	2.711	199.632	0.786
GBDT	Training set	0.000	0.003	0.002	15.7448	1.000
	Cross-validation set	37.919	6.528	5.231	110.918	0.692
	Test set	18.543	4.308	2.793	554.293	0.723
XGBoost	Training set	0.001	0.030	0.008	15.352	1.000
	Cross-validation set	36.210	5.190	4.007	50.681	0.850
	Test set	10.596	3.254	1.5715	72.679	0.867

5.2. Discussion

In this paper, five machine learning methods were used to construct traffic accident prediction models, considering safety violations and service violations. The model with the best prediction effect was chosen by comparing the MSE, RMSE, MAE, MAPE, and R^2 . Table 9 reveals that in terms of MSE, RMSE, MAE, and R^2 , the XGBoost model has the highest prediction accuracy, followed by the BP Neural Network model, SVM model, GBDT model, and ETs model. In terms of the MAPE, the prediction accuracies of the XGBoost model, ETs model, BP Neural Network model, SVM model, and GBDT model are in the order of highest to lowest. Therefore, the XGBoost model was finally chosen.

To investigate the effect of service violations on the predictive effect of the XGBoost model, only safety violations were used in the modeling. The model employed a total of five safety violation types, including bus GPS speeding violation, bus dangerous driving behavior, bus safety operation violation, bus safety violation, and bus general road access violation, as input variables, and the number of fleet traffic accidents as the dependent variable, with 70% of the data serving as the training set. Table 11 displays the final output model evaluation results.

Table 11. Evaluation results of XGBoost regression model considering only safety violations.

	MSE	RMSE	MAE	MAPE	R^2
Training set	0.934	0.966	0.359	32.101	0.995
Test set	37.358	6.112	4.274	485.989	0.534

Comparing the indicators of the XGBoost model in Table 9 with the results in Table 11, the model constructed by considering both safety violations and service violations performs better than the model constructed by considering only safety violations in terms of higher

prediction accuracy and better fit, indicating that the introduction of service violations improved the model's performance.

The XGBoost traffic accident prediction model, which incorporates both safety and service violations, has a high level of accuracy and interpretability.

Figure 10 depicts the critical percentages of various types of violations on the output of the model for predicting traffic accidents. According to the figure, three violations have a greater impact on the occurrence of traffic accidents: violation of bus safety regulations, violation of bus service specifications, and violation of bus safety operation regulations, with 27.9%, 20%, and 16.5%, respectively. It is followed by violations pertaining to bus facilities and equipment, GPS speeding violations, complaints, bus station transport orders, bus cleaning and sanitation, bus signs and markings, violations of general bus road traffic regulations, and dangerous driving behavior.

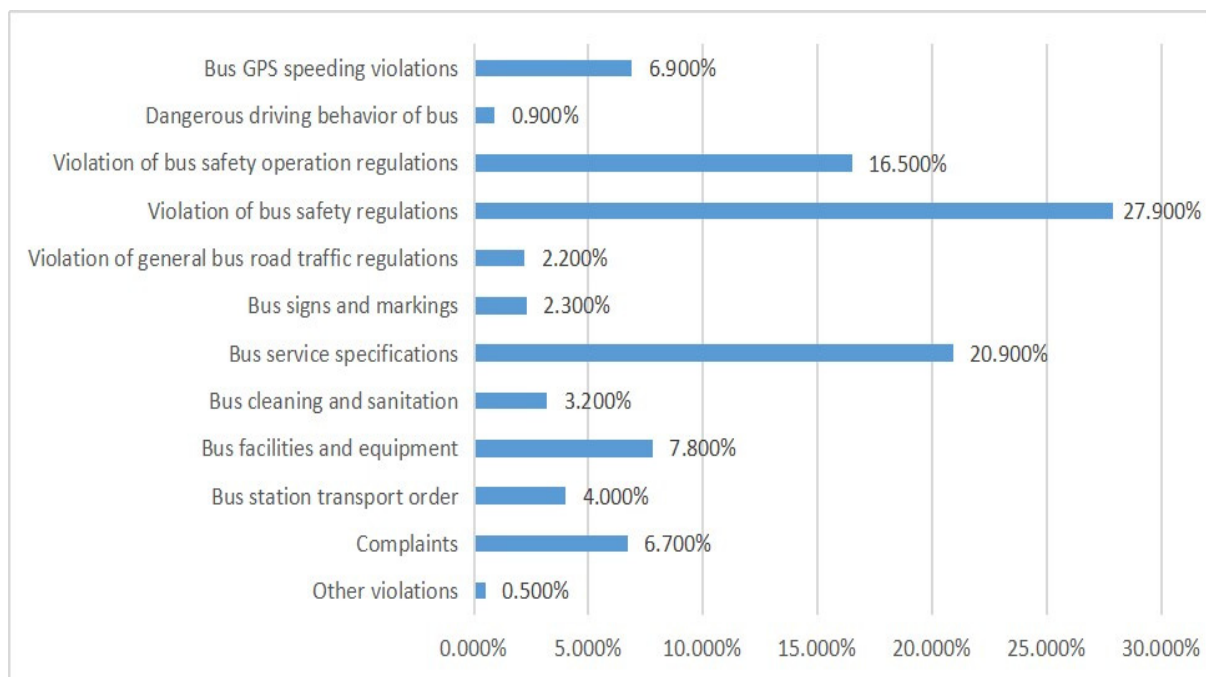


Figure 10. The importance ratio of each violation type.

The main violations of bus safety regulations include driving a sick vehicle, appearing to be “drunk” before starting work, turning off and disrupting GPS or 4G video surveillance equipment, not using a triangle after stopping, not waiting for passengers to pull up and sit down, and other general violations. Similar conclusions have been reached by other researchers regarding the effect of such violations on traffic accidents. Zhong discovered that drunk driving increased the likelihood of car accidents [47]. According to an article by Zhao, 4% of accidents involve vehicle safety, and driving a defective vehicle was more likely to result in traffic accidents [48].

Violations of bus safety operation regulations include driving while using electronic devices such as cell phones, driving with an open door, violating door opening and closing rules, improper use of seat belts, driving with one hand, and taking both hands off the steering wheel while driving. Other studies have confirmed the significance of these violations in the field of crash research. According to a study by Febres, the use of seat belts reduces the likelihood of fatal crashes [49]. Farmer discovered that the risk of collision increased by approximately 17 percent when drivers used cell phones [50].

Complaints, bus service regulations, and bus facilities and equipment are the three most significant influencing factors for service violations, especially, in this paper. The most common violations of bus service regulations involve non-driving time during the service

process for phone calls and other non-work-related activities, non-driving time smoking, failure to turn on the LED display, wearing irregularly or not wearing work number plate or shoulder (arm) badge, failure to dress according to the regulations, and failure to use Mandarin service. The occurrence of bus service regulation violations reflects the laxity and burnout of drivers' work attitudes, and this negative attitude increases the driving risk. Sergio and Bing Li reached the same conclusion regarding the connection between burnout and traffic accidents [51,52]. Most public transportation facilities and equipment violations involve damaged or missing facilities and equipment, such as lamps and seats, which drivers fail to notice or address promptly. It may cause accidents while passengers are in transit. In addition, 40% of bus accidents are in-vehicle accidents, which include, but are not limited to, passenger falls and injuries, as well as sudden accidents resulting from arguments between passengers and drivers. Following these occurrences, some passengers choose to lodge complaints with the bus company. It may be the reason for the high number of bus facility and equipment complaints and violations.

5.3. Management Recommendation

Figure 10 shows that violations of bus safety regulations, violations of bus service specifications, and violations of bus safety operating regulations are the main types of violations affecting traffic accidents. Based on the violations covered by these three violation types, the bus fleet prone to traffic accidents can be divided into two broad categories. The first category belongs to driving misconduct. These bus fleets violate bus safety regulations in addition to operating safety regulations. The second category, work attitude misconduct, is exemplified primarily by frequent bus service code violations. The bus company can add onboard monitoring to regulate violations such as drunk driving, failure to use seat belts, and open-door driving for fleets with improper driving behavior. They can also enrich the driver's training in driving skills by enhancing the details and implementing effective safety education [53]. Meanwhile, the bus company can also increase publicity and education regarding professional identity and responsibility for bus fleets with poor work attitudes.

Moreover, due to the popularity of unmanned ticketing, the bus driver is the only person providing service. It is essential to monitor the driver's service behavior and insist that he/she strictly adhere to service specifications and safety operating procedures. Managers of bus operations can adjust the severity of penalties based on the impact of violations (the number of percentages in Figure 10) on traffic accidents and enhance the driver point system to improve service levels and operational safety.

6. Conclusions

In this paper, Chongqing Liangjiang Public Transportation Co., Ltd. traffic accident data and violation data from January to June 2022 were used to study the violations, which are classified into five categories of safety violations and seven categories of service violations. Then, XGBoost, Extra Trees, BP Neural Network, Support Vector Machine, and Gradient Boosting Tree were utilized to develop five models for predicting traffic accidents. Analyzing the five-evaluation metrics of MSE, RMSE, MAE, MAPE, and R^2 reveals that the XGBoost model outperforms the other four models in terms of both prediction accuracy and fitting effect. Consequently, XGBoost was ultimately chosen to establish the model for predicting bus traffic accidents.

To verify the reasonableness of the service violation introduction, a prediction model using XGBoost based only on the data of safety violations was constructed. The five indicators of MSE, RMSE, MAE, MAPE, and R^2 were 37.358, 6.112, 4.274, 485.989, and 0.534, respectively, when only safety violations were considered, and 10.957, 3.255, 1.572, 72.688, and 0.868, respectively, when both safety violations and service violations were considered. The results demonstrate that the bus company's service regulations have increased passenger comfort while decreasing the risk of traffic accidents.

In addition, XGBoost was utilized to rank the influence of twelve distinct types of violations. The results indicate that three violations have the greatest impact on prediction: violation of bus safety regulations, violation of bus service specifications, and violation of bus safety operation regulations. Finally, based on the ranking of the severity of the impact of violations, the fleets prone to violations were divided into two categories, and governance measures from the perspective of bus operation managers were proposed.

The primary limitation of this study is the insufficient sample size. In this paper, only six months of accident data from a single company were used. If more data become available in the future, the study area will be examined from an international perspective to increase the model's generalizability. Moreover, in this study, only safety violation data and service violation data were used to predict traffic accidents, highlighting the importance of service violations in predicting bus accidents. Nevertheless, road conditions, weather, season, environment, and driver characteristics also influence the incidence of traffic accidents. Next, the authors will rely on a state-funded project to collect data on additional factors and build a traffic accident prediction model based on a hybrid approach by combining multiple influencing factors to reduce the risk of bus accidents.

Author Contributions: Conceptualization, L.Z. (Lili Zheng), L.Z. (Lianxin Zhang) and T.D.; methodology, L.Z. (Lianxin Zhang) and J.X.; analysis, L.Z. (Lianxin Zhang), and K.Z.; writing—original draft preparation, L.Z. (Lianxin Zhang); writing—review and editing, L.Z. (Lili Zheng) and L.Z. (Lianxin Zhang); supervision, Y.L.; project administration, L.Z. (Lili Zheng) and T.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key R&D Program of China(2021YFC3001500), Scientific and Technological Developing Scheme of Jilin Province (No.20200403049SF), and Supported by Graduate Innovation Fund of Jilin University (2022156).

Data Availability Statement: After signing a non-disclosure agreement, the data utilized in the study was obtained from a bus company. As a result, the data from the resource model cannot be shared.

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

Appendix A

Appendix A.1. Pseudocode for Extra Trees

Algorithm A1 Extra Trees splitting algorithm

- 1: **Split_a_node** (S)
- 2: Input: the local learning subset S corresponding to the node we want to split
- 3: Output: a split $[a < a_c]$ or nothing
- 4: If **Stop_split** (S) is TRUE then return nothing.
- 5: Otherwise select K attributes $\{a_1, \dots, a_K\}$ among all nonconstant (*in* S) candidate attributes;
- 6: Draw K splits $\{s_1, \dots, s_K\}$, where $s_i = \text{Pick_a_random_split}(s, a_i), \forall i = 1, \dots, K$;
- 7: Return a split s_* such that $\text{Score}(s_*, S) = \max_{i=1, \dots, K} \text{Score}(s_i, S)$.
- 8:
- 9: **Pick_a_random_split**(S, a)
- 10: Inputs: a subset S and an attribute a
- 11: Output: a split
- 12: Let a_{max}^S and a_{min}^S denote the maximal and minimal value of a in S ;
- 13: Draw a random cut-point a_c uniformly in $[a_{min}^S, a_{max}^S]$;
- 14: Return the split $[a < a_c]$.
- 15:
- 16: **Stop_split**(S)
- 17: Input: a subset S
- 18: Output: a boolean
- 19: If $|S| < n_{min}$, then return TRUE;
- 20: If all attributes are constant in S , then return TRUE;
- 21: If the output is constant in S , then return TRUE;
- 22: Otherwise, return FALSE.

Appendix A.2. Pseudocode for BP Neural Network Appendix

Algorithm A2 BP Neural Network

- 1: Inputs: a training set $D = \{(x_i, y_i)\} (i = 1, 2, \dots, m)$ and learning rate η
- 2: Output: BP Neural Network
- 3: Initializes the connection weights and thresholds of neurons in $(0, 1)$
- 4: Repeat
- 5: For all (x_k, y_k) in D
- 6: Calculate the output of the neural network based on the data of the current sample;
- 7: Calculate the error between the output of the neural network and the label value;
- 8: Calculate the descending gradient of each parameter according to the error;
- 9: Update parameters according to the descent gradient;
- 10: End
- 11: Until meet the end condition

Appendix A.3. Pseudocode for Support Vector Machines

Algorithm A3 Support Vector Machine

- 1: Input: a training set $D = \{(x_i, y_i)\} (i = 1, 2, \dots, m)$
- 2: Output: the classification function model
- 3: Set up an equation $w^T x + b = 0$
- 4: Set up the optimization problem as follows: $\min_{w,b} \left(\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i \right)$

$$\begin{aligned} s.t. & y_i(w^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, i = 1, 2, \dots, N \end{aligned}$$

where: w is the normal vector; b is a constant term; C is the punishment factor; ξ_i is the relaxation variable.

- 5: By converting the above formula into a quadratic programming problem and introducing the corresponding Lagrange function, the classification problem becomes:

$$L(w, b, \lambda) = \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [y_i(w^T x_i + b) - 1 + \xi_i] - \sum_{i=1}^N \beta_i \xi_i$$

where: α_i, β_i are Lagrange multipliers, $\alpha_i \geq 0, \beta_i \geq 0$.

- 6: According to the duality principle, the above formula is changed into:

$$\max_{\alpha} L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^N \alpha_i \alpha_j y_i y_j x_i x_j$$

$$s.t. \sum_{i=1}^N y_i \alpha_i = 0, 0 \leq \alpha_i \leq C$$

In this algorithm, the kernel function is:

$$K(x_j x_i) = x_j^T x_i$$

- 7: After solving α , the model can be obtained by solving w and b .

Appendix A.4. Pseudocode for GBDT Appendix

Algorithm A4 Gradient Tree Boosting Algorithm

- 1: Input: a training set $D = \{(x_i, y_i)\} (i = 1, 2, \dots, m)$
- 2: Output: $\hat{f}(x) = f_M(x)$.
- 3: Initialize $f_0(x) = \operatorname{argmin}_{\gamma} \sum_{i=1}^N L(y_i, \gamma)$.
- 4: For $m = 1$ to M :
- 5: For $i = 1, 2, \dots, N$ compute $r_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f=f_{m-1}}$
- 6: Fit a regression tree to the targets r_{im} giving terminal regions $R_{jm}, j = 1, 2, \dots, J_m$.
- 7: For $j = 1, 2, \dots, J_m$ compute $\gamma_{jm} = \operatorname{argmin}_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)$.
- 8: Update $f_m(x) = \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$. Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$.

Appendix A.5. Pseudocode for XGBoost

Algorithm A5 eXtreme Gradient Boosting1: Input: a training set $D = \{(x_i, y_i)\} (i = 1, 2, \dots, m)$

2: Output: the classification function model

3: Build the XGBoost model:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in F (i = 1, 2, \dots, n)$$

where:

$$F = \left\{ f(x) = w_{q(x)} \right\} (q: R^m \rightarrow \{1, 2, \dots, T\}, w \in R^T)$$

F is the set of CART decision tree structures, q is the tree structure of the sample mapped to the leaf node, T is the number of child nodes, and w is the real number fraction of the leaf node.

4: Construct objective function:

$$Obj = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T$$

where:

$$g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}), h_i = \partial_{\hat{y}^{(t-1)}}^2 l(y_i, \hat{y}^{(t-1)})$$

5: Use **exact greedy algorithm** for split finding

6: The optimal XGBoost model is built by searching for the optimal tree structure using the objective function and placing it into the existing model.

7:

8: **Exact Greedy Algorithm for Split Finding**9: Input: I , instance set of current node10: Input: d , feature dimension

11: Output: Split with max score

12: $g_{ian} \leftarrow 0$ 13: $G \leftarrow \sum_{i \in I} g_i, H \leftarrow \sum_{i \in I} h_i$ 14: For $k = 1$ to m do15: $G_L \leftarrow 0, H_L \leftarrow 0$ 16: For j in sorted($I, by x_{jk}$) do17: $G_L \leftarrow G_L + g_j, H_L \leftarrow H_L + h_j$ $G_R \leftarrow G - G_L, H_R \leftarrow H - H_L$

$$score \leftarrow \max \left(score, \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{G^2}{H + \lambda} \right)$$

18: End

19: End

References

1. The Ministry of Transport and Twelve Other Departments and Units on the Issuance of Green Travel Action Plans (2019–2022). Available online: https://xxgk.mot.gov.cn/2020/jigou/ysfws/202006/t20200623_3315926.html (accessed on 5 November 2022).
2. Sawalha, Z.; Tarek, S. Evaluating Safety of Urban Arterial Roadways. *J. Transp. Eng.* **2001**, *127*, 151–158. [CrossRef]
3. Miaou, H. Modeling Vehicle Accidents and Highway Geometric Design Relationships. *Accid. Anal. Prev.* **1993**, *25*, 689–709. [CrossRef] [PubMed]
4. Sawalha, Z. Transferability of accident prediction models. *Science* **2006**, *44*, 209–219. [CrossRef]
5. Lapedest, F. Accident Prediction Models for Urban Roads. *Accid. Anal. Prev.* **2003**, *35*, 273–285.
6. Shawky, M.; Al-Badi, Y.; Sahnoun, I.; Al-Harathi, H. The Relationship between Traffic Rule Violations and Accident Involvement Records of Drivers. *Adv. Intell. Syst. Comput.* **2017**, *484*, 745–755.
7. 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]
8. Feng, S.; Li, Z.; Ci, Y.; Zhang, G. Risk factors affecting fatal bus accident severity: Their impact on different types of bus drivers. *Accid. Anal. Prev.* **2016**, *86*, 29–39. [CrossRef]
9. Holubowycz, O.T.; Kloeden, C.N.; McLean, A.J. Age, sex, and blood alcohol concentration of killed and injured drivers, riders, and passengers. *Accid. Anal. Prev.* **1994**, *26*, 483–492. [CrossRef] [PubMed]
10. Delen, D.; Sharda, R.; Bessonov, M. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accid. Anal. Prev.* **2006**, *38*, 434–444. [CrossRef] [PubMed]
11. Pugh, N.; Park, H. Prediction of Red-Light Running using an Artificial Neural Network. In Proceedings of the IEEE SOUTH-EASTCON, St. Petersburg, America, 19–22 April 2018.

12. Tian, Y.; Robinson, J.D. Predictors of Cell Phone Use in Distracted Driving: Extending the Theory of Planned Behavior. *Health Commun.* **2017**, *32*, 1066–1075. [[CrossRef](#)]
13. Mark, J.M.; Sullman, M.; Karl, B.P. Aberrant driving behaviours amongst New Zealand truck drivers. *Transp. Res. F* **2022**, *5*, 217–232.
14. Zhang, G.; Yau, K.K.W.; Chen, G. Risk factors associated with traffic violations and accident severity in China. *Accid. Anal. Prev.* **2013**, *59*, 18–25. [[CrossRef](#)] [[PubMed](#)]
15. Ayuso, M.; Guillén, M.; Alcañiz, M. The impact of traffic violations on the estimated cost of traffic accidents with victims. *Accid. Anal. Prev.* **2010**, *42*, 709–717. [[CrossRef](#)] [[PubMed](#)]
16. Ebrahemzadieh, M.; Giahi, O.; Foroginasab, F. Analysis of Traffic Accidents Leading to Death Using Tripod Beta Method in Yazd, Iran. *Promet.* **2016**, *28*, 291–297. [[CrossRef](#)]
17. Alver, Y.; Mutlu, M.M.; Demirel, M.C. Young driver behaviors analysis: Relationship between traffic rule violations and socio-demographic structure. In Proceedings of the 18th International Conference of Hong Kong Society for Transportation Studies, HKSTS 2013–Travel Behaviour and Society, Hong Kong, China, 14–16 December 2013; pp. 391–398.
18. Mao, L.; Zhang, J.; Duan, L.; Mao, E. Analysis on road traffic accidents and its influencing factors in rural area of Guangxi. In Proceedings of the 8th International Conference of Chinese Logistics and Transportation Professionals—Logistics: The Emerging Frontiers of Transportation and Development in China, Chengdu, China, 8–10 October 2008; pp. 4577–4582.
19. Hadji, H.M.; Mahdavi, A.; Jabbari, N.M. Validation of the influencing factors associated with traffic violations and crashes on freeways of developing countries: A case study of Iran. *Accid. Anal. Prev.* **2018**, *121*, 358–366. [[CrossRef](#)]
20. 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](#)]
21. Anebonam, U.; Okoli, C.; Ossai, P.; Ilesanmi, O.; Nguku, P.; Nsubuga, P.; Abubakar, A.; Oyemakinde, A. Trends in road traffic accidents in Anambra State, South Eastern Nigeria: Need for targeted sensitization on safe roads. *Pan Afr. Med. J.* **2019**, *32*, 12. [[CrossRef](#)]
22. Assum, T. Reduction of the blood alcohol concentration limit in Norway—Effects on knowledge, behavior and accidents. *Accid. Anal. Prev.* **2010**, *42*, 1523–1530. [[CrossRef](#)]
23. Shinar, D.; Schechtman, E.; Compton, R. Self-reports of safe driving behaviors in relationship to sex, age, education and income in the US adult driving population. *Accid. Anal. Prev.* **2001**, *33*, 111–116. [[CrossRef](#)]
24. Iversen, H. Risk-taking attitudes and risky driving behaviour. *Transp. Res. Part F* **2004**, *7*, 135–150. [[CrossRef](#)]
25. Maycock, G. Sleepiness and driving: The experience of U.K. car drivers. *Accid. Anal. Prev.* **1997**, *29*, 453–462. [[CrossRef](#)] [[PubMed](#)]
26. Feraud, I.S.; Lara, M.M.; Naranjo, J.E. A fuzzy logic model to estimate safe driving behavior based on traffic violation. In Proceedings of the 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM), Salinas, Ecuador, 16–20 October 2017; pp. 1–6. [[CrossRef](#)]
27. Chen, M.; Zhou, L.; Choo, S.; Lee, H. Analysis of Risk Factors Affecting Urban Truck Traffic Accident Severity in Korea. *Sustainability* **2022**, *14*, 2901. [[CrossRef](#)]
28. Lorenzo, M.; Andrea, M.; Marcello, O. An analysis of urban collisions using an artificial intelligence model. *Accid. Anal. Prev.* **1999**, *31*, 705–718.
29. Mostafa, S.M.; Salem, S.A.; Habashyis, S.M. Predictive Model for Accident Severity. *IAENG Int. J. Comput. Sci.* **2022**, *49*, 110–124.
30. Wu, D.; Wang, S. Comparison of road traffic accident prediction effects based on SVR and BP neural network. In Proceedings of the 2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence, ICIBA 2020, Chongqing, China, 6–8 November 2020.
31. Ali, G.A.; Tayfour, A. Characteristics and Prediction of Traffic Accident Casualties In Sudan Using Statistical Modeling and Artificial Neural Networks. *Int. J. Transp. Sci. Technol.* **2012**, *1*, 305–317. [[CrossRef](#)]
32. Farhangi, F.; Sadeghi-Niaraki, A.; Razavi-Termeh, S.V.; Choi, S.-M. Evaluation of Tree-Based Machine Learning Algorithms for Accident Risk Mapping Caused by Driver Lack of Alertness at a National Scale. *Sustainability* **2021**, *13*, 10239. [[CrossRef](#)]
33. Nouh, R.; Singh, M.; Singh, D. Safedrive: Hybrid recommendation system architecture for early safety predication using internet of vehicles. *Sensors* **2021**, *21*, 3893. [[CrossRef](#)]
34. Kashifi, M.T.; Ahmad, I. Efficient Histogram-Based Gradient Boosting Approach for Accident Severity Prediction with Multisource Data. *Transp. Res. Rec.* **2022**, *2676*, 236–258. [[CrossRef](#)]
35. Yang, J.; Duan, A.; Li, K.; Yin, Z. Prediction of Vehicle Casualties in Major Traffic Accidents Based on Neural Network. In Proceedings of the AIP Conference Proceedings, Wuhan, China, 19–20 January 2019.
36. Yu, B.; Wang, Y.T.; Yao, J.B.; Wang, J.Y. A comparison of the performance of ANN and SVM for the prediction of traffic accident duration. *Neural Network World.* **2016**, *26*, 271–287. [[CrossRef](#)]
37. Zhi, L.; Pan, L.; Wei, W.; Cheng, X. Using support vector machine models for crash injury severity analysis. *Accid. Anal. Prev.* **2012**, *45*, 478–486.
38. Cong, C.; Guohui, Z.; Zhen, Q.; Rafiqul, A.; Zong, T. Investigating driver injury severity patterns in rollover crashes using support vector machine models. *Accid. Anal. Prev.* **2016**, *90*, 128–139.
39. Mehta, K.; Jain, S.; Agarwal, A.; Bomnale, A. Road Accident Prediction Using Xgboost. In Proceedings of the 2022 International Conference on Emerging Techniques in Computational Intelligence, ICETCI 2022, Hyderabad, India, 25–27 August 2022.

40. Boo, Y.; Choi, Y. Comparison of mortality prediction models for road traffic accidents: An ensemble technique for imbalanced data. *BMC Public Health*. **2022**, *22*, 1476. [[CrossRef](#)]
41. Hussain, G.; Batool, I.; Kanwal, N.; Abid, M. The moderating effects of work safety climate on socio-cognitive factors and the risky driving behavior of truck drivers in Pakistan. *Transp. Res. F*. **2019**, *62*, 700–715. [[CrossRef](#)]
42. Wang, Q.; Zhang, W.; Yang, R.; Huang, Y.; Zhang, L.; Ning, P.; Cheng, X.; Schwebel, D.C.; Hu, G.; Yao, H. Common traffic violations of bus drivers in urban China: An observational study. *PLoS One* **2015**, *10*, e0137954. [[CrossRef](#)] [[PubMed](#)]
43. Passenger Transport Services Specifications for Urban. Available online: <https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=B2E4E7C08AA5BF70C3949B2BCFCAAF> (accessed on 19 January 2023).
44. Sikdar, P.; Rabbani, A.; Dhapekar, N.K. Hypothesis of data of road accidents in India-review. *Int. J. Civ. Eng. Technol.* **2017**, *8*, 141–146.
45. Hauke, J.; Kossowski, T. Comparison of Values of Pearson’s and Spearman’s Correlation Coefficients on the Same Sets of Data. *Quaest. Geogr.* **2011**, *30*, 87–93. [[CrossRef](#)]
46. Gu, Y.; Deng, L. STAGCN: Spatial-Temporal Attention Graph Convolution Network for Traffic Forecasting. *Mathematics*. **2022**, *10*, 1599. [[CrossRef](#)]
47. Zhong, M.; Hong, H.; Wu, P.; Fang, Q. Traffic accident tendency measurement method for drunk driving based on uchida kraepelin psychological test. *Shanghai Jiaotong Daxue Xuebao/J. Shanghai Jiaotong Univ.* **2016**, *50*, 413–418.
48. Zhao, G.; Li, J.; Zhou, W. Regression analysis of association between vehicle performance and driver casualty risk in traffic accidents. In Proceedings of the ICTIS 2015—3rd International Conference on Transportation Information and Safety, Wuhan, China, 25–28 June 2015.
49. Febres, J.D.; García-Herrero, S.; Herrera, S.; Gutiérrez, J.M.; López-García, J.R.; Mariscal, M.A. Influence of seat-belt use on the severity of injury in traffic accidents. *Eur. Transp. Res. Rev.* **2020**, *12*, 9. [[CrossRef](#)]
50. Farmer, C.M.; Klauer, S.G.; McClafferty, J.A.; Guo, F. Relationship of Near-Crash/Crash Risk to Time Spent on a Cell Phone While Driving. *Traffic Inj. Prev.* **2015**, *16*, 792–800. [[CrossRef](#)]
51. Useche, S.; Alonso, F.; Cendales, B.; Autukeviciute, R.; Serge, A. Burnout, Occupational Stress, Health and Road Accidents among Bus Drivers: Barriers and Challenges for Prevention. *J. Environ. Occup. Sci.* **2017**, *6*, 1–7. [[CrossRef](#)]
52. Li, B.; Liu, J.X. Research on Preventive Management of Freeway Traffic Accident. *Adv. Mater. Res.* **2013**, *779*, 763–768. [[CrossRef](#)]
53. Wang, L.; Wang, Y.; Shi, L.; Xu, H. Analysis of risky driving behaviors among bus drivers in China: The role of enterprise management, external environment and attitudes towards traffic safety. *Accid. Anal. Prev.* **2022**, *168*, 106589. [[CrossRef](#)] [[PubMed](#)]

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