SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Analytical Models for Traffic Congestion and Accident Analysis

Hongrui Liu
Rahul Ramachandra Shetty

CSUTC California State University Transportation Consortium

CSU TRANSPORTATION CONSORTIUM transweb.sjsu.edu/csutc

Mineta Transportation Institute

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) funded by the U.S. Department of Transportation and the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 21-27

Analytical Models for Traffic Congestion and Accident Analysis

Hongrui Liu, PhD
Rahul Ramachandra Shetty

November 2021

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 2102	**2. Government Accession No.**	**3. Recipient's Catalog No.**	
4. Title and Subtitle Analytical Models for Traffic Congestion and Accident Analysis		**5. Report Date** November 2021 **6. Performing Organization Code**	
7. Authors Hongrui Liu: 0000-0002-2540-2765 Rahul Ramachandra Shetty: 0000-0002-6664-8112		**8. Performing Organization Report** CA-MTI-2102	
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219		**10. Work Unit No.** **11. Contract or Grant No.** 69A3551747127	
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology University Transportation Centers Program 1200 New Jersey Avenue, SE Washington, DC 20590		**13. Type of Report and Period Covered** **14. Sponsoring Agency Code**	
15. Supplemental Notes			
16. Abstract In the US, over 38,000 people die in road crashes each year, and 2.35 million are injured or disabled, according to the statistics report from the Association for Safe International Road Travel (ASIRT) in 2020. In addition, traffic congestion keeping Americans stuck on the road wastes millions of hours and billions of dollars each year. Using statistical techniques and machine learning algorithms, this research developed accurate predictive models for traffic congestion and road accidents to increase understanding of the complex causes of these challenging issues. The research used US Accidents data consisting of 49 variables describing 4.2 million accident records from February 2016 to December 2020, as well as logistic regression, tree-based techniques such as Decision Tree Classifier and Random Forest Classifier (RF), and Extreme Gradient boosting (XG-boost) to process and train the models. These models will assist people in making smart real-time transportation decisions to improve mobility and reduce accidents.			
17. Key Words Transportation, Traffic Congestion, Data Analysis, Predictive Models, Machine Learning	**18. Distribution Statement** No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 27	**22. Price**

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2021.2102

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565
Email: mineta-institute@sjsu.edu

transweb.sjsu.edu/research/2102

ACKNOWLEDGMENTS

The work was sponsored in part by the Mineta Transportation Institute. This research was presented at the 2021 INFORMS Conference on Service Science.[1]

[1] Hongrui, L. & Shetty, R. (2021). "Accidents Analysis and Severity Prediction using Machine Learning Algorithms. Proceedings of the 2021 INFORMS Conference on Service Science, August 10-12, 2021. View the presentation at: https://www.icss2021.servicescienceglobal.org/wp-content/uploads/Shetty-1570724346.mp4

CONTENTS

Acknowledgments iv
List of Figure vi
List of Tables vii
1. Introduction 1
2. Data Source 3
    2.1. Exploratory Data Analysis 3
    2.2. Data Preprocessing 8
3. Methodology 9
4. Results & Future Work 10
Bibliography 15
About the Authors 16

List of Figures

Figure 1. (a) Frequency of Accident Record with Respect to Severity and (b) Locations and Severity of Accidents Across CA 3
Figure 2. Number of Accidents due to (a) all Weather Conditions and (b) Amount of Precipitation 4
Figure 3. Number of Accidents Due to (a) Direction and (b) Visibility 5
Figure 4. Number of Accidents Due to (a) Pressure and (b) Humidity 5
Figure 5. Number of the Accidents by (a) Day of the Week, (b) Month in the Year, and (c) Time of the Day 6
Figure 6. Number of Accidents with POI Attributes 7
Figure 7. Number of Accidents by Major City in California 8

List of Tables

Table 1. Relationship Between Accident and Congestion 11
Table 2. Summary of Accuracy Scores with the Imbalanced and the Balanced Data Sets 11
Table 3. Model Results Summary from Original Imbalanced Data 12
Table 4. Model Results Summary with Balanced Data (SMOTE) 13
Table 5. Selected Features in the Predictive Model 14
Table 6. Summary of Accuracy Score with 14 Features 14

1. Introduction

Traffic congestion and road accidents are major public challenges and affect peoples' daily lives in different ways. The situation is getting worse in recent years due to population growth and the boom in ecommerce in many urban areas (Sardjono et al., 2019). According to the 2019 annual study from INRIX, a company that provides location-based data and analytics on traffic and parking, the average time Americans lost to traffic has increased by two hours from 2017 to 2019. Road accidents can have an even more significant impact on a persons' life depending on the severity of the event. In the US, over 38,000 people die in road crashes each year, and 2.35 million are injured or disabled, according to the statistics report from the Association for Safe International Road Travel (ASIRT) in 2020. It is important to understand the root cause of traffic congestion and road accidents and their association so that effective policies and transportation decision tools can be implemented to improve roadway safety and relieve traffic congestion.

The analysis of traffic congestion and road accidents are very complex, as they are affected by many factors, such as road characteristics, time of the day/week, weather conditions, and events that may change in real-time (Yuan et al., 2018; Lu et al., 2017). The development of advanced technologies such as GPS (Global Positioning System) and IoT (Internet of Things) offers great visibility and transparency of roadway conditions nowadays. This information, if utilized appropriately, can help people understand the factors impacting traffic congestion and make timely transportation decisions to improve the mobility of people and goods. Since 2015, the Department of Transportation (DOT) launched a series of Smart City Challenges, asking for ideas to create an integrated, first-of-its-kind smart transportation system that would use data, applications, and technology to revolutionize the transportation systems to help improve people's lives.

In response to the aforementioned needs, some research has been conducted in the field of road accident and traffic congestion analysis using data analytics models. For example, a data mining technique used k-means algorithms to identify locations with high number of accidents and the affecting factors (Kumar and Toshniwal, 2016). The k-means algorithm was used to cluster the location into high, medium, and low-frequency locations and road attributes features were analyzed to find causes for the accidents. Dogru and Subasi (2018) developed a detection system that uses Artificial Neural Network (ANN), Support Vector Machine (SVM), and RF machine learning algorithms to understand driver behaviors and detect accidents on freeways. The performance of RF algorithm, in terms of its accuracy, was found superior to ANN and SVM algorithms in that application. Machine learning algorithms Adaboost, RF, and naïve Bayes were used to predict the severity of injuries due to accidents (AlMamlook et al., 2019). The findings of this study indicate that the RF model can be a promising tool for predicting the injury severity of traffic accidents. The K-means clustering algorithm, together with association rule mining, has been used to find various patterns in road accidents (Nandurge and Dharwadkar, 2017). Artificial

neural network models, utilizing Feedforward Multilayer Perceptron (FFMLP), was used to predict road accidents for time series forecasting by Kouziokas (2018). Several parameters such as the number of the neurons in the hidden layers and the nature of the transfer functions, were taken into consideration to build model and the optimal prediction model was tested in the study area and the results have shown a very good prediction accuracy. Another study used a LSTM-GBRT (long short-term memory, gradient boosted regression trees) model to predict the safety level of traffic accidents (Zhang et al., 2020). Compared with various regression models, the experimental results show that the LSTM-GBRT model has a better fitting effect and robustness.

We propose to use statistical techniques and machine learning algorithms to process and train the large amount of data offered by advanced information technologies to obtain robust predictive models of traffic congestion and road accidents. The machine learning algorithms used in the study include logistics regression, decision tree classifier, random forest classifier, and XG-Boost. The predicted target is the severity of the accidents, which is a measure of the impact of road accidents on traffic congestion. The study also includes analyses to identify the factors that have a significant impact on road accidents. We used data from Kaggle on US accidents with information on weather, location, period, and point of interest (POI) attributes for our analysis. The objective is to build a robust model using machine learning algorithms to predict the severity of road accidents. This information can be used to make important decisions to minimize time spent on the road. This study contributes the existing literature by analyzing a large amount of data (4.2 million records) with high dimensional features (49 features) and comparing four machine learning algorithms to obtain an accurate predictive model. In addition, we studied the trade-off between the computation effort and model accuracy to provide an insight on implementing a practical model in real life.

This paper is organized as follows. In Section II, we present the data source and perform data exploratory analysis and pre-processing. Next, we introduce the machine learning methodologies used to train the predictive models in Section III. The numerical results of the case study, conclusions, and future work are discussed in Section IV.

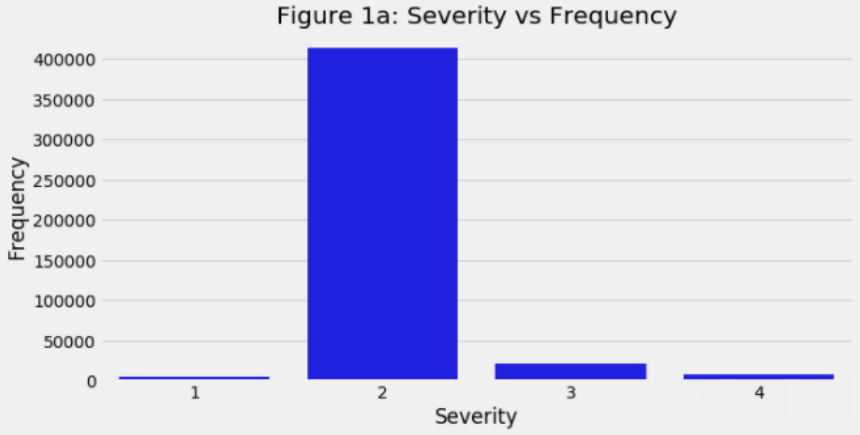
2. Data Source

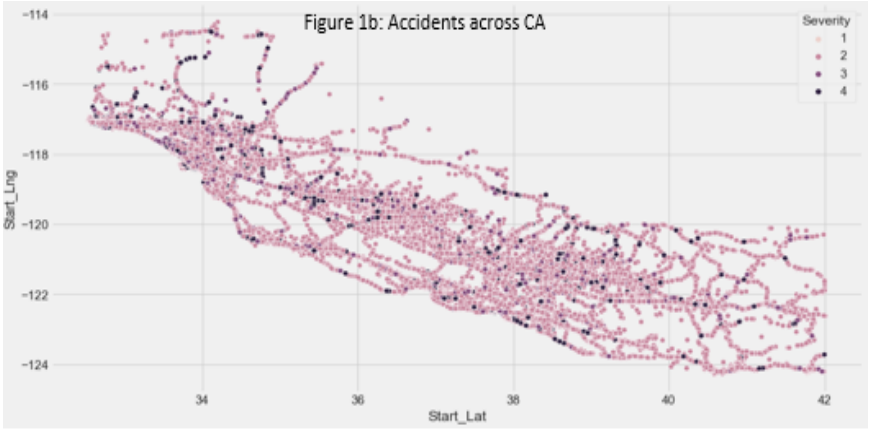
Our dataset consists of 4.2 million records of car accidents from February 2016 to December 2020 across 49 states of the United States, collected from Kaggle under the title US Accidents. The data set has 49 variables with 17 categorical variables, 13 Boolean variables, 16 numerical variables, 2 Date Time stamps, and 1 string. Out of the 49 variables, 12 of them are traffic attributes, 9 of them are location attributes, 11 of them are weather attributes, and 13 of them are point of interest (POI) attributes.

2.1 Exploratory Data Analysis

The data entries on Kaggle are from three sources, MapQuest, Bing and other. We selected the data from Bing for our analysis as the target variable "severity" is better classified with less conflicts from this source. The accident severity from Bing has four levels, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay) (Moosavi et al., 2019). The average impacted area on the road is 0.01 miles by severity 1 accidents, 0.27 miles by severity 2 accidents, 0.55 miles by severity 3 accidents, and 0.96 miles by severity 4 accidents. To limit the scope of the study, the data set was reduced to a state level and California was identified to be the state that has the highest number of accidents for further study. The number and severity of the accidents in California are shown in Figure 1a. There are 5534 accidents with severity of 1, 414069 accidents with severity of 2, 20592 accidents with severity of 3, and 8316 accidents with severity of 4. It is observed that the data is imbalanced with the highest number of severity level 2 accidents. Figure 1b shows how the accidents are distributed across the state of California.

Figure 1. (a) Frequency of Accident Record with Respect to Severity and
(b) Locations and Severity of Accidents Across CA





The correlation between different attributes and the target variable, severity, was then studied to have an initial understanding of the impact of various attributes on the target variable. We found that a higher number of accidents occurred when the weather conditions are fair (Figure 2a), with 0 precipitation (Figure 2b), when the wind is calm (Figure 3a), and with less wind speed, and higher visibility (Figure 3b). Other weather features such as pressure, humidity, and windchill did not show any clear impact on the accident severity (Figure 4a and Figure 4b).

Figure 2. Number of Accidents due to (a) all Weather Conditions and (b) Amount of Precipitation

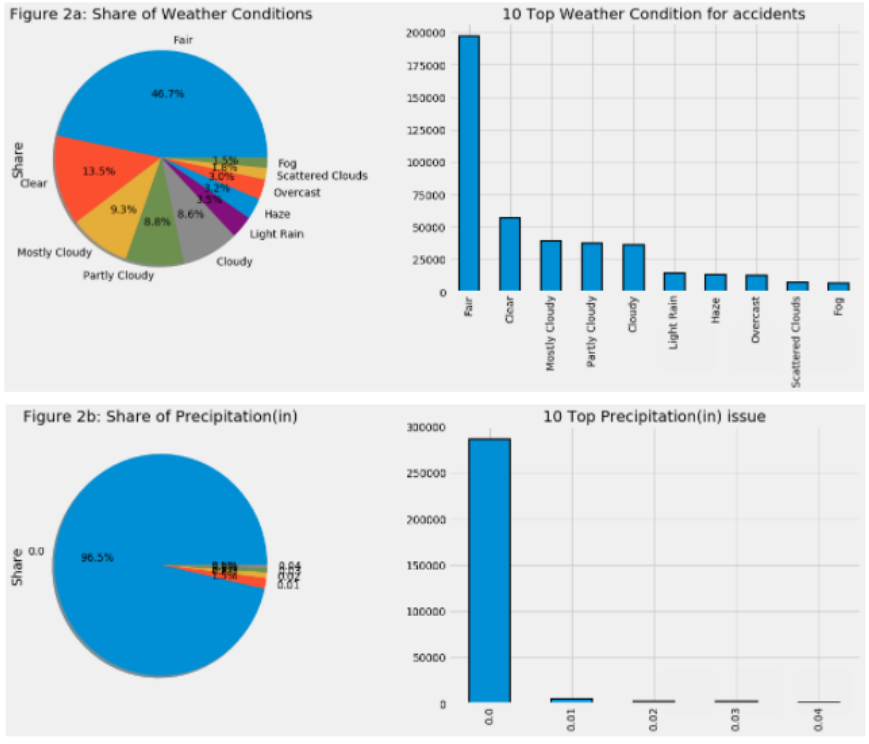


Figure 3. Number of Accidents Due to (a) Wind Direction and (b) Visibility

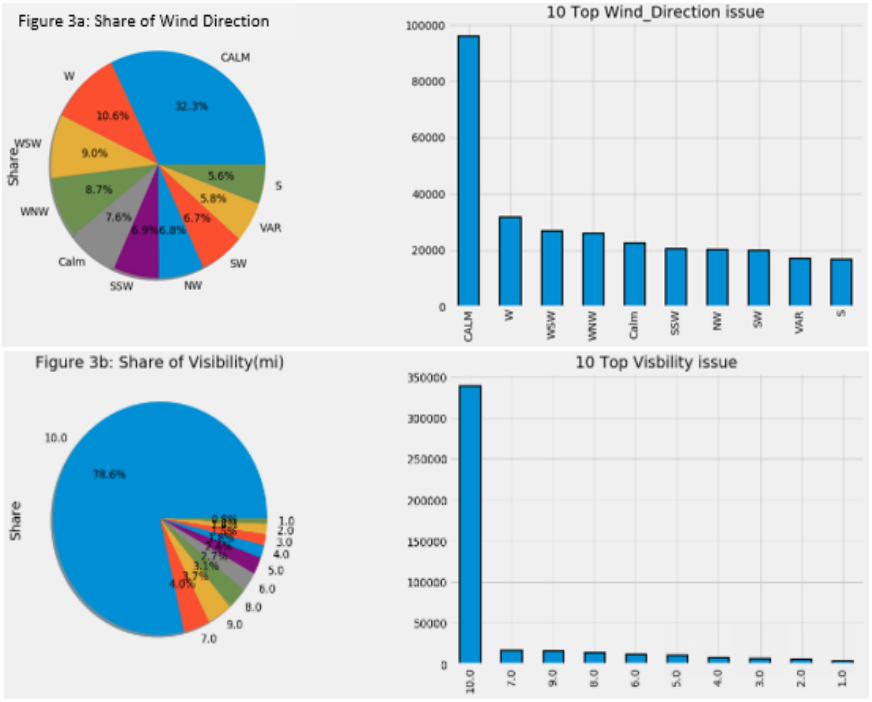
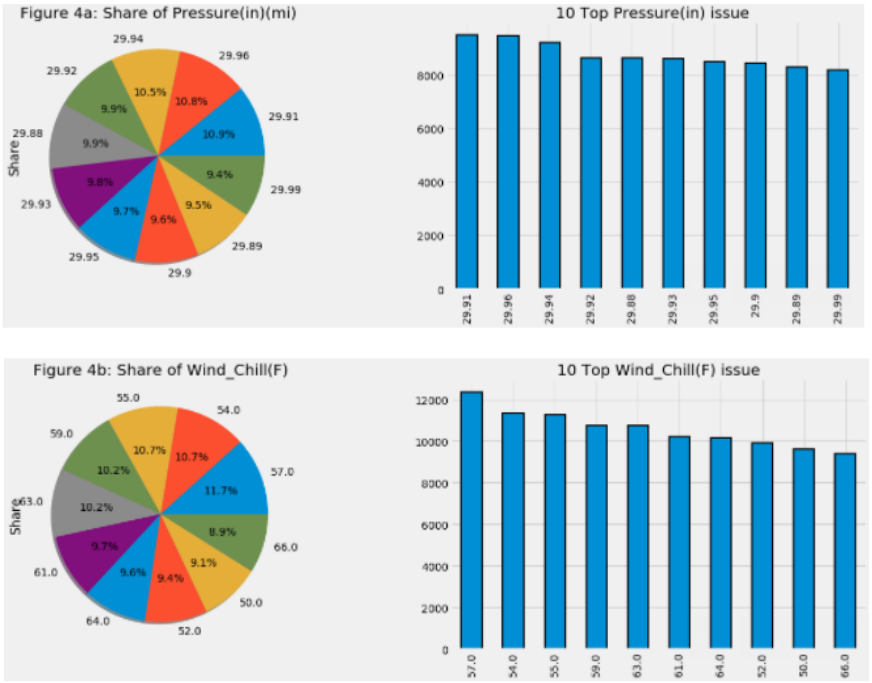
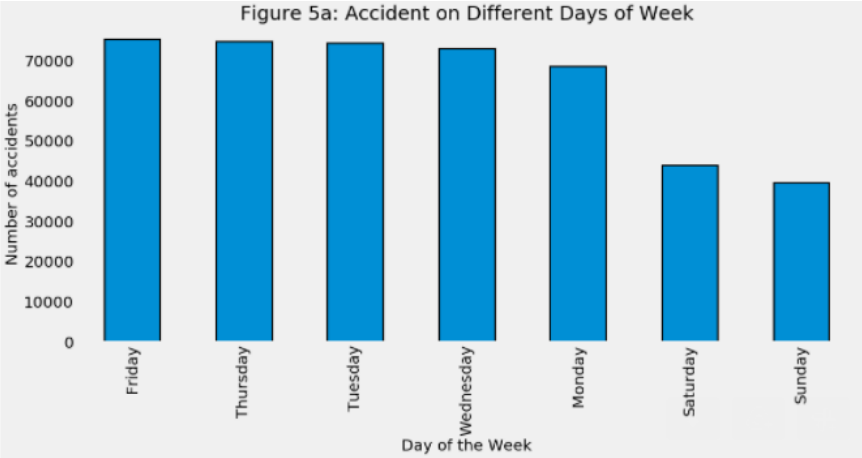


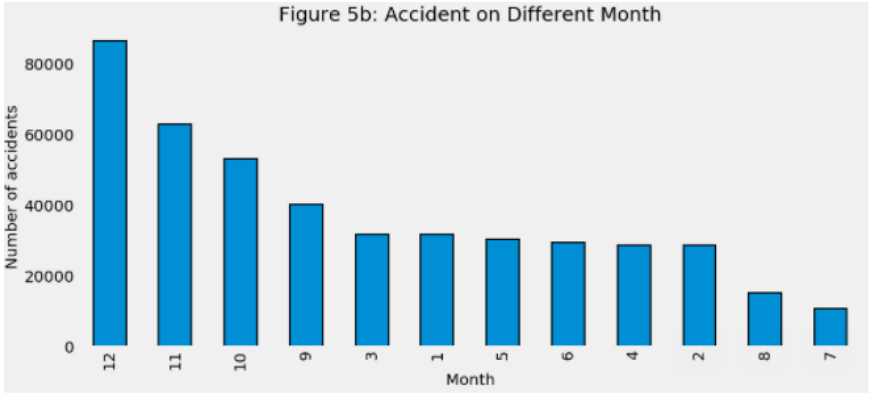
Figure 4. Number of Accidents Due to (a) Pressure and (b) Humidity

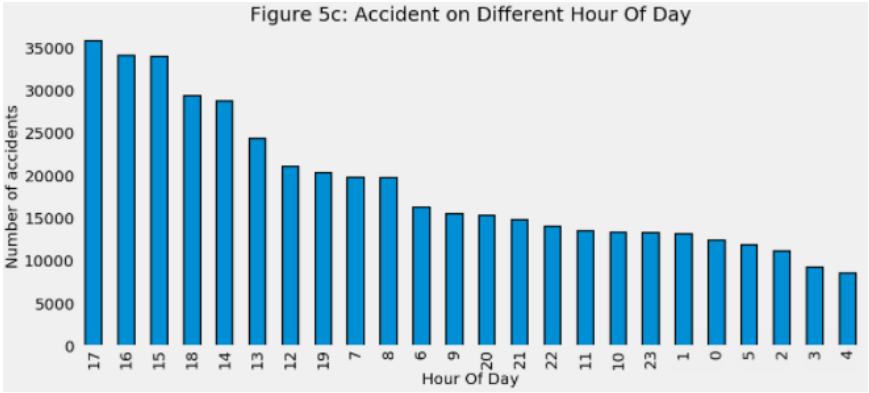


Feature extraction was then used to extract the hour, week, month, and year information from the period of day attribute. The results show that a large number of accidents took place on weekdays rather than on weekends (Figure 5a). Over a year, October, November, and December had the most accidents compared to the other months, as shown in Figure 5b. The times between 4 pm to 6 pm had a higher number of accidents during a day as shown in Figure 5c.

Figure 5. Number of the Accidents by (a) Day of the Week, (b) Month in the Year, and (c) Time of the Day

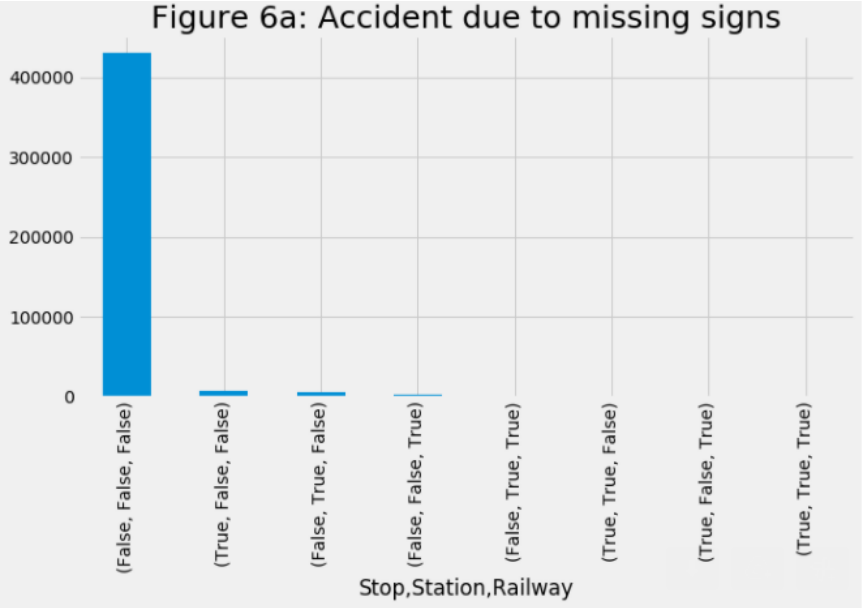


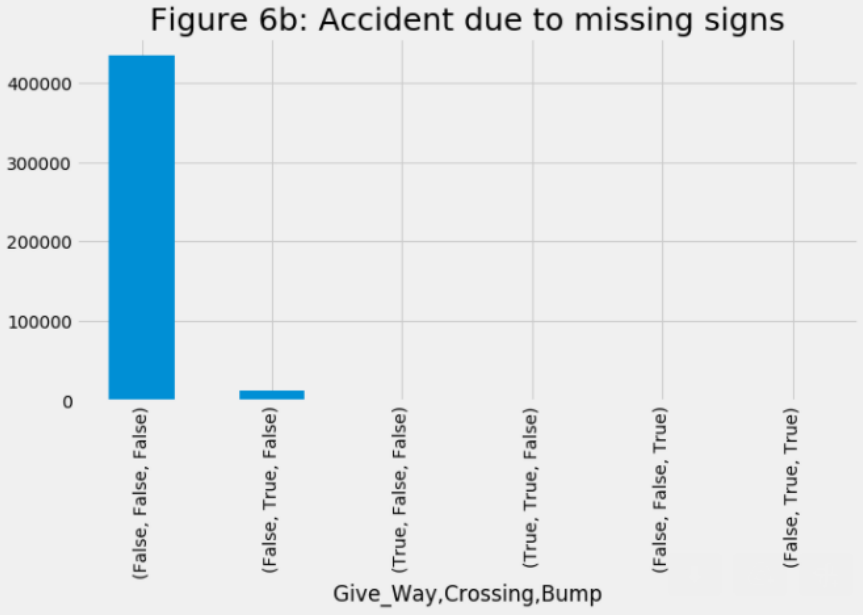




A multivariate analysis was conducted on the POI attributes by grouping stops, stations, and railways. There were a significant number of accidents when all the three attributes are missing, or there is a missing sign in giving way, crossing, and bump, as shown in Figure 6.

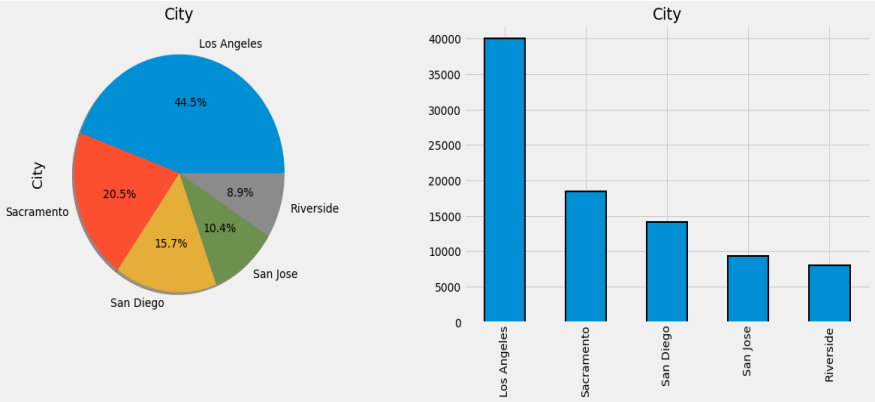
Figure 6. Number of Accidents with POI Attributes





The analysis of the location attributes shows that LA (Los Angeles) is the city that has the highest number of accidents in California, with 44.5% of overall accidents as shown in Figure 7.

Figure 7. Number of Accidents by Major City in California



2.2 Data Preprocessing

The data was cleaned and prepared for the different machine learning algorithms to train the predictive models. Feature engineering techniques are used to extract time features and to analyze any trends or patterns over time. Feature selection techniques are used to handle the high data dimensionality by identifying features of high importance. Features such as precipitation that contain too many null values were dropped. Features such as country, state, turning loop, and source, which have a single unique value each, and those such as 'ID', and 'description' which have no impact to the accident severity were removed from the data set. The chi-square test was used to evaluate categorical variables, Pearson correlation was used to remove highly correlated features, and single or constant value features are dropped using variance thresholds. As a result of data cleaning, 25 variables were retained in the analysis. The data was next split into x (the 24 independent features) and y (prediction target, severity). We use 70% of the data as the training set and 30% for testing model accuracy. Feature encoding is performed on train and test sets separately to avoid data leakage and overfitting. The prediction target, the accident severity levels were converted into categorical values and all the other features were converted to numerical values using pandas categorical code. Data imputation was used to fill any missing values with the median value. Synthetic minority oversampling technique (SMOTE) was used to balance the data set by duplicating the minority data from the minority data population.

3. Methodology

We used four machine learning algorithms to train the predicative models as described below.

- Multivariable logistic regression: a classification algorithm that takes less effort and time to train the model. It converts the loss function to cross-entropy, and probability prediction to multinomial probability distribution using maximum likelihood estimation to predict different severity classes. The target variable has 4 classes and hence Multinomial logistic regression is used over logistic regression.
- Decision tree: this is a supervised technique that uses different algorithms to split a node into multiple sub-nodes on all variables and selects the split that provides the best result. The tree splits at the topmost node and the outcome is represented by the branches, and the leaf consists of the class. The tree is stopped once it reaches the defined stopping criteria. The terminal node represents the mode which is the predictor for that region. The decision tree requires minimal data cleaning which is helpful in our data set as there are 25 multidimensional variables and over 448 thousand entries. A decision tree is also easy to interpret.
- Random Forest (RF): the RF algorithm eliminates high variance by adding randomness to the model. The algorithm works by creating multiple decision trees using a bootstrapping method by random sampling with replacement. A bagging technique is used for the output from each decision tree. The final output is the mode of the output of all the individual decision trees. To eliminate overfitting and to overcome bias RF is used.
- Extreme Gradient boosting (XG-Boost): the algorithm is a specific implementation of the Gradient Boosting method. Boosting builds models from individual so called "weak learners" in an iterative way. Decision tree is the the most common type of weak model used in boosting. A final model will be created based on a collection of individual models. The predictive power of these individual models is weak and prone to overfitting but combining many such weak models in an ensemble will lead to an overall much improved result. The gradient is used to minimize a loss function, which utilize gradient descent to optimize ("learn") weights. XG-Boost uses more accurate approximations, for example, second-order gradients and advanced regularization to find the best tree mode.

Python packages including Pandas, NumPy, matplotlib, seaborn, and date-time from python were used for data analysis and data visualization. Synthetic Minority Oversampling Technique (SMOTE) is used to randomly oversampling the data that utilizes the k-nearest neighbor algorithm to create synthetic data by increasing the number of records without adding any new information or variation to the model. The upscaled data is then fitted to the different machine learning algorithms to train models to predict accident severity.

4. Results & Future Work

The training data set that included 25 variables (24 features and 1 target) was used to train the four machine learning algorithms to obtain predictive models of accident severity. The 24 features are listed in Table 1. The test data set was then used to test the accuracy of the four predictive models. The four prediction models were first trained using the original imbalanced data set, and then the balanced data set by applying the SMOTE technique. The accuracy score, precision, recall, and F1 score of the four predictive models are summarized in Tables 2, 3 and 4. The four models trained from the original imbalanced data have similar accuracy. The accuracy of the linear regression model dropped significantly after applying the SMOTE technique, when the data set is expanded to have 268139 more entries to balance the data set. The Decision Tree, RF and XG-boost algorithms exhibited greater robustness with both the imbalanced and the balanced data, and they improved with the balanced data when predicting the non-dominant severity levels 3 and 4 as shown in Tables 3 and 4. Additional analysis was conducted to further reduce the features from 24 to 14, using the balanced data, to understand the trade-offs between computational expense and model accuracy. The remaining 14 features used for the analysis are shown in Table 5. Again, Decision Tree, RF and XG-Boost were demonstrated to have a greater robustness in terms of the accuracy score as shown in Table 6.

The RF model handles noise, multi-class variables, and high variances well, while the boosting technique learns from past errors to improve prediction accuracy. The tuning of hyperparameters may have important impact to the prediction results. Further studies will be conducted to explore the optimal settings for the parameters. In the study of the trade-off between the computational effort and model accuracy, we simply removed 10 of the 24 features used in the full data analysis. The accuracy score didn't change significantly for all the models. Further studies will be conducted to investigate the impact of the features in detail, so we can establish a removal sequence depending on the required model accuracy. With the availability of the large data set, we will also explore the use of deep learning algorithms in the future.

The obtained predictive models together with the real time visibility of the environmental conditions offered by advanced information technologies can help build a smart transportation system to guide people to make decisions to minimize their time spent on the road.

Table 1. Relationship Between Accident and Congestion

Start_Lat	Temperature(F)	Wind_Direction	Amenity
Junction	Civil_Twilight	Distance(mi)	Pressure(in)
Wind_Speed(mph)	Crossing	Traffic_Signal	Nautical_Twilight
Side'	Visibility(mi)	Precipitation(in)	Give_Way
Sunrise_Sunset	Astronomical_Twilight	Station	Stop
Traffic_Calming	Duration	Railway	No_Exit

Table 2. Summary of Accuracy Scores with the Imbalanced and the Balanced Data Sets

	Multivariable Linear Regression	Decision Tree	Random Forest	XG- Boost
Original Imbalanced Data	95.30	95.03	96.50	96.38
Balanced Data—SMOTE	52.24	93.86	95.22	93.89

Table 3. Model Results Summary from Original Imbalanced Data

Multivariable Logistic regression	Precision	Recall	F1 Score	Support
Severity 1	0.60	0.27	0.38	1546
Severity 2	0.96	1	0.98	85184
Severity 3	0.23	0.01	0.01	1914
Severity 4	0.35	0.01	0.01	984
Decision Tree	Precision	Recall	F1 Score	Support
Severity 1	0.69	0.71	0.70	1546
Severity 2	0.98	0.98	0.98	85184
Severity 3	0.29	0.30	0.29	1914
Severity 4	0.30	0.36	0.33	984
Random Forest	Precision	Recall	F1 Score	Support
Severity 1	0.73	0.74	0.73	1546
Severity 2	0.97	0.99	0.98	85184
Severity 3	0.47	0.16	0.24	1914
Severity 4	0.60	0.23	0.34	984
XG-Boost	Precision	Recall	F1 Score	Support
Severity 1	0.71	0.79	0.75	1546
Severity 2	0.91	0.99	0.98	85184
Severity 3	0.47	0.10	0.17	1914
Severity 4	0.64	0.22	0.32	984

Table 4. Model Results Summary with Balanced Data (SMOTE)

Multivariable Logistic regression	Precision	Recall	F1 Score	Support
Severity 1	0.16	0.92	0.27	1546
Severity 2	0.98	0.52	0.68	85184
Severity 3	0.04	0.47	0.08	1914
Severity 4	0.02	0.30	0.04	984
Decision Tree	Precision	Recall	F1 Score	Support
Severity 1	0.64	0.71	0.68	1546
Severity 2	0.98	0.96	0.97	85184
Severity 3	0.24	0.37	0.29	1914
Severity 4	0.22	0.36	0.27	984
Random Forest	Precision	Recall	F1 Score	Support
Severity 1	0.66	0.80	0.72	1546
Severity 2	0.98	0.98	0.98	85184
Severity 3	0.30	0.37	0.33	1914
Severity 4	0.43	0.34	0.38	984
XG-Boost	Precision	Recall	F1 Score	Support
Severity 1	0.72	0.80	0.76	1546
Severity 2	0.97	0.99	0.98	85184
Severity 3	0.49	0.10	0.17	1914
Severity 4	0.61	0.23	0.33	984

Table 5. Selected Features in the Predictive Model

Duration	Start_Lat	Distance(mi)	Temperature(F)
Wind_Speed(mph)	Sunrise_Sunset	Civil_Twilight	Pressure(in)
Junction	Astronomical_Twilight	Side	Visibility(mi)
Traffic_Signal'	Station		

Table 6. Summary of Accuracy Score with 14 Features

Model	Multivariable Linear Regression	Decision Tree	Random Forest	XG- Boost
Accuracy Score	52.73	93.81	95.19	94.66

Bibliography

AlMamlook, R. E., K. M. Kwayu, M. R. Alkasisbeh, and A. A. Frefer. "Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity." *IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT).* (2019): 272-276. doi: 10.1109/JEEIT.2019.8717393.

Dogru, N. and A. Subasi. "Traffic Accident Detection Using Random Forest Classifier." *15th Learning and Technology Conference (L&T)*. (2018): 40-45. doi: 10.1109/LT.2018.8368509.

Kouziokas, G. N. "Neural Network-Based Road Accident Forecasting." In *Transportation and Public Management. Data Analytics: Paving the Way to Sustainable Urban Mobility*, 98–103. 2018. doi: 10.1007/978-3-030-02305-8_12.

Kumar, S. and D. Toshniwal. "A Data Mining Approach to Characterize Road Accident Locations." *Journal of Modern Transportation* 24, no. 1 (2016): 62–72. doi: 10.1007/s40534-016-0095-5.

Lu, W., D. Luo, and M. Yan. "A Model of Traffic Accident Prediction Based on Convolutional Neural Network." *IEEE International Conference on Intelligent Transportation Engineering (ICITE)*. (2017): 198-202.

Moosavi, S. M., H. Samavatian, S. Parthasarathy, R. Teodorescu, and R. Ramnath. "Accident Risk Prediction based on Heterogeneous Sparse Data." *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.* (2019). doi:10.1145/3347146.3359078.

Nandurge, P. A. and N. V. Dharwadkar. "Analyzing Road Accident Data Using Machine Learning Paradigms." *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. (2017): 604-610. doi: 10.1109/I-SMAC.2017.8058251.

Sardjono1, W., E. Selviyanti, and W. G. Perdana. "Modeling the Relationship Between Public Transportation and Traffic Conditions in Urban Areas: A System Dynamics Approach." J*ournal of Physics: Conference Series* Ser. 1465, (2019): 012023/

Yuan, Z., X. Zhou, and T. Yang, (2018). "Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data." In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining,* 984–992. New York, NY, USA: ACM, 2018.

Zhang, Z., W. Yang, and S. Wushour. (2020). "Traffic Accident Prediction Based on LSTM-GBRT Model." *Journal of Control Science and Engineering* 2020, (2020): 1–10. doi: 10.1155/2020/4206919.

About the Authors

Hongrui Liu

Hongrui Liu is an assistant professor in Industrial and Systems Engineering at San Jose State University (SJSU). Her research interests include optimization modeling, algorithms, data analytics, and their applications in the supply chain and energy industry.

Rahul Ramachandra Shetty

Rahul Ramachandra Shetty received his MS degree in Industrial and Systems Engineering at San Jose State University in 2021. He is currently a Supply Chain Business Manager at Lam Research Corporation.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors