

Traffic Accident Density Prediction Considering Injury Severity Based on Random Forest and GAM

Yizhen Wang^{1,#}, Jinglin Cao^{2,#,*}, Liqi Kou^{3,#}

¹ Central University of Finance and Economics, Beijing, China

² Ningbo University of Technology, Ningbo, China

³ Jiangxi University of Finance and Economics, Nanchang, China

*Corresponding Author Email: mrcaojl@gmail.com

#These authors contributed equally

Abstract. Traffic accident density prediction is an significant topic of road safety. This study uses some data in the census tracts in California to develop traffic accident density models using two different modeling approaches. In this paper, the dependent variable was the density of traffic crash which is a measure of the relative distribution of traffic crashes. The random forest regression was used to predict crash density. But many machine learning models like this are like a black-box, which can't find the causal relationship between features and targets. For many practical problems, it is often more important to explain why a phenomenon happens than improving the model's predictability. As a result, differed from pure predictive study, the generalized additive model (GAM) was introduced which was an interpretable statistical model to achieve explanatory analysis. The GAM explores which independent variables will have an impact on response variables and the extent of their impact. Thus, the traffic accident density prediction of this practical problem has more credibility analysis.

Keywords: traffic accident density, road safety, random forest regression, generalized additive model, cause-and-effect study.

1. Introduction

With the increasingly common travel of motor vehicles, there are traffic accidents in the world every year, causing millions of people to be injured and killed. Every year, 1.35 million people around the world are killed on the road [1]. It is estimated that from 2015 to 2030, fatal and non-fatal road injuries will cause the world economy to lose about 1.8 trillion US dollars [2]. Consequently, the traffic accident has become a global problem.

The purpose of the prediction of traffic accidents is to master the future development trend of traffic accidents, so as to take effective countermeasures to reduce the number of accidents and reduce economic losses. Prediction to predict traffic accidents alone does not help improve road safety issues, so we need to analyze the influencing factors of traffic accidents. The current research on this type is mainly focused on correlation analysis. Many researchers' approaches are to use statistical analysis methods for the analysis of the influencing factors of traffic accidents. By building a mathematical statistical model, some factors, such as traffic flow and speed limit, have a significant correlation with the number of traffic accidents. Correction is a statistical concept, that is, changes in a certain factor will cause other factors to change, but the change of the former changes in the former cannot be determined. There is no doubt that causality will bring us certainty and help the solution of road security issues. Therefore, in this paper, we introduced causal research.

We chose a random forest algorithm in machine learning and GAM in the econometrics to establish a prediction model respectively. Although the random forest model has a good prediction performance, it is not explained. And GAM just made up for this. In the subsequent experiments, we used the data of the census tracts collected from California in the United States to develop a machine learning and statistical model and compared the predictive performance of the two through coefficient of determination and root mean square error. In addition, interpretability is also important. Through

the explanation and analysis of GAM, we have concluded that the density of traffic accidents (especially causing severe damage and death) in a census tract with large population density are more likely to occur in traffic accidents in the future. This may be related to the local incomplete transportation infrastructure and defective road geometric design. Through this study, we hope to provide the California government with the overall situation of the density of traffic accidents in all the census tracts and help them better improve their road safety.

Our contribution is two-fold. First, we show the importance of traffic accident injury data, which is reflected in the feature importance of random forest model and the result image of GAM. This is not available in many related studies. Second, the innovation of this paper lies in the use of econometric methods with strong interpretability of models to carry out structural analysis. Our results evidence that the population density of a census tract combined with the density of historical traffic accident injuries has a causal relationship with the traffic accident density. And we provide constructive suggestions to the local government.

The rest of this paper is constructed as follows. Section 2 reviews the existing literature corresponding to the traffic accident prediction. Section 3 presents the methodology of the study, including introduction to data set, data preprocessing and two prediction models. Section 4 reports the descriptive statistics and predictions. Section 5 summarizes the results. Section 6 concludes the study with main findings and directions for future work.

2. Literature review

In order to exactly evaluate the relationship among injury-severity-related crashes density and each kind of factors, we first review the factors contributed to a normal traffic accident in general. Quantities of studies have investigated the contributing factors to traffic crashes. Based on existing datasets, these studies usually consider the factors into two categories: one is the external condition that under which circumstance the accidents happened, another is the intrinsic characteristic of accident itself. The external condition include the geographic environment such as the census tract [3], roadway geometry [4], traffic analysis zone [5,6], and road type like whether the accident happened on the main roads in cities or on country roads. The intrinsic characteristics of crashes are the number of injured people, ethnic identity of casualties [7], the type of casualties like whether they are pedestrians or bicyclists. For the two aspects of factors mentioned above distinctly described the full view of traffic accidents, it's better to combine both external conditions and the intrinsic characteristics to give a comprehensive knowledge of traffic crashes which is helpful for the policy makers take long-term advice.

However, the studies mentioned above don't take injury severity, a signal internal factor of traffic accidents, into account. Even if the former articles take the injury severity as one of the means to evaluate the severity of crashes, take Tang(2019) [8], Heejin Jeong(2018) [9] and Zheng(2018) [10] for instance, they divided the origin data based on injury severity into several levels ,they use number and percentage to describe the proportion of each injury severity level ,however, it doesn't mean that the numbers can depict the exact situation of an accident contains different levels of injuries. Besides, it is too obscure to give some suggestions to the policy makers through these numbers to reduce the density of traffic crashes. In conclusion, to access the severity of traffic crashes and reduce both the numbers and percentage of accidents, it's essential to combine several levels of injury severity into a certain attribute to facilitate the description of a complex traffic crash.

3. Methodology

3.1. Data Collection and Pre-processing

The study area for this paper is 7,998 census tracts in the California, containing data for a total of ten years from 2011 to 2020.

This dataset contains fifteen independent variables and one dependent variable for prediction. Among all the independent variables, the density of injury severity is rather special because it is indirectly derived from our original data and we have described the process of its creation in INTRODUCTION. Finally, the names and descriptions of all the variables are shown in the Table 1:

Table 1. Variables Descriptions for The Experiment

Variable	Description
crash_dens	Traffic accidents density (dependent variable)
AADT_dens	The density of Average Annual Daily Traffic
AADTT_dens	The density of Annual Average Daily Truck Traffic
lowest_not_NHS_dens	The density of lowest grade highway not in the NHS (National Highway System) system
Interstate_dens	The density of interstate
NIS_dens	The density of Non-Interstate STRAHNET
SC_dens	The density of STRAHNET Connector
lowest_NHS_dens	The density of lowest grade highway in the NHS (National Highway System) system
aIC_dens	The density of Approved Intermodal Connector
PAaM_dens	The density of Principal Arterial
pop_dens	Population density
emp_dens	Employment density
mfg_emp_dens	Employment density of Manufacturing
rt_emp_dens	Employment density of Retail Trade
tw_emp_dens	Employment density of Transportation and Warehousing
extent_injury_dens	A number representing road safety grade

Selecting relevant, informative, and independent features is a critical step in building a predictive model. The model can learn well from data described by quality features, and even if it is not the optimal model, quality features can get better results. Through feature selection, the model can also be run more efficiently and reduce overfitting.

First, we expect that each feature should be independent of the others, i.e., there is no multicollinearity between them. We use a variance inflation factor (VIF) to determine multicollinearity and remove features based on high VIF values. Therefore, we calculate the VIF of each feature and remove the features with VIF greater than 10. The Fig. 1 shows the features with VIF less than 10.

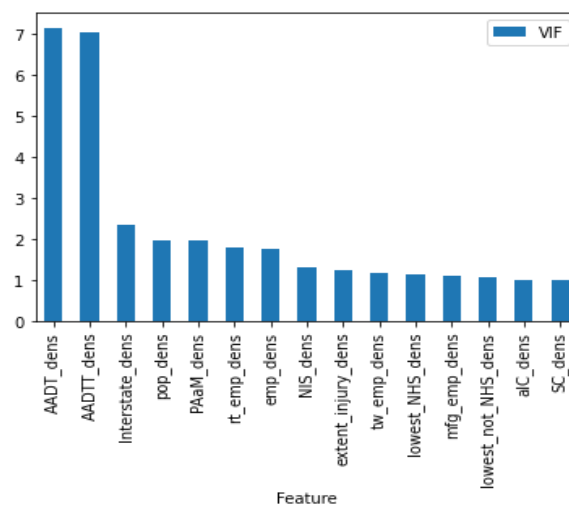


Fig 1. VIF of all features

From the above figure, we can see that there is no multicollinearity in our features.

Then we start to remove some features that are almost irrelevant to the prediction target. The random forest algorithm has a built-in method for detection of important features, which allows us to achieve this goal.

The idea of random forest for feature importance assessment is to calculate how much each feature contributes to the different decision trees of the random forest, and then take its average and compare the magnitude of contribution between different features. We train a simple unoptimized random forest regression model, which in turn yielded the feature_importances_ attribute, and then rank and plot it which is shown in the Fig. 2.

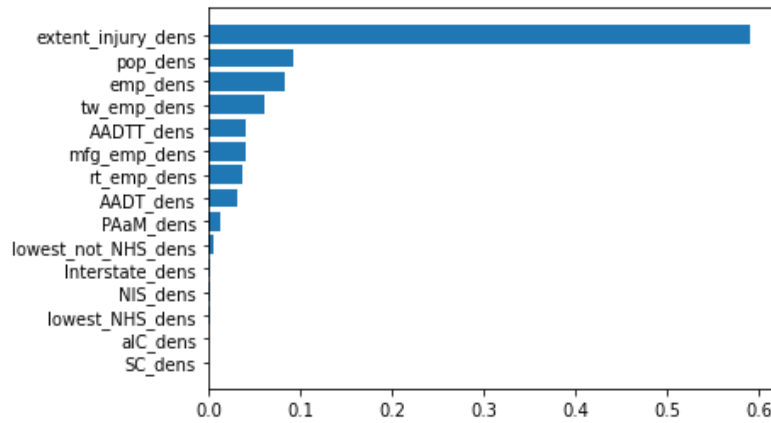


Fig 2. Feature importance plot

Considering the results of the above figure, we decide to remove the last six features with very low importance.

Meanwhile, after our statistics, the proportion of zero values of these six features are 77.007%, 83.108%, 93.386%, 86.284%, 98.150%, 99.000% respectively. Therefore, we decided to remove these six features. Finally, the dataset we use for training has only nine independent variables and one dependent variable.

3.2. Random Forest Model

Random forest is a supervised learning algorithm that uses ensemble learning method for regression. It can be used to solve regression and classification problems [11]. In this paper, we use random forest regression to build a regression forecasting model.

The random forest regression algorithm is a combinatorial model consisting of a set of regression decision subtrees $\{f(x, \theta_t), t=1, 2, \dots, T\}$, where θ_t is a random variable subject to independent identical distribution, x represents the independent variable, T denotes the number of regression decision trees and $\bar{f}(x)$ is the average of all decision tree predictions. We take the average of each decision tree as the prediction result of random forest regression:

$$\bar{f}(x) = \frac{1}{T} \sum_{t=1}^T \{f(x, \theta_t)\} \quad (1)$$

Random forest is an ensemble learning model using the Bagging [12] method. The Bagging method randomly draws multiple training samples from the original sample and each training sample is equal to the original sample size. And it constructs a regression decision subtree for each training sample, and finally takes the average of each tree as the final prediction result. In an ideal case, about 36.8 % of the total training data forms the OOB (out-of-bag) sample. This can be shown as follows:

If we suppose N is the number of samples in the training dataset, the probability of not picking a row in a random draw is $\frac{N-1}{N}$. Then the probability of not picking N rows in random draws using

sampling-with-replacement is $\left(\frac{N-1}{N}\right)^N$, which with infinity N equals:

$$\lim_{N \rightarrow \infty} \left(1 - \frac{1}{N}\right)^N = \frac{1}{e} \approx 0.368 \quad (2)$$

Random forest also utilizes random subspace method. In the process of constructing the regression decision subtree, each splitting node randomly selects a feature subspace from the total feature space as the candidate feature set of the node and selects the best feature for splitting. This method ensures that the nodes between trees and the feature subsets between each tree node are different, but also ensures the independence and diversity of the trees, which in turn improves the randomness of node splitting. In random forest regression, the number of features in the regression decision subtree and feature subspace determines the final prediction performance of the model.

3.3. Generalized Additive Model

The generalized additive model (GAM) [13] represents a non-parametric smooth regression framework serving a set of response variables from a family of exponential distributions (such as normal, exponential, Poisson, binomial, negative binomial, etc.). The generalized form is as follows:

$$g(\mu_Y) = \beta_0 + f_1(X_1) + f_2(X_2) + \dots + f_n(X_n) \quad (3)$$

$f_n(X)$ is a nonparametric function, while the response variable Y follows some distribution in the family of exponential distributions (not limited to normality). $g(\mu_Y)$ represents the function of the conditional mean of the response variable Y (exponential, Poisson, binomial, negative binomial, etc.), also known as the connection function. And β_0 is the intercept term. We convert the conditional mean of all kinds of non-normal exponential distribution family response variables into the conditional mean of normal form to establish the non-parametric additive response relationship with independent variables.

The advantage of generalized additive models lies in their outstanding ability to solve the highly nonlinear and nonmonotonic relationship between response variables and predictors. GAM is a model based on data that determines the relationship between response variables and predictors, it allows us to use nonlinear smoothing terms to fit the model without prior knowledge of the relationship between dependent variables and independent variables.

4. Results analysis

4.1. Statistical Analysis

Descriptive statistics are brief informational coefficients that summarize a given data set, which describes the characteristics of the data set. In this paper, we choose the mean as a measure of central tendency, and the standard deviation is used as a measure of variability. At the same time, the minimum value and maximum value of all variables are listed in the Table 2.

Table 2. Descriptive statistics

Variable	Min	Max	Mean	Std.D
AADT_dens	0	1.294	0.030	0.078
AADTT_dens	0	0.065	0.002	0.004
PAaM_dens	0	2.704×10^{-6}	1.171×10^{-7}	1.722×10^{-7}
pop_dens	0	0.006	3.022×10^{-4}	3.412×10^{-4}
emp_dens	0	0.017	1.079×10^{-4}	3.982×10^{-4}
mfg_emp_dens	0	8.071×10^{-4}	4.523×10^{-6}	1.957×10^{-5}
rt_emp_dens	0	0.001	1.087×10^{-5}	2.963×10^{-5}
tw_emp_dens	0	0.001	2.200×10^{-6}	1.894×10^{-5}
extent_injury_dens	0	1.443×10^{-5}	6.543×10^{-8}	3.979×10^{-7}
crash_dens	0	6.328×10^{-6}	3.217×10^{-8}	1.397×10^{-7}

4.2. Random Forest Regression Model Results

Random forest is a machine learning algorithm, and the setting of the hyperparameters of the algorithm not only affects the prediction accuracy of the model, but also influences the good or bad training effect of the model. Therefore, hyperparameter tuning is necessary. Strategies commonly used to search for optimal configurations of hyperparameters include Grid Search with Cross Validation and Randomized Search with Cross-Validation. The grid search method divides each dimensional hyperparameter into multiple grid points in advance, and then performs an exhaustive search for the combination of grid points in all dimensions. This method is widely used in practice because of its simplicity and ease of operation. Bergstra and Bengio proposed an equally simple hyperparameter tuning method for random search [14], in which test points are randomly added to the search space. The main advantage of the random search over the grid search method is that it has a much larger search space. For each hyperparameter dimension, the random search will no longer be limited to a finite number of grid points, but can try more different values.

In this paper, we combine the advantages of both methods and avoid the shortcomings by first finding the approximate range of the optimal configuration of hyperparameters through Randomized Search CV, and then using Grid Search CV to search for the optimal configuration of hyperparameters within that range.

Table 3. Hyperparameter table of random forest

Hyperparameter	Description	Optimal value
n_estimators	Number of trees in the forest	1000
max_depth	The maximum depth of the tree	490
min_samples_split	The minimum number of samples required to split an internal node	5
min_samples_leaf	The minimum number of samples required to be at a leaf node	2

The Table 3 shows the optimal combination of hyperparameters, which will be used to generate an optimal random forest regression model. We can then evaluate the performance of the model.

There are many metrics for regression models, and here we choose two representative ones: R^2 (coefficient of determination) and RMSE (root mean square error). These are calculated as follows.

Let y be the actual value of the target variable and the mean value of y is $\bar{y}$ and the predictive value is $\hat{y}$. Then the variability of the data set can be measured with two formulas:

The residual sum of squares (RSS):

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

The total sum of squares (TSS):

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (5)$$

The definition of the coefficient of determination is

$$R^2 = 1 - \frac{RSS}{TSS} \quad (6)$$

The root mean squared error (RMSE) is

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

The performance of the random forest regression model using the above metrics shows $R^2 = 0.604$ and $RMSE = 6.629 \times 10^{-8}$. The model's feature importance suggests that the density of injury extent is the most important feature.

4.3. Generalized Additive Model Results

Before constructing the GAM, it is necessary to explore the nonlinear relationship between the predictive variables and the response variables. The generalized additive model Fit1 is established on the whole dataset. The descriptive function of R language will test the hypothesis of the generalized additive model, and the null hypothesis is that there is a linear relationship between the predictive variables and response variables. The alternative hypothesis is that there is a nonlinear relationship between the predictor variables and the response variables. After calculation, p values of 9 variables were all less than 0.01. The results of ANOVA test are shown in the Table 4:

Table 4. Anova for nonparametric effects

	Df	F	Pr (F)
(Intercept)			
s(AADT_dens)	3.0	10.225	1.023×10^{-6} ***
s(AADTT_dens)	3.0	19.526	1.313×10^{-12} ***
s(PAaM_dens)	3.0	4.207	0.005559 **
s(pop_dens)	3.0	5.362	0.001097 **
s(emp_dens)	3.3	19.303	2.160×10^{-13} ***
s(mfg_emp_dens)	7.0	7.007	2.441×10^{-8} ***
s(rt_emp_dens)	6.0	10.554	1.179×10^{-11} ***
s(tw_emp_dens)	5.9	96.571	$< 2.2 \times 10^{-16}$ ***
s(extent_injury_dens)	3.0	267.831	$< 2.2 \times 10^{-16}$ ***

Signif. codes: 0 ‘***’ 0.001 ‘**’

Thus, there is a significant nonlinear relationship between the predictor variable and the response variable at this time, the predictive variables after data processing have a nonlinear relationship with the response variable crash_dens of traffic accident density.

After the whole dataset was divided into test set and training set, the generalized additive model was established based on the data of the training set. The predictive variables appeared in the generalized additive model in the form of smooth splines with 3 degrees of freedom, and the partial dependence plots of the nine predictive variables was drawn. (Partial dependence plots can be used to show how a feature affects model predictions)

By analyzing the partial dependence plot, this paper found that the regression curve of population density and injury degree had a good fitting degree. By analyzing the Fig. 3-4, this paper found that the regression curve of population density and injury degree had a good fitting degree. From the partial dependence plot of injury degree, a greater density of injury degree often means a greater density of traffic accidents, population density also showed the same trend as the extent of injury. However, other variables showed varying degrees of overfitting. To avoid this situation, the complexity of the model should be reduced, or the influence of the predictor variable on the response variable should be inferred by observing the variation trend of the variable in a certain interval.

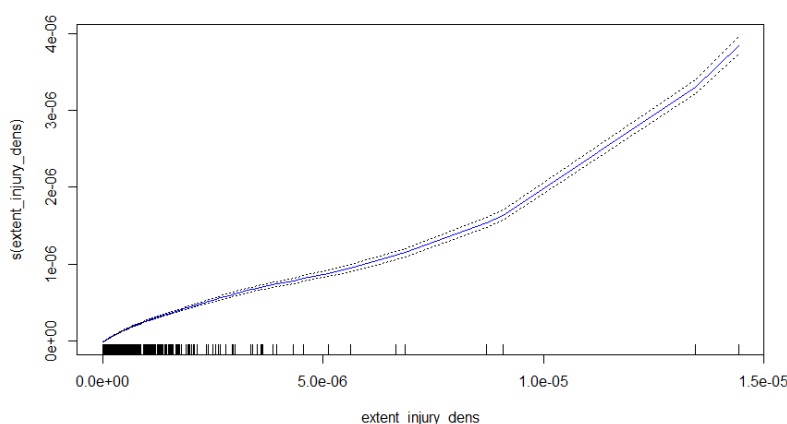


Fig 3. Partial dependence plot of extent_injury_dens

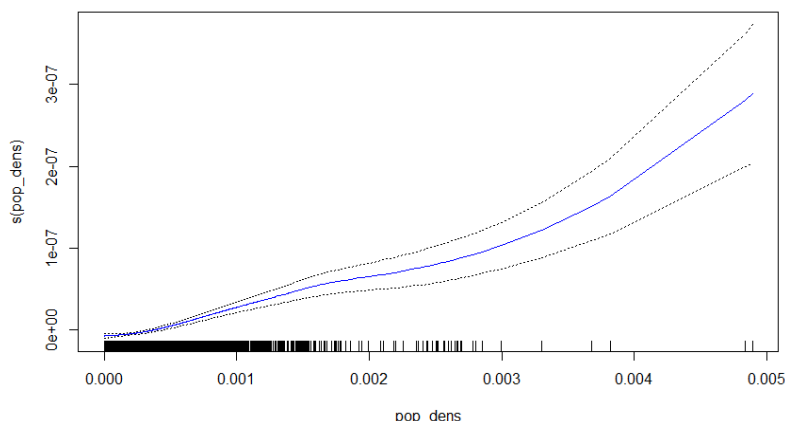


Fig 4. Partial dependence plot of pop_dens

However, other variables showed varying degrees of overfitting. To avoid this situation, the complexity of the model should be reduced, or the influence of the predictor variable on the response variable should be inferred by observing the variation trend of the variable in a certain interval. For instance, the regression curves of the annual average daily truck traffic density within the interval of 0-0.03 are well fitted, which is shown in Fig. 6. With the increase of the annual average daily truck traffic density, the traffic accident density also increases. However, when the annual average daily truck traffic density exceeds 0.03, serious overfitting occurs, which is caused by too sparse data. By observing the regression curve of the density of principal arterial roads, as is shown in Fig. 5, it can be found that when the density of arterial roads is large, the data is sparse and overfitting occurs. The confidence of the regression curve is large, which cannot be used to speculate the influence of the density of principal arterial roads on the density of traffic accidents.

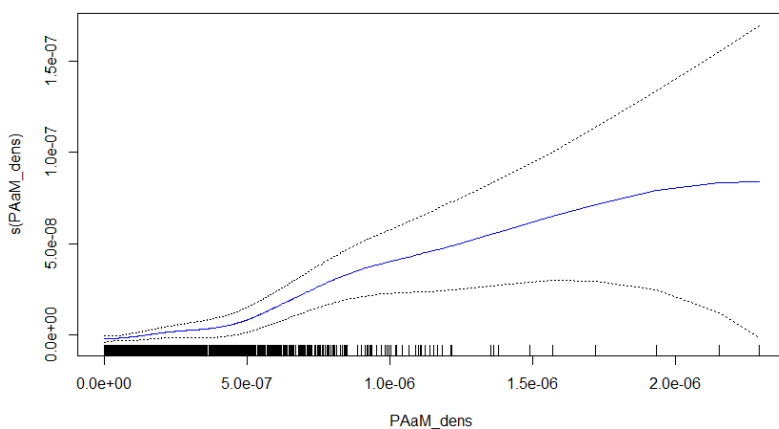


Fig 5. Partial dependence plot of PAaM_dens

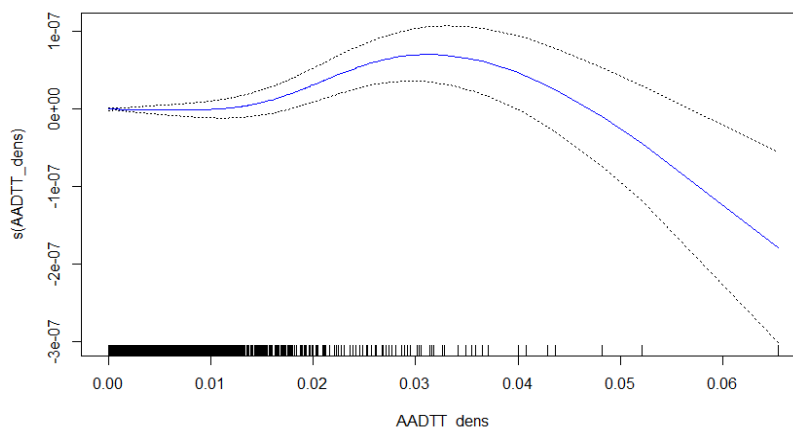


Fig 6. Partial dependence plot of AADTT_dens

Then the generalized additive model Fit2 is applied to the test set to see the prediction accuracy of the model and compare with the simple linear model Fit3. After calculation, the RMSE of the generalized additive model prediction test set is 8.201×10^{-8} , and the RMSE of the simple linear model is 8.290×10^{-8} . Obviously, the prediction accuracy of GAM is higher than that of the simple linear model.

However, compared with the random forest, the prediction accuracy of GAM is slightly insufficient. The comparison results are in the Table 5:

Table 5. Comparing metrics of Random Forest and GAM

Model	R ²	RMSE
Random Forest	0.604	6.629×10^{-8}
GAM	0.486	8.201×10^{-8}

5. Discussion

Comparing the results of models, random forest regression model, whose R2 is 0.604 and RMSE is 6.629×10^{-8} , produced better prediction than the generalized additive model with R2 equals to 0.486 and RMSE equals to 8.201×10^{-8} . It is manifest that the accuracy of random forest regression model is better than the latter one. It suggests that the good performance of random forest regression model in predicting crash density could be generalized to more homogenous datasets researching on crash density.

However, the weaker performance of the GAM model doesn't mean that it is useless in predicting the crash density. Based on the training set which is part of the whole dataset, and due to the inherent limitation, the GAM model didn't address the interaction among explanatory variables, which results a poorer prediction accuracy than machine learning model, but it proves each variable's influence on the final result. According to the study, the results of GAM model show that among all the factors that we have evaluated, the density of main roads, the extent injury density and population density are positively related to crash density, which is consistent with common knowledge. To reduce unnecessary costs and injuries caused by traffic accidents, policy makers should take actions on both the main road density and population density. But it is impossible for the government to reduce the number of main roads, which is incompatible with the idea of strengthening the infrastructure. The government should apply other alternative actions, such as enhancing traffic signal control on main roads, setting speed limits for vehicle and strengthening public's awareness of traffic safety to reduce the potential probability of traffic accidents. As for census tracts with large population density and severe extent injury density, the methods mentioned above equally applicable.

According to the partial dependence plot of all the variables, we found that some variables like employment density have been over-fitted in the study. The reason for the phenomena is the numerous noise data in the primitive dataset interference the results of both random forest regression model and the GAM model. For example, there are almost 70 percent of the population is 0 in the primitive dataset, which due to statistics on uninhabited areas. Also, all independent and dependent variables are non-linearly related to each other, which can also cause overfitting.

6. Conclusion

The study analyzed crash density in account of various factors especially injury severity, with the help of random forest regression model and the generalized additive model. Based on the processed data of 7998 census tracts in California from 2011 to 2020, the predicting accuracy of both models are calculated and compared. The study conducted random forest (RF) regression modeling analysis to present the predicting accuracy of RF is better than the generalized additive model (GAM), and the application of GAM is to emphasize the role of econometric models in explaining realistic situations. In order to optimize the final results of both models, the study used different methods to refine the models. In order to improve the accuracy of RF, the study applied Grid Search with Cross

Validation and Randomized Search with Cross-Validation for hyperparameter tuning. As for GAM, the study observed the variation trend of variables in a certain interval to avoid overfitting. To facilitate a policy analysis at the macro-level, the results of both models are computed. In addition, it is proved that extent injury density is an important criterion to crash density. In terms of planning perspective, it is important for policy makers to enhancing traffic signal control on main roads, setting speed limits for vehicle and strengthening public's awareness of traffic safety.

In future studies, it's important to avoid overfitting the dataset by further filtering of primitive dataset and setting penalty function or reduce the complexity of models. Also, it's manifest that it's valid to combine machine learning models with econometrics models to analyze crash density, which could be generalized to more homogenous datasets researching on crash density. Authors recommend that future studies may focus on these issues.

References

- [1] World Health Organization (WHO). Global Status Report on Road Safety 2018. December 2018. [cited 2020 October 28]. Available online: <https://www.who.int/publications/i/item/9789241565684>
- [2] Chen, Simiao, et al. "The global macroeconomic burden of road injuries: estimates and projections for 166 countries." *The Lancet Planetary Health* 3.9 (2019): e390-e398.
- [3] Merlin, L. A., Guerra, E., & Dumbaugh, E. (2020). Crash risk, crash exposure, and the built environment: A conceptual review. *Accident Analysis & Prevention*, 134, 105244.
- [4] Lee, M., & Khattak, A. J. (2019). Case study of crash severity spatial pattern identification in hot spot analysis. *Transportation research record*, 2673(9), 684-695.
- [5] Lee, J., Yasmin, S., Eluru, N., Abdel-Aty, M., & Cai, Q. (2018). Analysis of crash proportion by vehicle type at traffic analysis zone level: A mixed fractional split multinomial logit modeling approach with spatial effects. *Accident Analysis & Prevention*, 111, 12-22.
- [6] Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. *Accid. Anal. Prev.* 78, 146–154.
- [7] Yuan, Quan, and Jueyu Wang. "Goods movement, road safety, and spatial inequity: Evaluating freight-related crashes in low-income or minority neighborhoods." *Journal of Transport Geography* 96 (2021): 103186.
- [8] 1. Tang, J., Liang, J., Han, C., Li, Z., & Huang, H. (2019). Crash injury severity analysis using a two-layer Stacking framework. *Accident Analysis & Prevention*, 122, 226-238.
- [9] Jeong, H., Jang, Y., Bowman, P. J., & Masoud, N. (2018). Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. *Accident Analysis & Prevention*, 120, 250-261.
- [10] Zheng, Z., Lu, P., & Lantz, B. (2018). Commercial truck crash injury severity analysis using gradient boosting data mining model. *Journal of safety research*, 65, 115-124.
- [11] Biau, Gérard, and Erwan Scornet. "A random forest guided tour." *Test* 25.2 (2016): 197-227.
- [12] Altman, Naomi, and Martin Krzywinski. "Ensemble methods: bagging and random forests." *Nature Methods* 14.10 (2017): 933-935.
- [13] Baayen, R. Harald, and Maja Linke. "An introduction to the generalized additive model." *A practical handbook of corpus linguistics*. New York: Springer (2020): 563-591.
- [14] Feurer, Matthias, and Frank Hutter. "Hyperparameter optimization." *Automated machine learning*. Springer, Cham, 2019. 3-33