

ACCIDENTALLY_LATE

J. KIRK, S. MCNAIR, N. FOSTER, K. HOLMBERG,
M. ONEAL

SELECTED TOPIC

- Traffic accidents 2016-2022
- Impact of Covid: increase in accidents in Covid years vs. Pre-Covid years
- Impact of weather: various feature available in data set



QUESTIONS EXPECTED TO BE ANSWERED

What impact has Covid had on the severity of car accidents in the US?

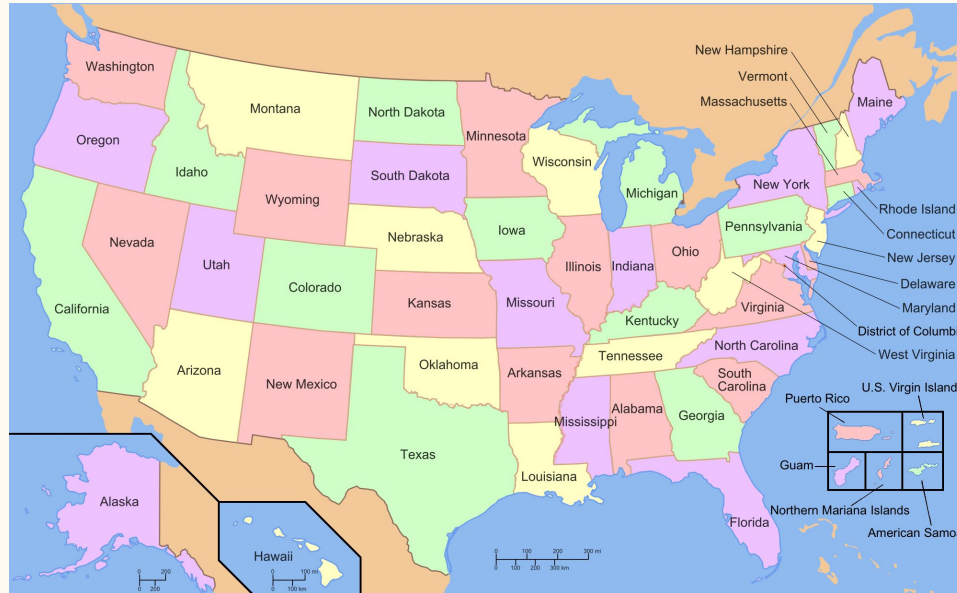
Has weather played a role in the information?

Top 5 Weather Conditions with the most accident occurrences

	<u>Accidents updated</u>	<u>PreCOVID accidents</u>	<u>COVID accidents</u>
<i>Fair</i>	1044151	115198	927250
<i>Cloudy</i>	335487	41604	293072
<i>Mostly Cloudy</i>	288935	32253	253466
<i>Partly Cloudy</i>	196112	31529	163688
<i>Light Rain</i>	121335	18454	89987

REASON TOPIC SELECTED

**We were curious to see the effect Covid had on traffic accidents
across the country**



TECHNOLOGIES



DESCRIPTION OF DATA SOURCE

- US-Accidents: A Countrywide Traffic Accident Dataset - Sobhan Moosavi
(smoosavi.org)
 - Dataset has 47 columns of information; we included the following columns for the purpose of the project: severity, start time, state, temperature, visibility, wind speed, precipitation, weather condition, sunrise vs sunset, year, and date
- cphalpert/census-regions: US Census Bureau Regions and Divisions by State
(github.com)
 - US Census Bureau Regions and Divisions by State; 4 regions, 7 divisions

DATA EXPLORATION PHASE INFORMATION

Dataframe Relationship



DATA EXPLORATION PHASE CON'T

```
1 # Get count of unique values in the 'year' column from 'accidents_updated'  
2 print(accidents_updated['year'].value_counts())
```

```
2021    1419863  
2020     575031  
2019     198861  
2018      21139  
2017      19517  
2016      11009  
Name: year, dtype: int64
```

```
1 # Get count of unique values in the 'year' column from 'PreCOVID_accidents'  
2 print(PreCOVID_accidents['year'].value_counts())
```

```
2019     198861  
2020      65722  
2018      12458  
Name: year, dtype: int64
```

```
1 # Get count of unique values in the 'year' column from 'COVID_accidents'  
2 print(COVID_accidents['year'].value_counts())
```

```
2021     1419863  
2020      508568  
Name: year, dtype: int64
```


DATA EXPLORATION PHASE CON'T

Severity Count

<u>Accidents updated</u>	<u>PreCOVID accidents</u>	<u>COVID accidents</u>
2 (2,084,426)	2 (228,702)	2 (1,828,676)
3 (71,340)	3 (27,134)	4 (39,448)
4 (66,076)	4 (21,055)	3 (36,885)
1 (23,578)	1 (150)	1 (23,422)

Supervised Learning

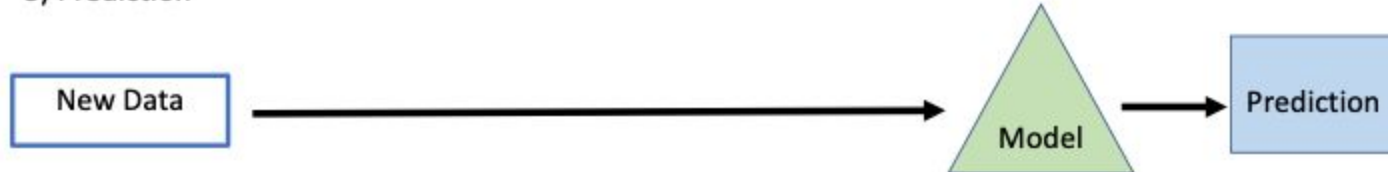
1) Training



2) Testing/Validation

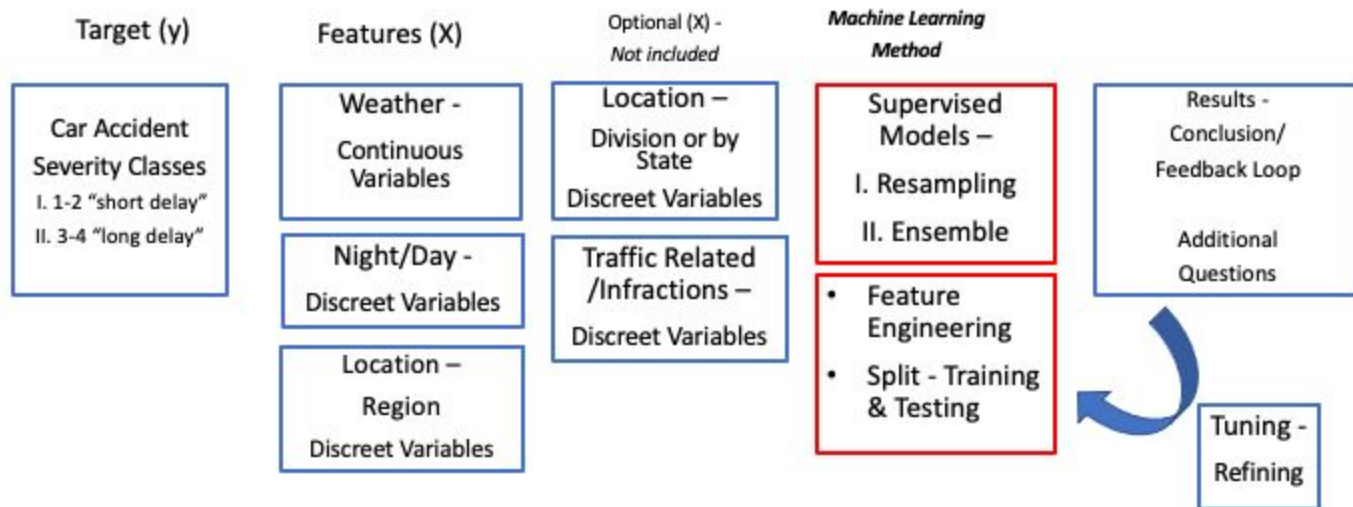


3) Prediction

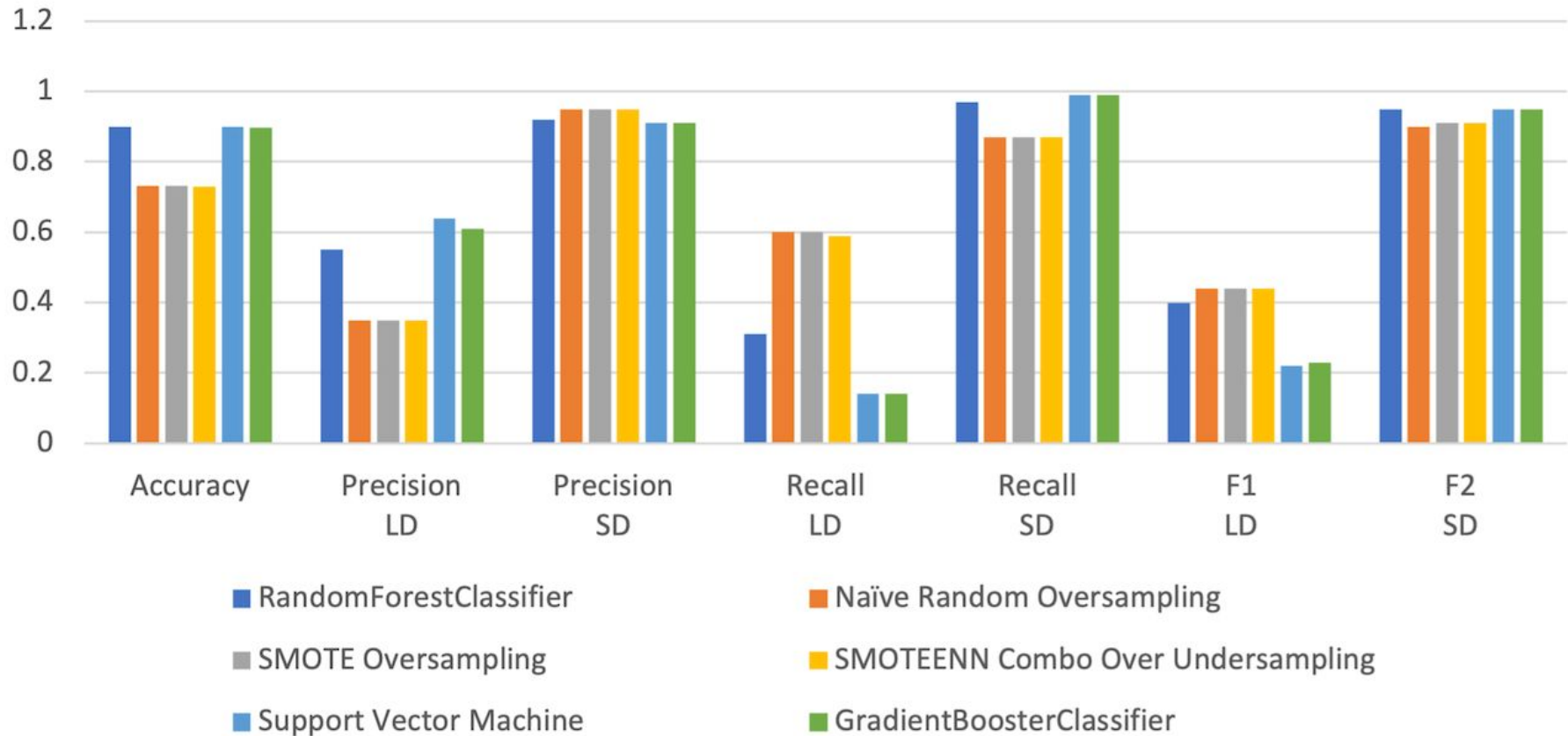


Machine Learning Diagram

Question: *Were U.S. car accidents less severe during Covid year 2020-2021 compared to years prior?*



MLM Scores Summary



TOP MODEL CONFUSION MATRICES

- Low F1 score and high false positives and false negatives is not good at predicting severity
- However, the question is if these features affect severity. Because of this, we did feature importance on the RFC model.

Random Forest Classifier

	Predicted long_delay	Predicted short_delay
Actual long_delay	4068	9079
Actual short_delay	3363	108490

Naive Random Oversampling

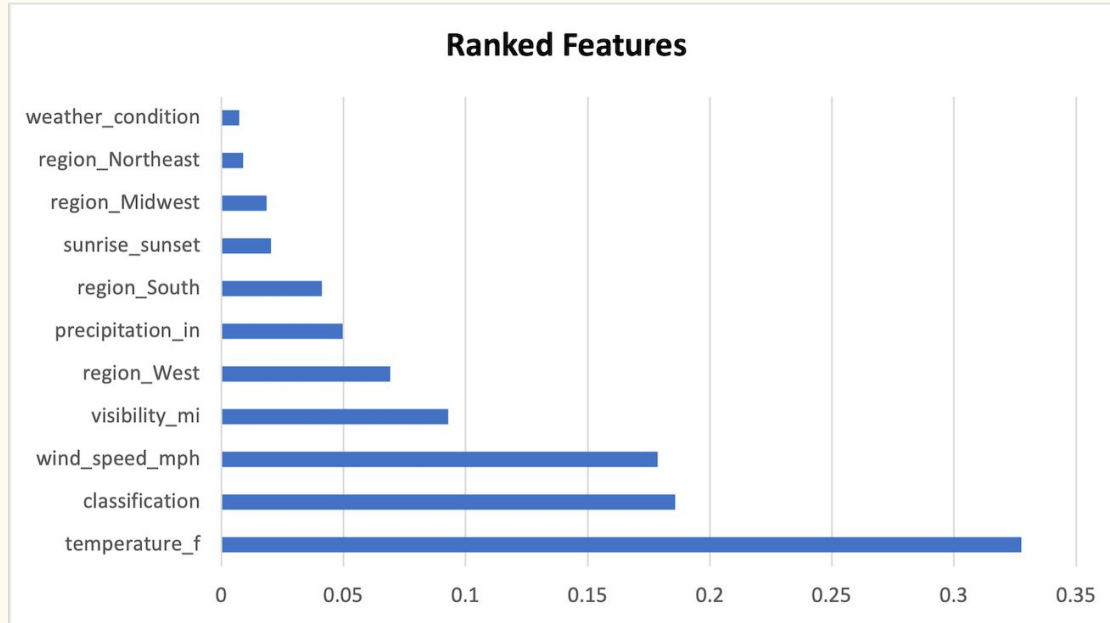
	Predicted long_delay	Predicted short_delay
Actual long_delay	8032	5367
Actual short_delay	14957	96644

Smote Oversampling

	Predicted long_delay	Predicted short_delay
Actual long_delay	7983	5416
Actual short_delay	14552	97049

MACHINE LEARNING CONCLUSION

RFC tells us *temperature* is the top ranked feature



RESULTS-LINK TO VISUALS IN TABLEAU

https://public.tableau.com/app/profile/kurt.holmberg/viz/ProjectBook2_16592207802480/Story1?publish=yes

RECOMMENDATIONS FOR FUTURE ANALYSES

- Why is there > 1 million difference between PreCovid and Covid accidents?
- What impact did the COVID lockdown/quarantine have on the number and severity of car accidents in each state in the US?
- Impact of severity of individual states with lockdown vs. states without lockdown during Covid years.

WHAT TO DO DIFFERENTLY NEXT TIME

- Change Github file structure at the beginning
- Save more frequently in Tableau and other programs
- Meet more frequently at the start of the project
- Do Anova study vs Pandas

QUESTIONS

