TRAFFIC DELAY PREDICTION OF ROAD ACCIDENTS IN TEXAS

Akshata Bhandiwad | Yash Tushar Chopada | Abeer Haruray

25TH NOVEMBER 2020

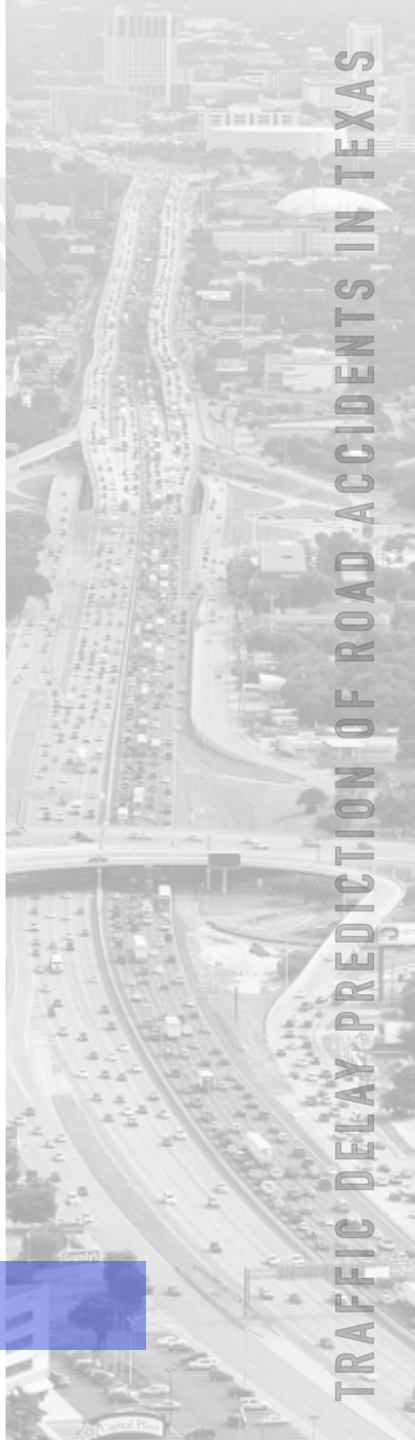
BUSINESS SCENARIO

12 million
vehicles involved in crashes in US (2018)

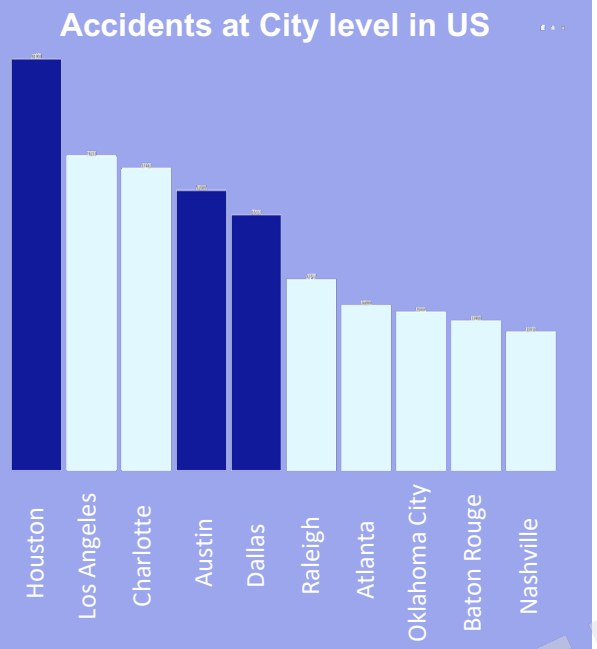
$242 billion
Estimated Economic cost Of road accidents in 2010

$28 billion
Congestion cost Post road accidents

- Need to **study the traffic delays** as a function of roadway and weather characteristics at the accident spots.
- Essential for a **Transportation Management Center (TMC)** to analyze and predict the traffic delay level post accident to efficiently manage their resources.
- Help the state-licensed **Emergency Medical Services (EMS)** Dispatch centers to recognize the requirement of airlift to save the life.



DATA SET : US ACCIDENTS (kaggle.com)



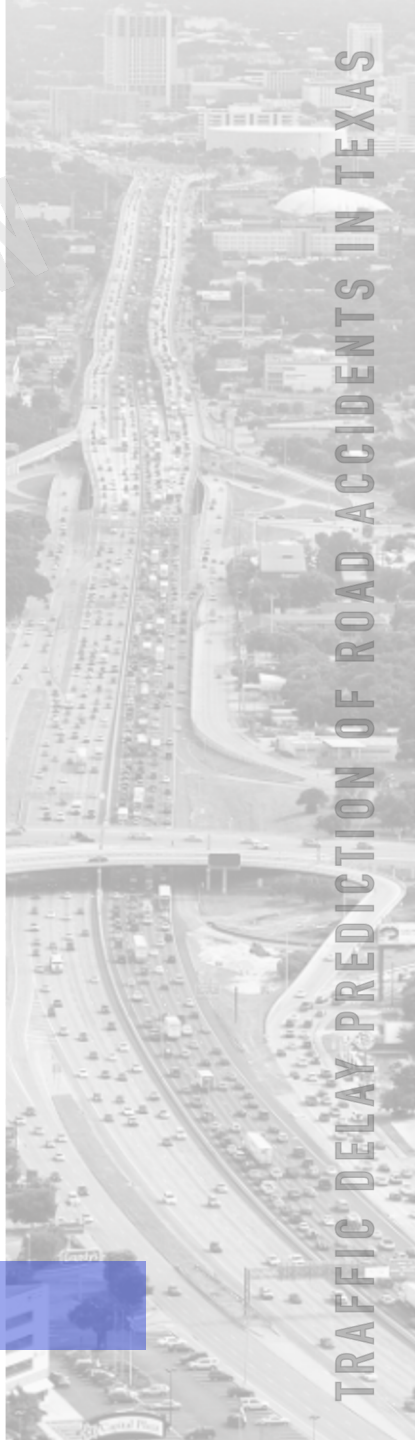
Texas

300k Observations
47 variables

- Original data set: **3.5 million** road accident records with **49 variables** describing city and state wise information, accident severity, time when impact of accident on traffic was dismissed, GPS coordinates, weather conditions, and nearby points of interest.

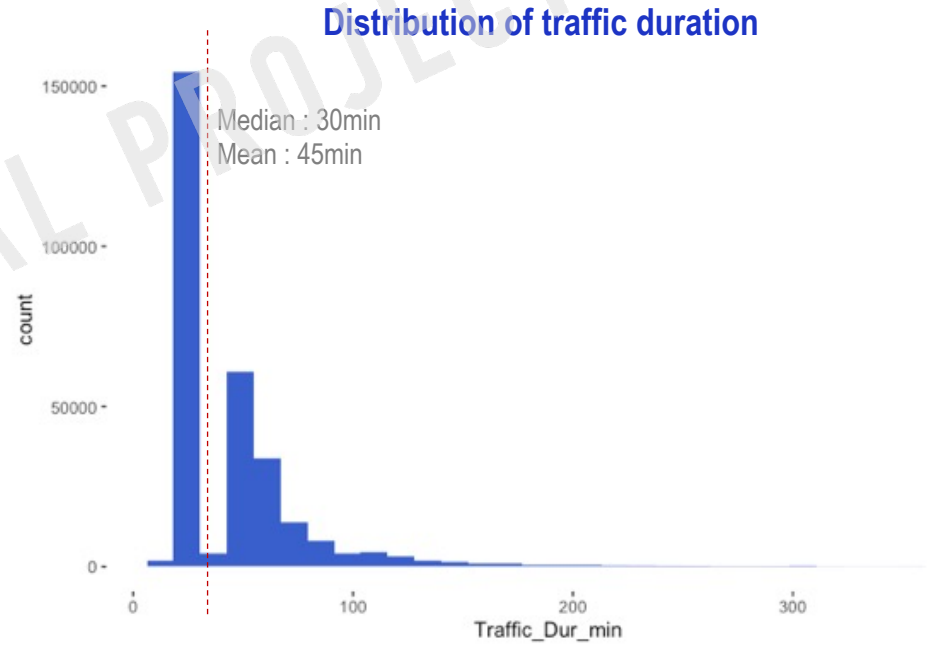
- **Calculated fields**: Traffic Duration (in min), Weekday/weekend, Time of day (STHR)

- **Data acknowledgements:** Moosavi, Sobhan, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, and Rajiv Ramnath. “A Countrywide Traffic Accident Dataset.”, arXiv preprint arXiv:1906.05409 (2019).



TEXAS : TRAFFIC DELAY LEVEL

- **Discrepancy in traffic delay level categorization** for duration of traffic accidents (in minutes) reported by different sources. Hence, there was a **need to reclassify** the traffic delay levels.



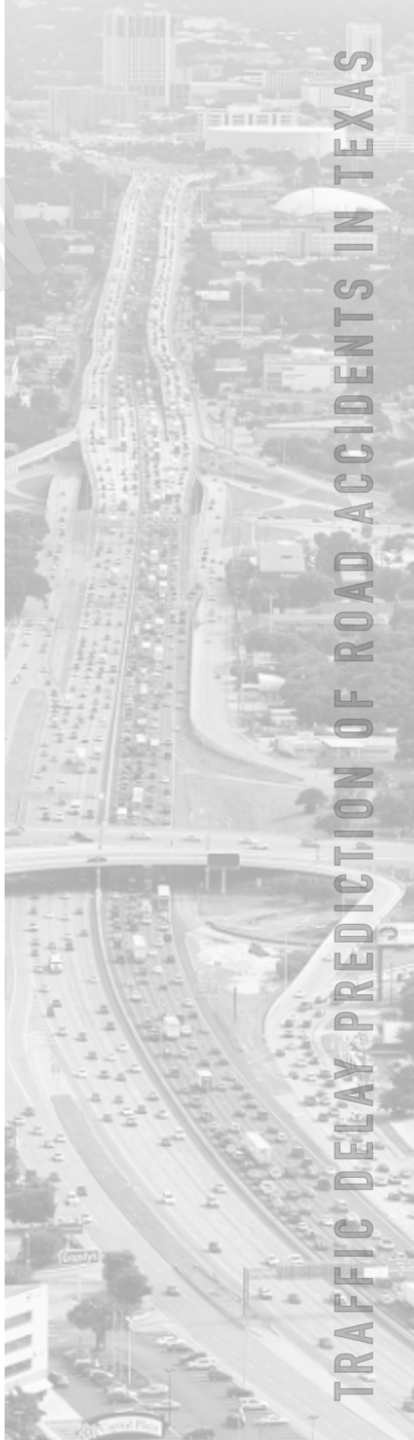
Binary Classification

Traffic duration <= 45min : 1
Traffic duration > 45min : 2

ALGORITHMS

- **Input Variables** considered, whose data will be available at the time an accident is reported at TMC
- Class of interest : **Traffic Delay level 2 (High)**

Algorithms	Sensitivity	Specificity	Accuracy
Binary logistic regression (probability threshold : 0.25)	0.76	0.60	64.5%
Discriminant Analysis (probability threshold : 0.25)	0.74	0.63	66.2%
Decision Tree	0.79	0.69	71.7%
K-Nearest Neighbors	**0.72**	**0.96**	**89.7%**
Random Forest	0.69	0.88	83.2%



KNN RESULTS

72.2 %

Sensitivity : proportion of delay level 2 that are correctly identified

96.7 %

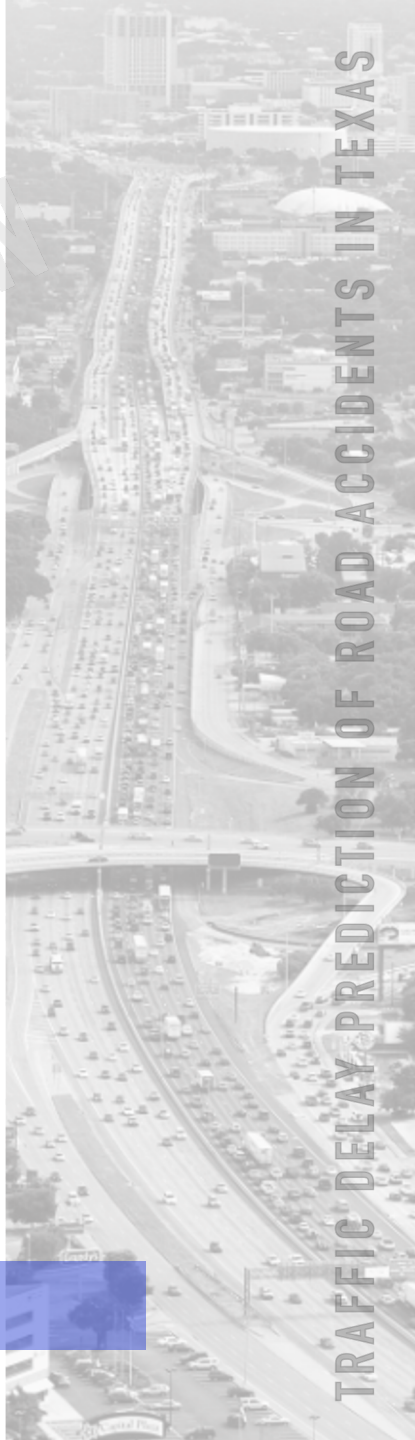
Specificity: proportion of delay level 1 that are correctly identified

89.7 %

Accuracy

Variables considered: Start_Lat, Start_Lng, Crossing, Give_Way, Junction, Railway, Roundabout, Station, Stop, Traffic_Signal', Day_Night, Precipitation, Pressure, Temperature, WindSpeed, Humidity, Weekend

```
Confusion Matrix and Statistics

          Reference
Prediction     1     2
         1 41180  4656
         2  1415 12122

               Accuracy : 0.8977
                 95% CI : (0.8953, 0.9002)
    No Information Rate : 0.7174
    P-Value [Acc > NIR] : < 0.0000000000000022

                  Kappa : 0.7321

 Mcnemar's Test P-Value : < 0.0000000000000022

            Sensitivity : 0.7225
            Specificity : 0.9668
         Pos Pred Value : 0.8955
         Neg Pred Value : 0.8984
             Prevalence : 0.2826
         Detection Rate : 0.2042
   Detection Prevalence : 0.2280
      Balanced Accuracy : 0.8446

       'Positive' Class : 2
```

IDEAS CAN CHANGE THE CITY

Thank you!