

Flatiron Data Science Mod 3 Final Project

Machine Learning Classification Model: Chicago Car Crashes

John J. Cho
April 3, 2020



Dataset: 2M+ Datapoints, 151 Features



- Crashes: 400,000+
- Vehicles: 800,000+
- People: 850,000+
- Features:
 - Crashes: 49
 - Vehicle: 72
 - People: 30



Predict: SEVERE Crashes



- ❏ SEVERE: at least ONE (1) incapacitating injury or fatality
- ❏ Features: reduced to 28 including:

- weather

- road conditions

- drivers, passengers, pedestrians

- damage costs

- lane count

- types of vehicles

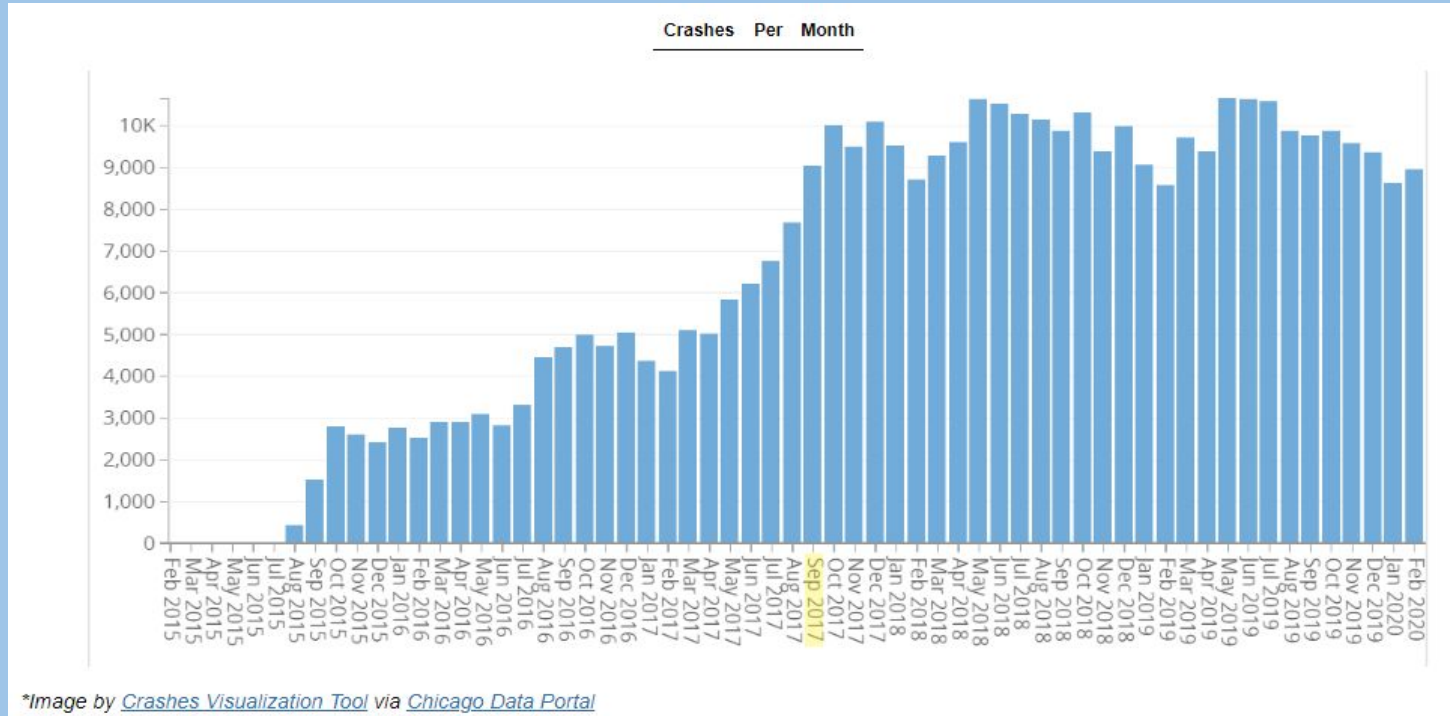
- alcohol involved

- airbag deployment

- traffic control devices

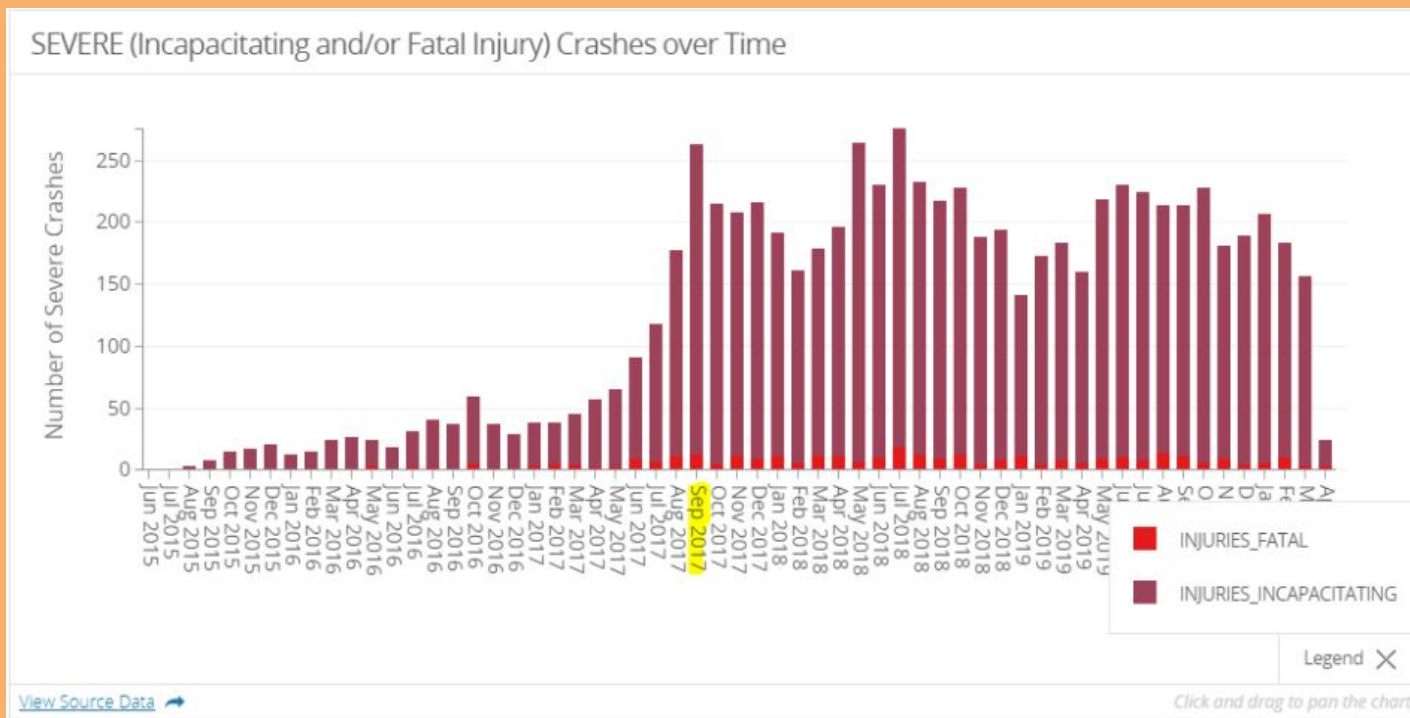
Exploring Our Data (Sep 2017-Feb 2020)

- How many car crashes are we talking about? ~8-11k per month



Exploring Our Data (Sep 2017-Feb 2020)

➤ How about SEVERE crashes? ~150-250 per month

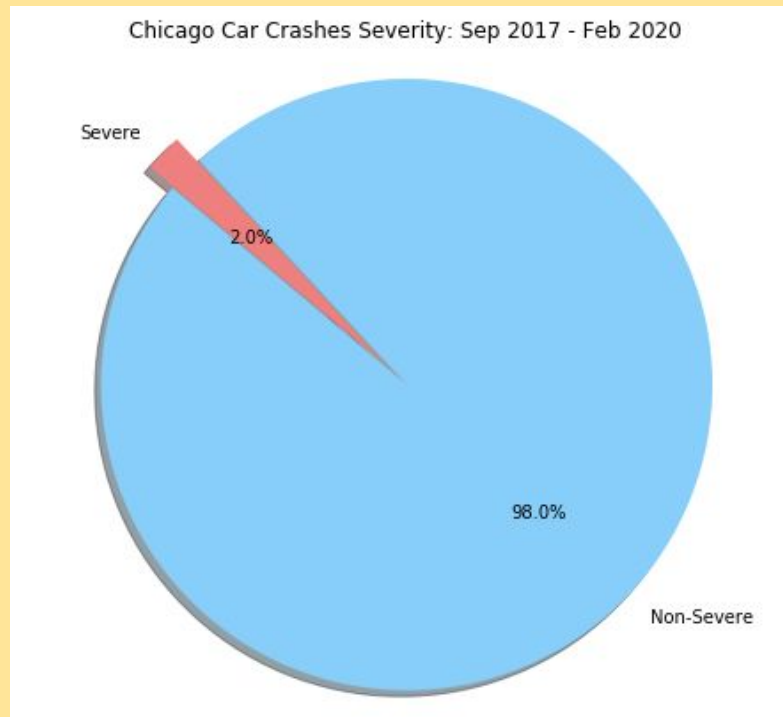


Machine Learning Prediction Models

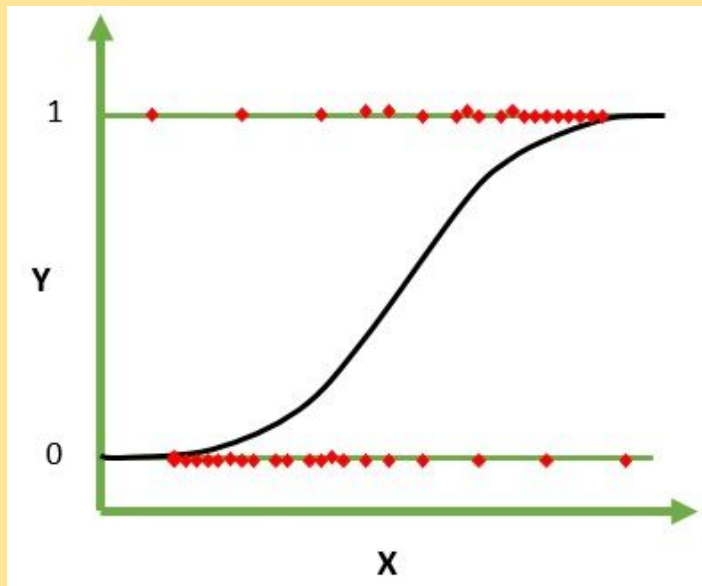
- Logistic Regression, Decision Trees, Boosting, SVM

Model Scoring Metrics - some caveats

- **Accuracy:** predicting ONLY non-severe would be 98% accurate!
- **Precision:** how correct SEVERE crash predictions are
- **Recall** ← correctly predicting EVERY ACTUAL SEVERE crash
 - Be okay with many severe crash predictions turning out to be non-severe



Best Model: Logistic Regression



Accuracy: 89%

Precision: 14%

Recall: 98%

❖ Out of **1,338 severe crashes** in our test set, this model correctly predicted **1,310** of them.



Highest SEVERE crash factors:

- Total number of injuries
- Involving pedestrian / bicyclist
- Ejected from / trapped in vehicle
- Speeding and/or alcohol involvement
- Airbag deployment

Recommendations and Future Work

- **City of Chicago** already does a **stellar** job at minimizing severe crashes.
- Number of overall crashes and severe crashes have remained consistent. **2% severity rate might represent a combined 'luck + human error' element that is already as low as it can go.**
- **Only recommendation would be to see if pedestrian, bicyclist safety could be further improved:**
 - Additional crosswalks, road markings, signs, audio/visual cues for when pedestrians/bicyclists are sharing the road
 - Safety awareness campaigns
 - Self driving cars? :-)
- Future work: Include location data that could possibly pinpoint specific neighborhood hotspots / police districts.



Thank you for your time!