

Foundations of Machine Learning Results

Sam O'Brien, Colin Bitz, Lucas Vallarino, Will Mitchell

December 10, 2024

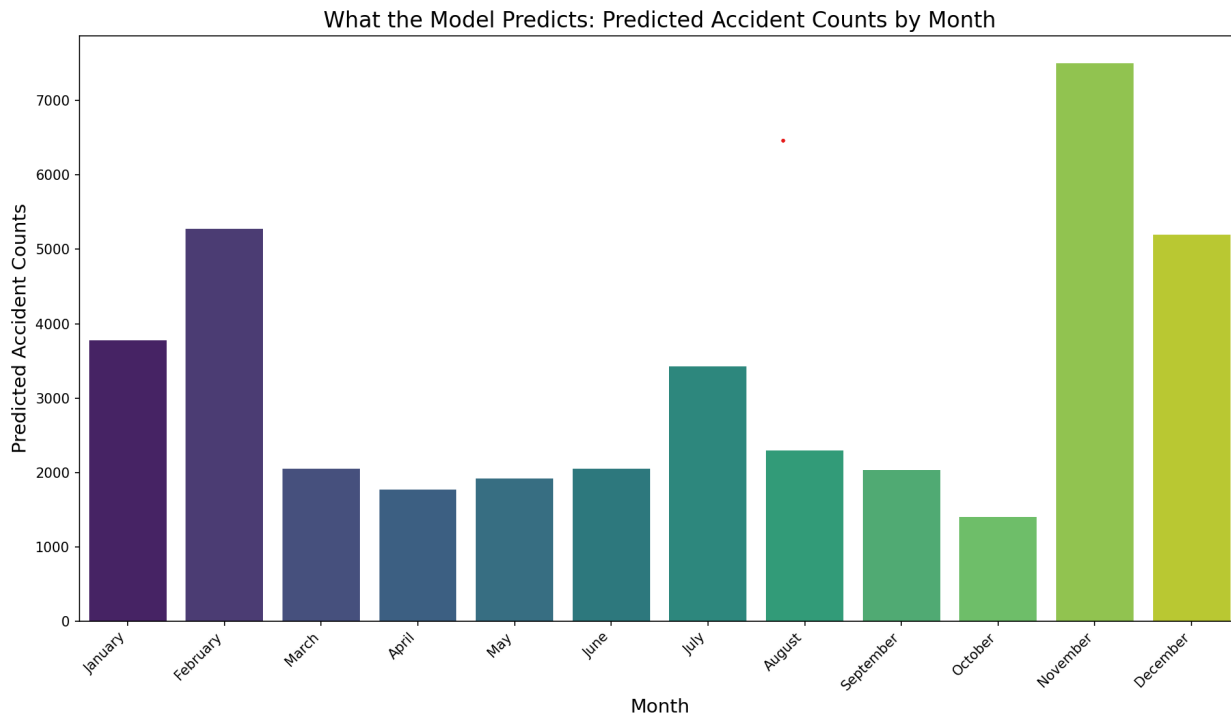
Question:

“For people planning road-related travel, what are the worst months to drive in Virginia based upon accidents per month?”

Method:

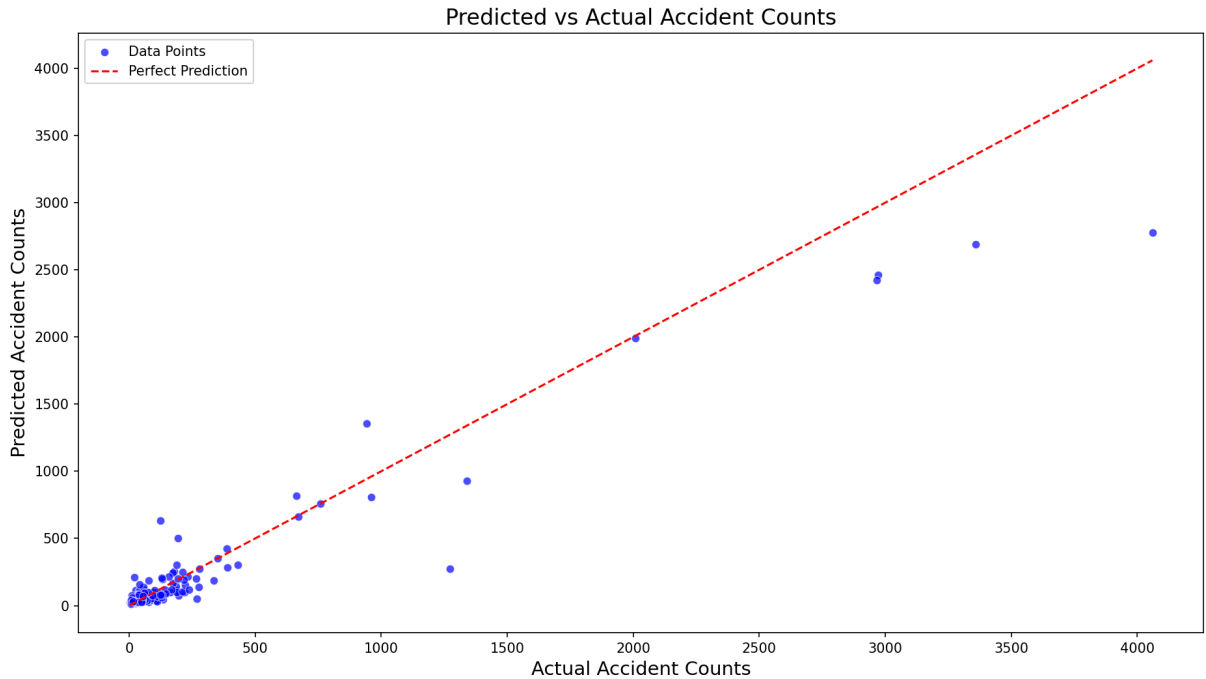
We used a random forest regressor to predict accident counts in the state of Virginia for a given month. We trained the model using a multitude of weather, location, and time factors to predict the accident counts.

Results:



The bar plot shows the model's predictions of accidents per month.

- The model's results follow common logic as the holiday seasons appear to be the worst months of the year to drive in Virginia.
- You can observe 2 tails at either end of the chart, demonstrating the increase of accidents
- The key takeaways from this graph and the overall model purpose are:
 - Holidays, specifically November through January, appear to be the worst times to drive in Virginia.
 - Try also to avoid road-related travel in July, as there appears to be an increase in accidents during that specific month.



This graph demonstrates the model's performance. There are a few key takeaways to note:

- Most of the data points are clustered near the bottom-left where the values are close to 0
 - Indicator that there is a high concentration of months with lower accident counts in Virginia with less months having higher counts.
- The model shows a few outliers, which may explain why there are some instances of the model failing to predict instances of higher months more accurately. (Less training data for these instances.)

Overall the model has a strong performance and is useful for predicting the accident counts per month in Virginia. See below:

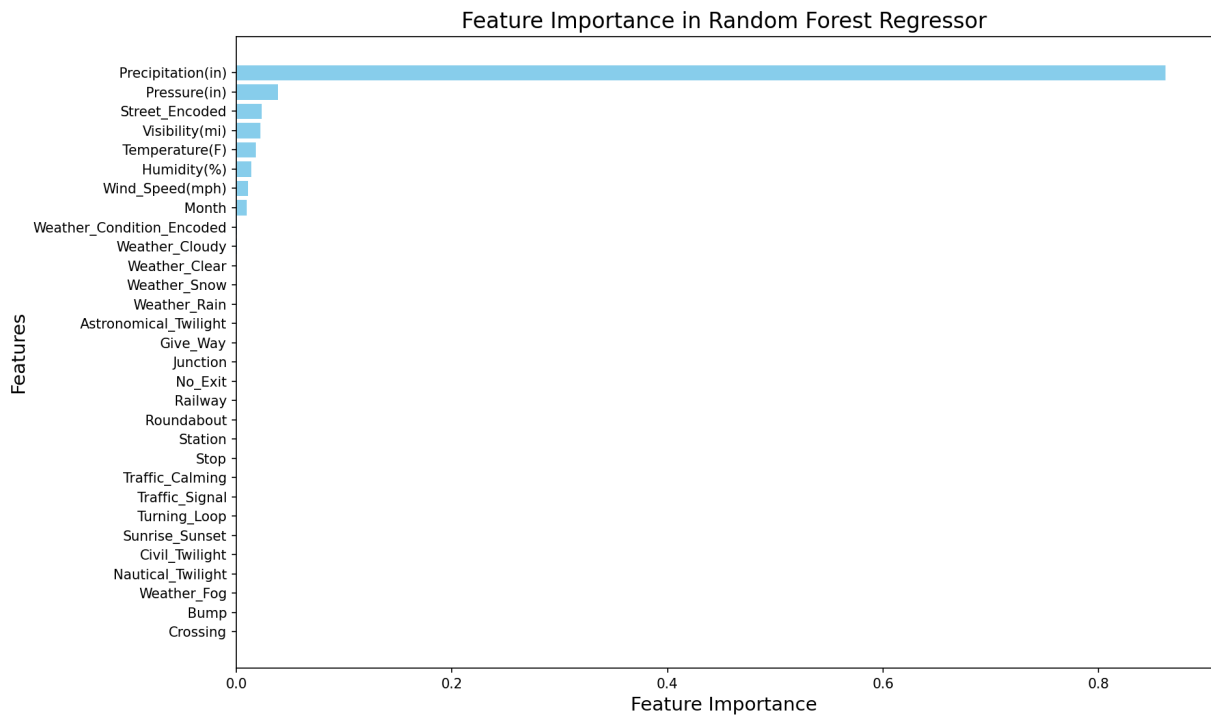
Model Performance Metrics:

Train R^2 : 0.9825

Test R^2 : 0.9095

Train MAE: 15.1399

Test MAE: 39.6383



This graph demonstrates what features the random forest model most often selects:

- The precipitation amount dominates the prediction of accident counts in this model. This is extremely logical as it is common knowledge that rain, snow, sleet, and other forms of weather make driving much more dangerous.
- The model uses other variables like street, and wind speed, and month, to help round out the decisions, but almost all of the prediction is associated with the precipitation amount.
 - This is useful because it suggests that precipitation amount is a heavy determinator of accident occurrences.

Conclusion:

- It appears that the worst months to drive are during the holiday season, this includes November, December, and January.
- Precipitation is a superior predictor of accident counts for a given month.
- In general, be careful when driving during the holiday season, especially if there is a lot of weather.
 - As a criticism, it should be noted that we filtered the model to select the top 150 roads based upon accident count. We did this because a vast majority of the roads in the model have a single instance of an accident, or very few.
 - In short, this model generalized to the busiest roads, or most dangerous roads, rather than all roads in Virginia.
 - So, this model is essentially useful if you are planning to travel on a major interstate.