

Machine Learning Models for Car Accidents

Spring 2023 IST 718

Group 1 Harim Shin, Kishan Rathor, Rich Granato

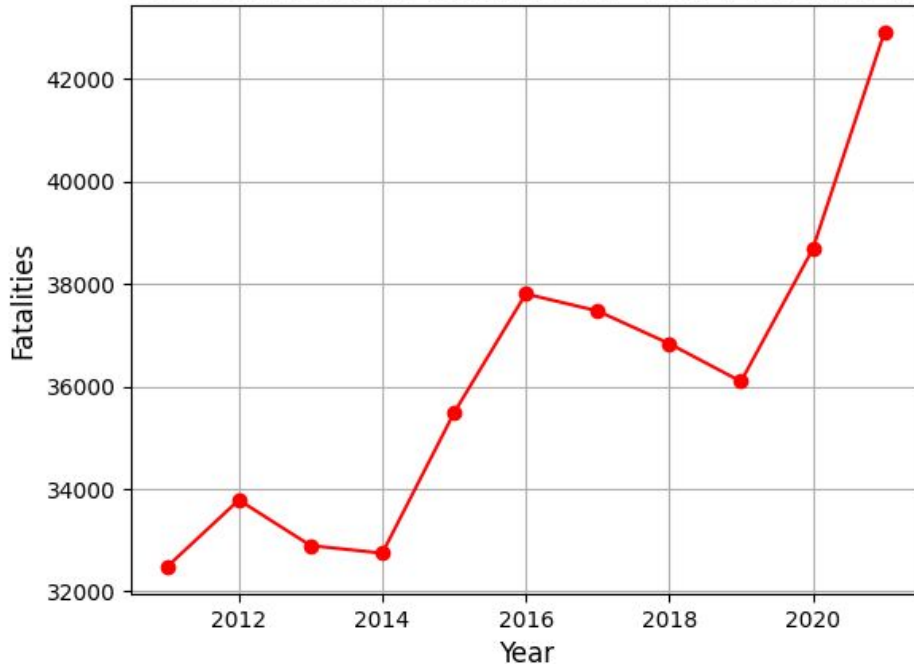


Contents

- Project Introduction
- Data Introduction
- Data Exploration Insights
- Models
 - Unsupervised Learning
 - Supervised Learning
 - Random Forest & Gradient Boosted Trees
- Conclusion

Project Introduction - Big Picture

U.S. Car Accident Fatalities by Year (2011-2021)



- 42,915 people died in car accidents in 2021
- Car accident fatalities have increased by 32.13% from 2011 to 2021
- What can we do to decrease the number of yearly U.S. car accidents?

Project Introduction - Objectives

- **Objective 1:** Predict severity of car accidents to help protect people after a car accident has taken place.
- This information would be utilized by:
 - Fire Departments
 - Police
 - News Media Outlets

1

Project Introduction - Objectives

- **Objective 2:** Identify common causes of car accidents to help prevent car accidents before they happen.
- This information would be utilized in:
 - Academia
 - Automobile Industry

2

An isometric illustration of a city street scene. It features various elements like buildings, trees, pedestrians, cars, a truck, and a bicycle. The scene is viewed from an elevated perspective, showing the layout of the street and surrounding urban environment. The colors are muted, with greys for buildings and roads, greens for trees, and various colors for the people and vehicles.

Data Introduction

- Period: Feb 2016 ~ Dec 2021
- Location: 49 states of USA (except Alaska and Hawaii)
- Number of datapoints: 2.8M

Data Introduction

Features

Time/Location



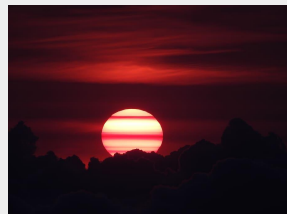
Weather



Traffic



Period



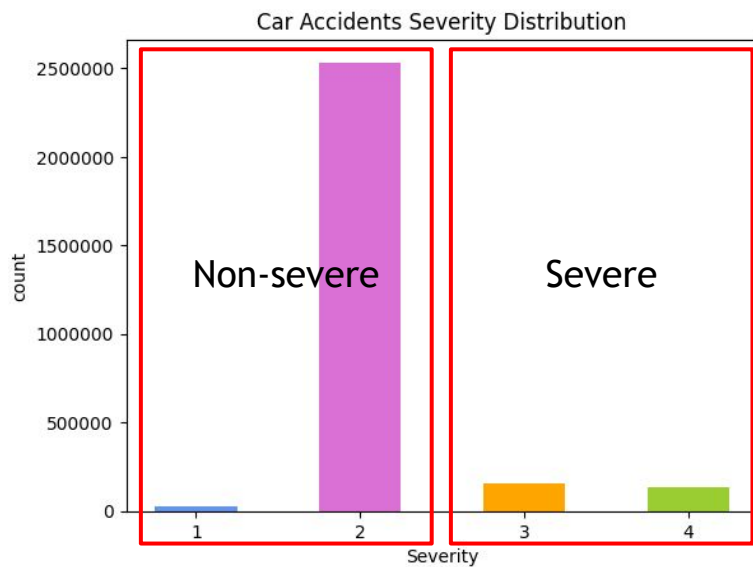
Predict

Target: Severity of car accidents

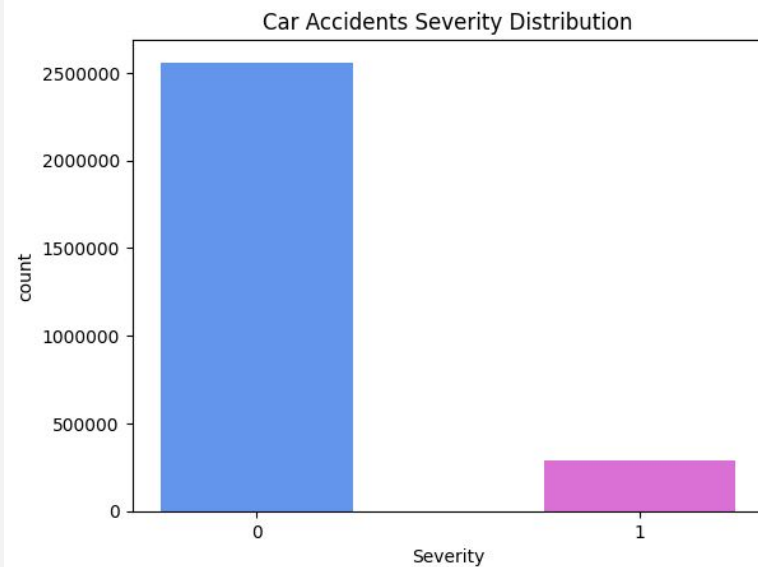


Data Introduction

Original Data

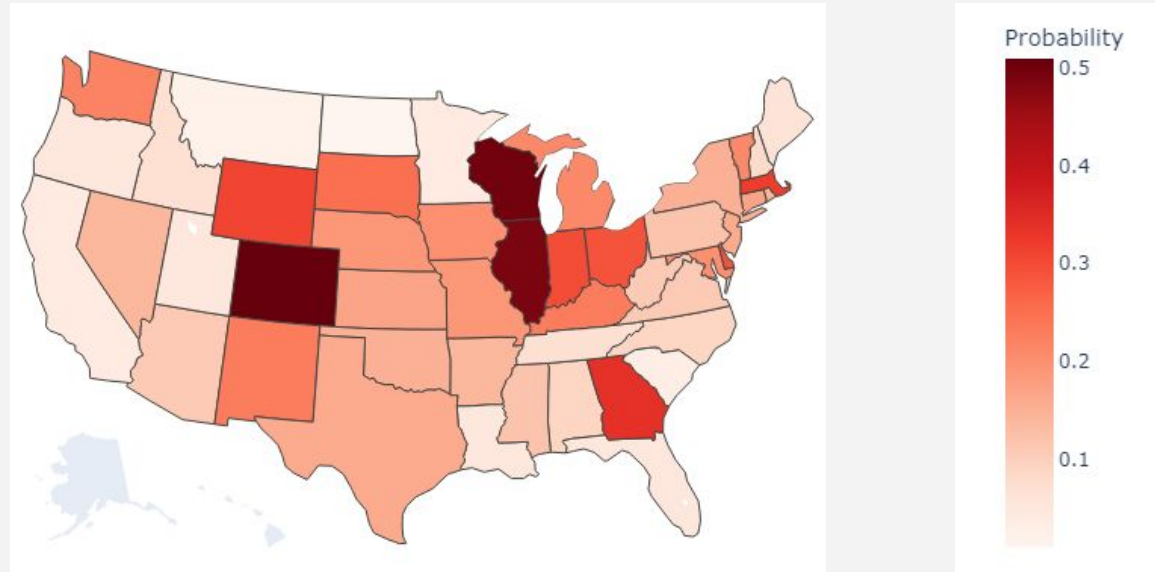


After Manipulation

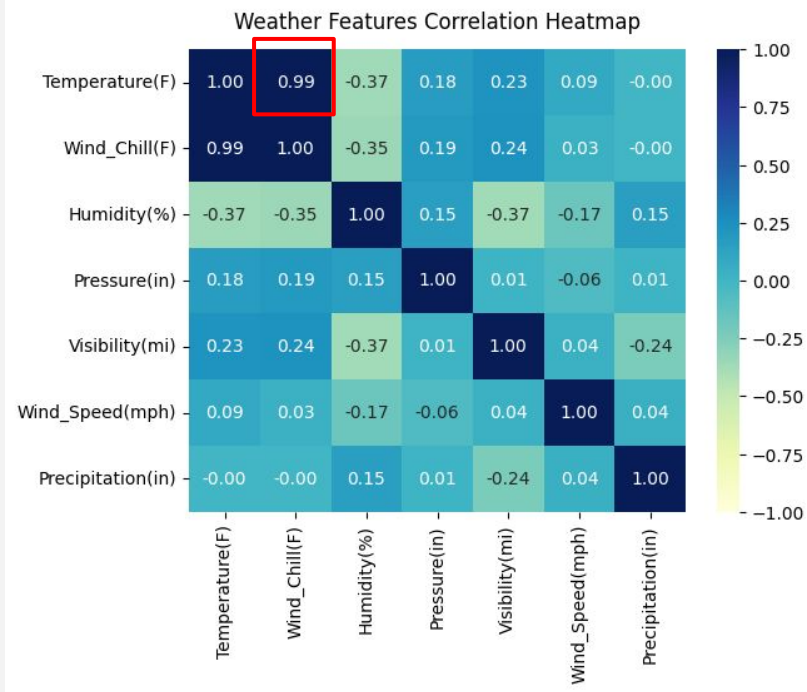


Data Exploration Insights - Time/Location

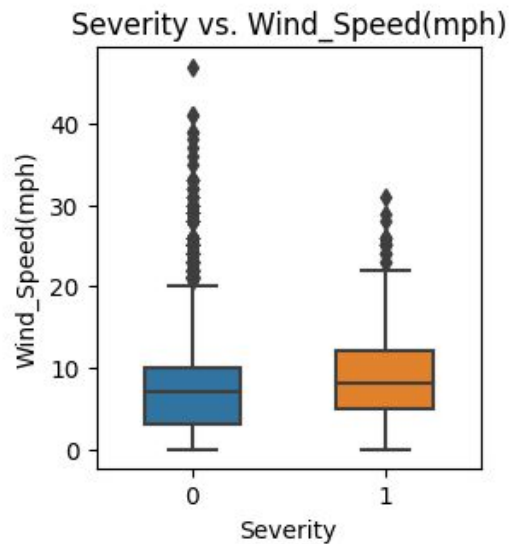
Severe Car Accident Probability by State



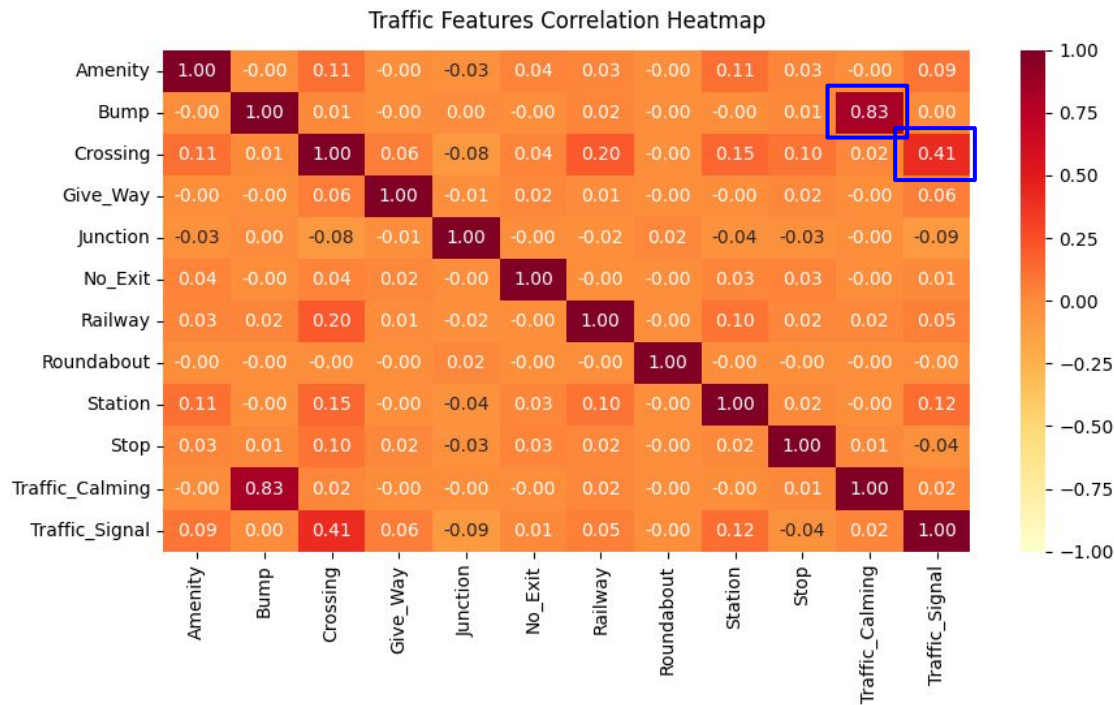
Data Exploration Insights - Weather



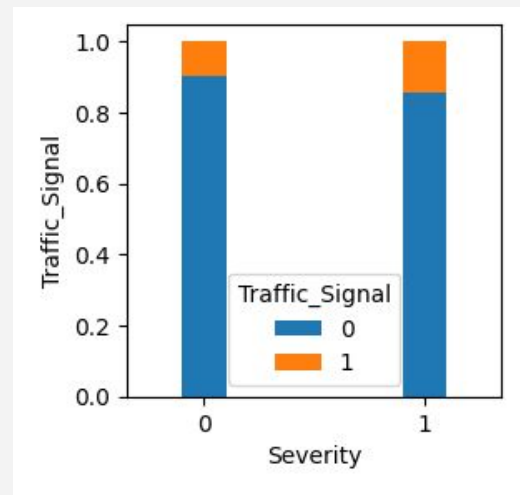
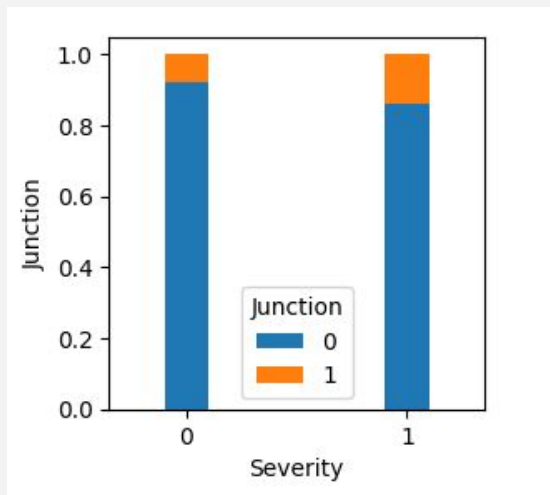
Data Exploration Insights - Weather



Data Exploration Insights - Traffic

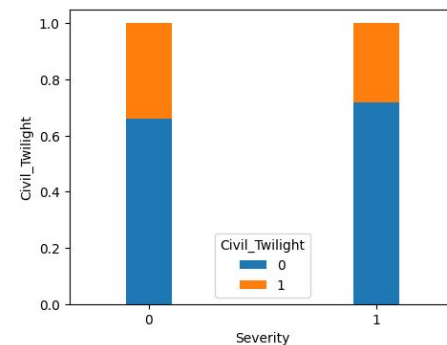
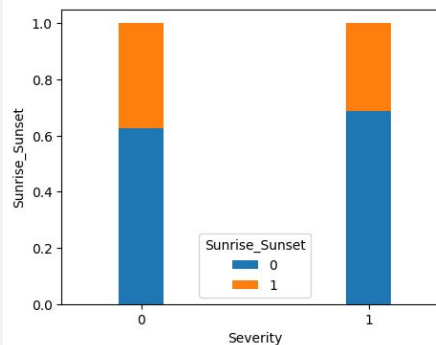
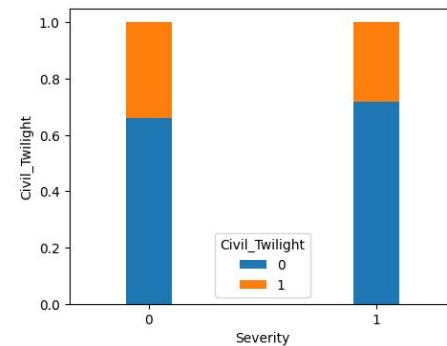
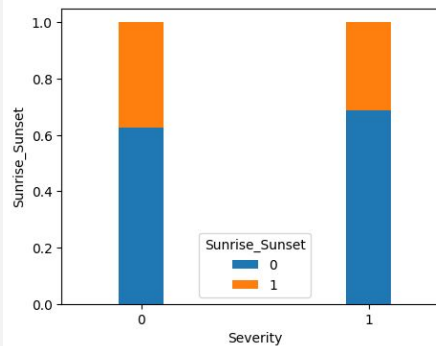
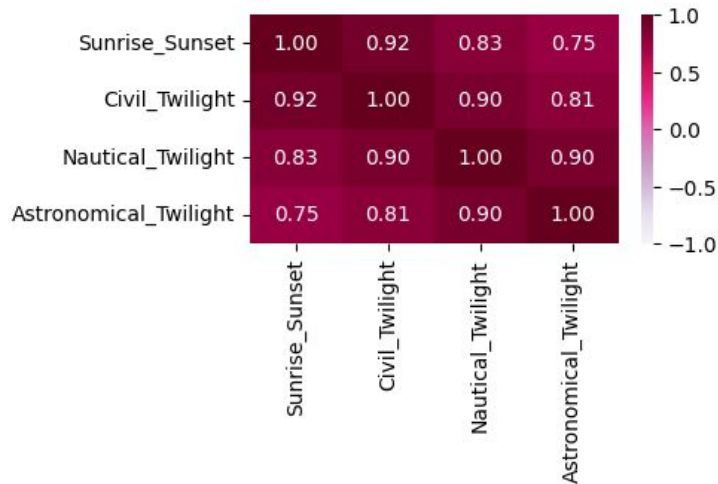


Data Exploration Insights - Traffic



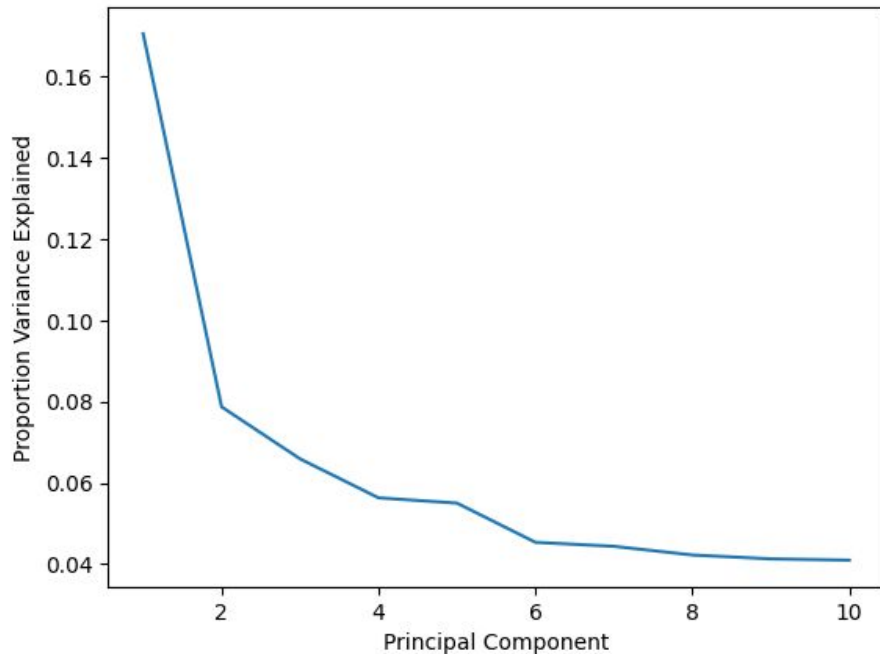
Data Exploration Insights - Sunlight

Sunlight Features Correlation Heatmap

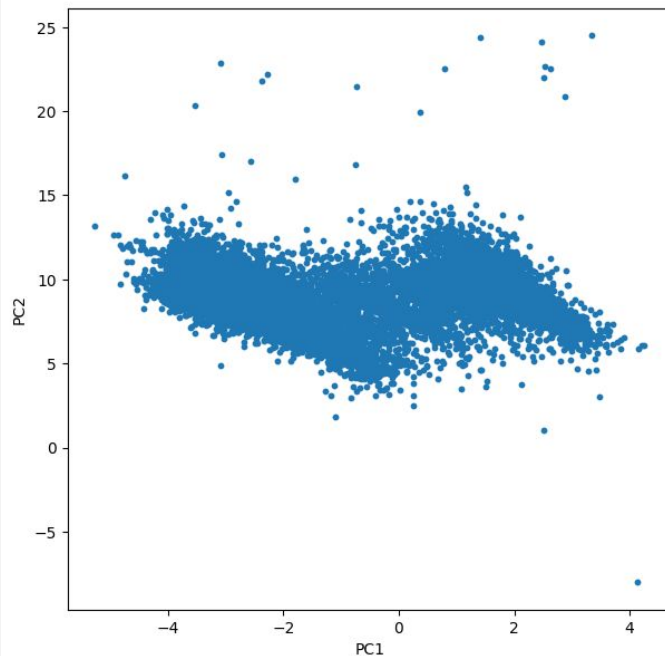


Unsupervised Learning - PCA

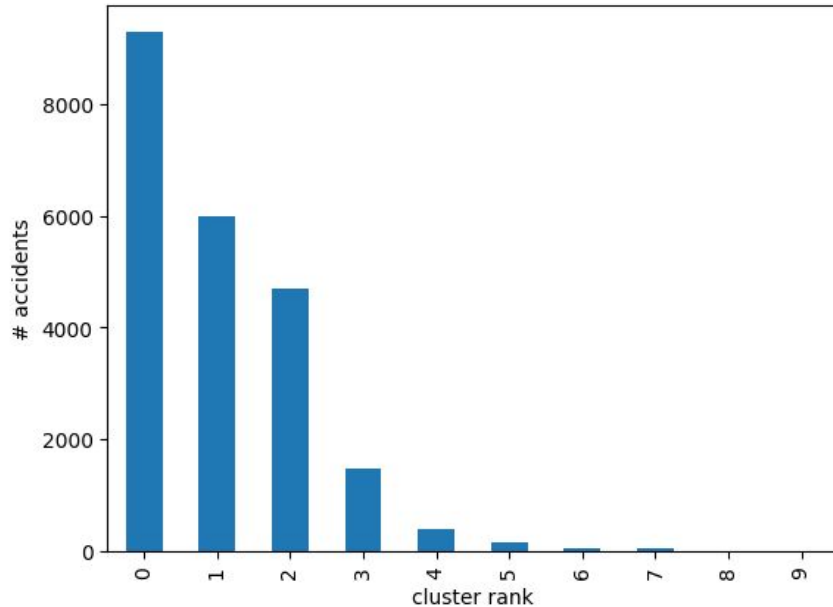
Car Accidents PCA Scree Plot



Car Accidents ID PC2 Scores vs. PC1 Scores



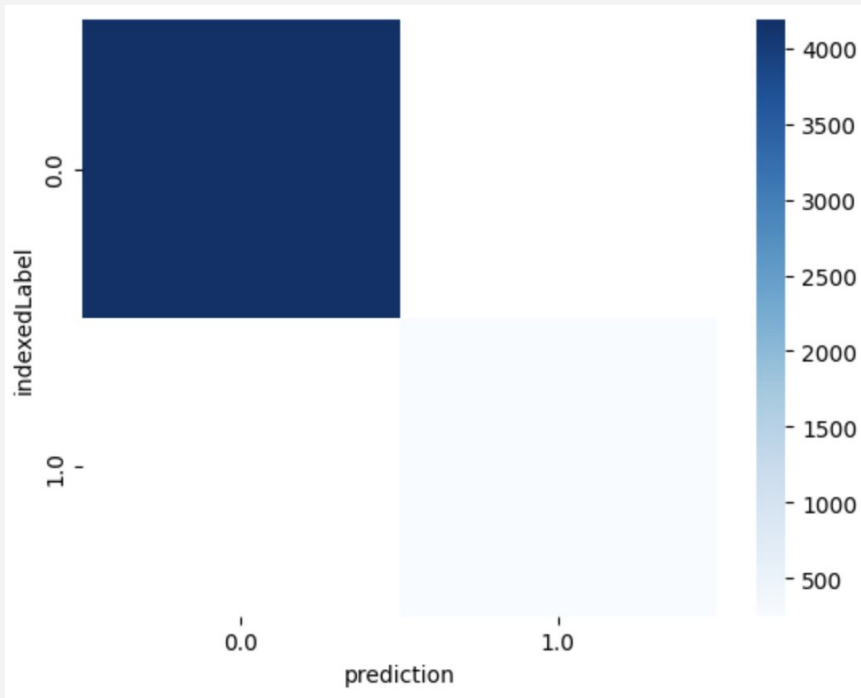
Unsupervised Learning - Clustering



cluster	accident ID
0	[A-6167, A-11427, A-30229, A-30727, A-48123, A...
1	[A-177650, A-226739, A-233440, A-241868, A-246...
2	[A-9622, A-171957, A-225704, A-226161, A-22782...
3	[A-230878, A-510717, A-615765, A-693623, A-746...
4	[A-229774, A-233646, A-233715, A-233935, A-240...
5	[A-33485, A-37199, A-43927, A-44963, A-106437,...
6	[A-95680, A-168806, A-222225, A-225232, A-2327...
7	[A-29445, A-33213, A-34599, A-37419, A-38643, ...
8	[A-219, A-15193, A-33645, A-34018, A-34051, A-...
9	[A-166749, A-455630, A-820816, A-830190, A-943...

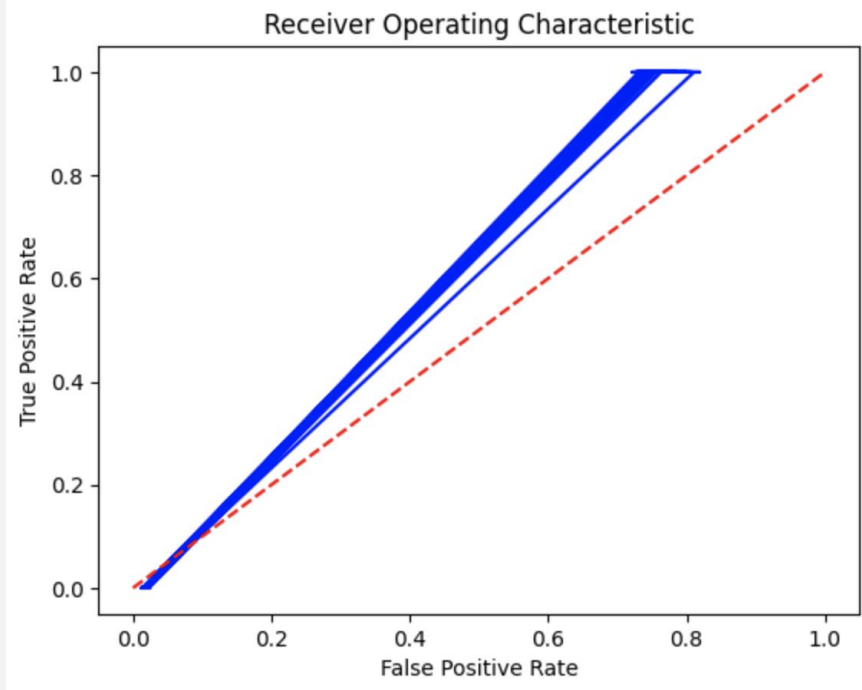
Supervised Learning - Ordinal Logistic Regression

Confusion Matrix



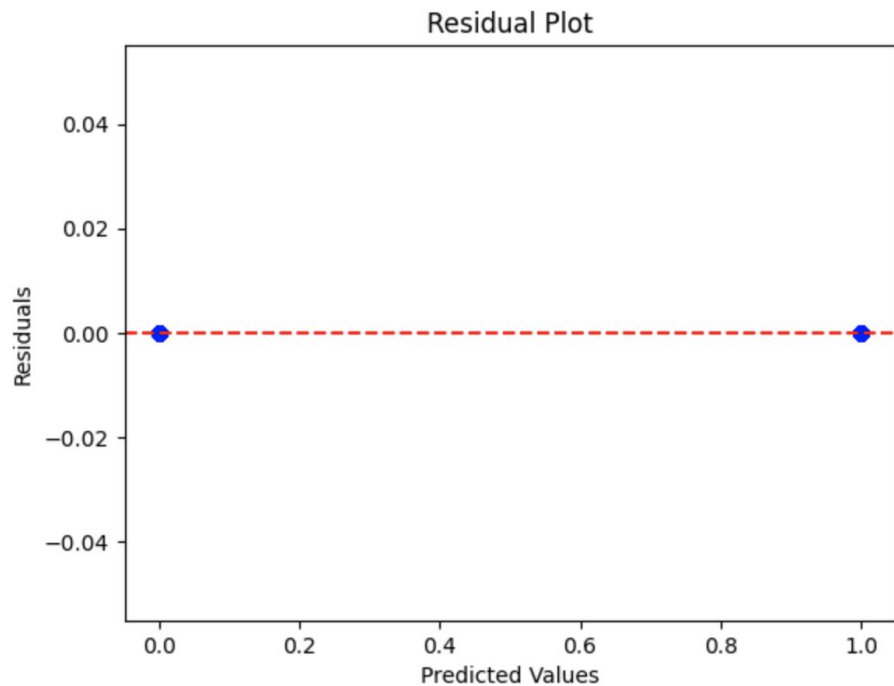
Supervised Learning - Ordinal Logistic Regression

ROC Curve

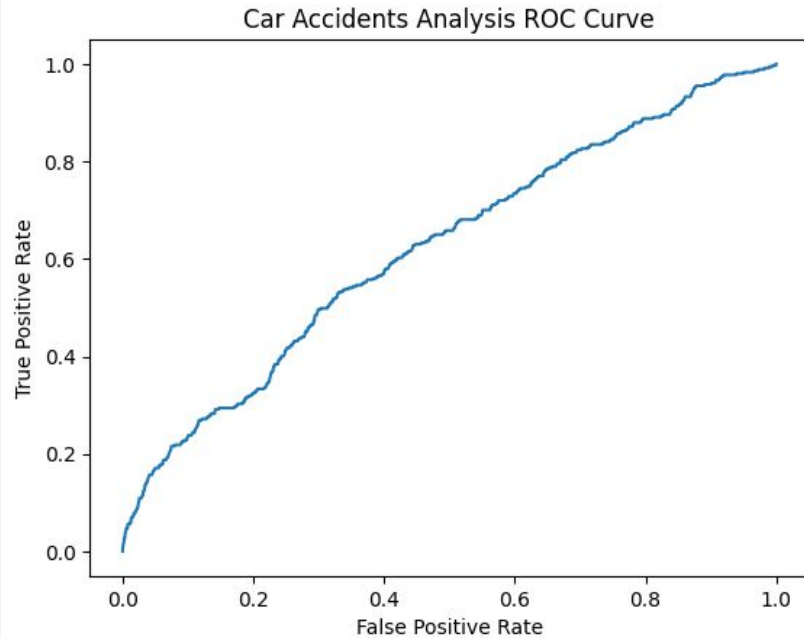


Supervised Learning - Ordinal Logistic Regression

Residual Plot



Supervised Learning - Logistic Regression



BinaryClassificationEvaluator: 0.6229

Distance	0.0112
Temperature	0.0017
Humidity	0.0044
Pressure	-0.2081
Visibility	0.0065
Wind Speed	0.0348
Precipitation	0.1192
Crossing	-0.5776
Junction	0.7239
Traffic Signal	0.6251
Sunrise_Sunset	-0.1742
Nautical Twilight	0.1324
Astronomical Twilight	-0.2928

Models - Decision Trees

Model	CV Folds	Evaluation Metric	Evaluation Value
Random Forest	3	AUC	0.71
Gradient Boosted	3	AUC	0.67

Models - Random Forest Predictions

		ACTUAL	
PREDICTED		TP 35	FP 155
		FN 213	TN 4045

Models - Random Forest Inferences

Feature	Importance
Distance	0.34
Pressure	0.25
Temperature	0.14
Wind Speed	0.10
Humidity	0.10

An isometric illustration of a city street scene. It features various elements like buildings, trees, pedestrians, cars, and a bicycle. The scene is viewed from an elevated perspective, showing the layout of the street and surrounding urban environment. The colors are muted and the style is clean and modern.

Conclusion

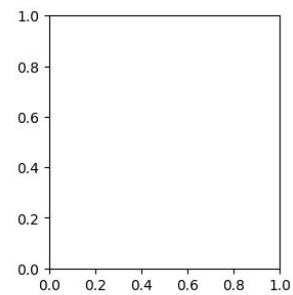
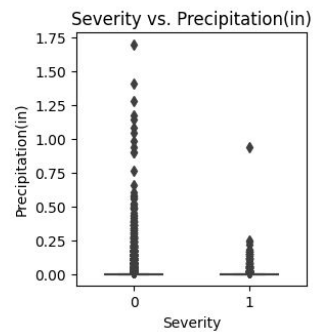
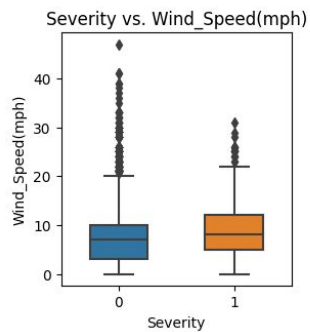
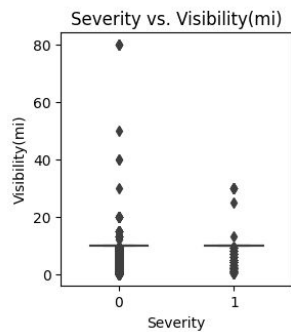
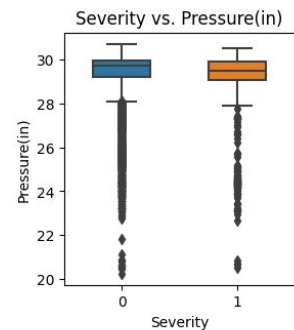
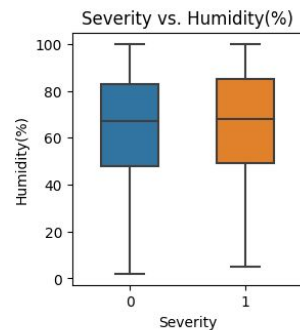
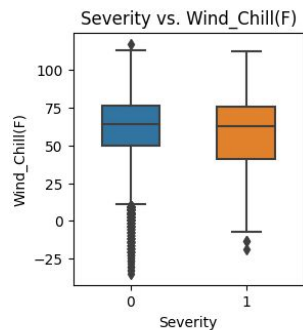
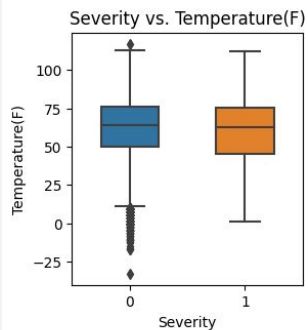
- Using satellite or some other means to detect distance traveled after impact can increase odds of successfully identifying severe accidents
- Random Forest yielded the best results for predicting severity. Once fine-tuned, it can be shared with companies and industries previously mentioned
- Self-driving cars would be able to utilize the finalized machine learning algorithm to lead to safer driving during conditions that could potentially lead to a severe car accident
- High wind speeds are more dangerous than many people may think



References

- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. "A Countrywide Traffic Accident Dataset.", 2019.
- Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, and Rajiv Ramnath. "Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights." In proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM, 2019.
- NHTSA's National Center for Statistics and Analysis. (n.d.). Early Estimate of Motor Vehicle Traffic Fatalities for the First 9 Months (January-September) of 2021. Retrieved April 22, 2023, from <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813240>
- Simon, S. (2023, January 27). How many people die from car accidents each year? Forbes. Retrieved March 4, 2023, from <https://www.forbes.com/advisor/legal/auto-accident/car-accident-deaths/>

Appendix



Appendix

