PREDICTION OF THE CAR ACCIDENT SEVERITY FOR THE CITY OF SEATTLE

SILVIYA NINOVA

BACKGROUND INFORMATION

- 1.35 million people died in car accidents, according to the World Health Organization;
- Main cause of death for people 5-29 years;
- 8th leading cause of death globally;
- Thousands of people suffering from injuries of car accidents.

An efficient prevention is desperately needed to save lives on the road.

THIS PROJECT

- Goal?
  - Predict the severity of an accident and distinguish between property-damage and injury collisions.
- How?
  - Using supervised machine learning algorithms.
- Why?
  - So as to avoid accidents, in which people get injured.
  - Help the local authorities to:
    - warn the local population about high-risk conditions;
    - prepare a strategy for road—accident prevention.

DATASET

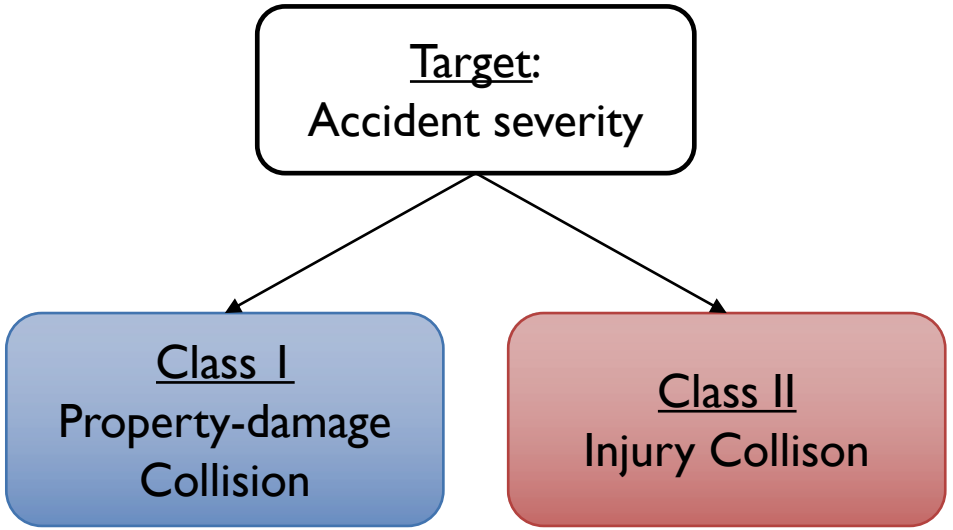
- Dataset on all collisions for the city of Seattle in the period 2004-2020, provided by the Seattle Police Department.
- Readily available online on the Seattle GeoData.
- The original dataset contains 194673 records and 37 features.

- Irrelevant or similar features were removed.
- All records with missing/incomplete information were discarded.

- The final dataset contains 142198 records and 21 features.

PREDICTION MODELS

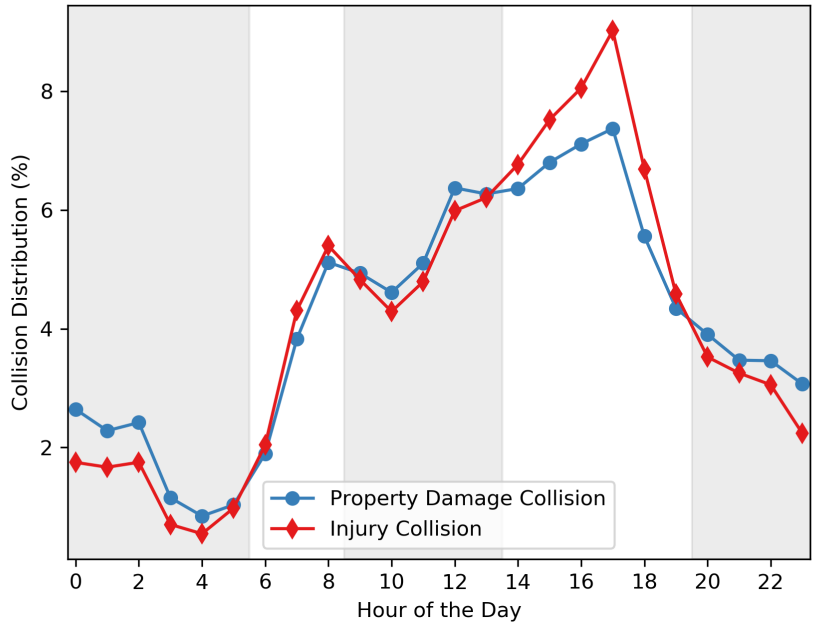
- Supervised machine learning models.
- Three binary classification techniques were tested:
  - Decision Tree
  - Random Forest
  - Extreme Gradient Boosting
- The data was corrected for imbalance.

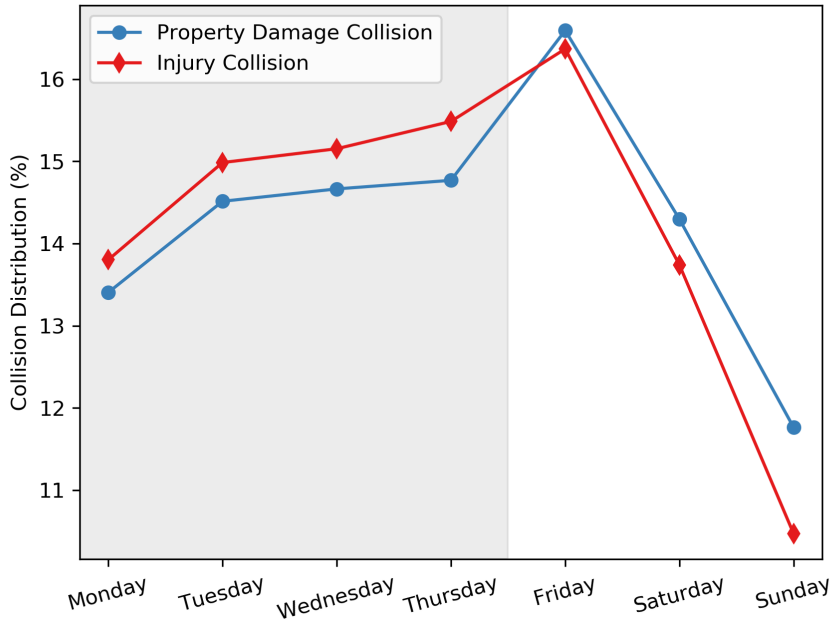


PREDICTION FEATURES

- The features used for the predictive model are separated into groups:
  - Time period;
  - Weather and road conditions;
  - Road segment and street illumination;
  - Participants in the accident;
  - Collision type;
  - Reckless driving.

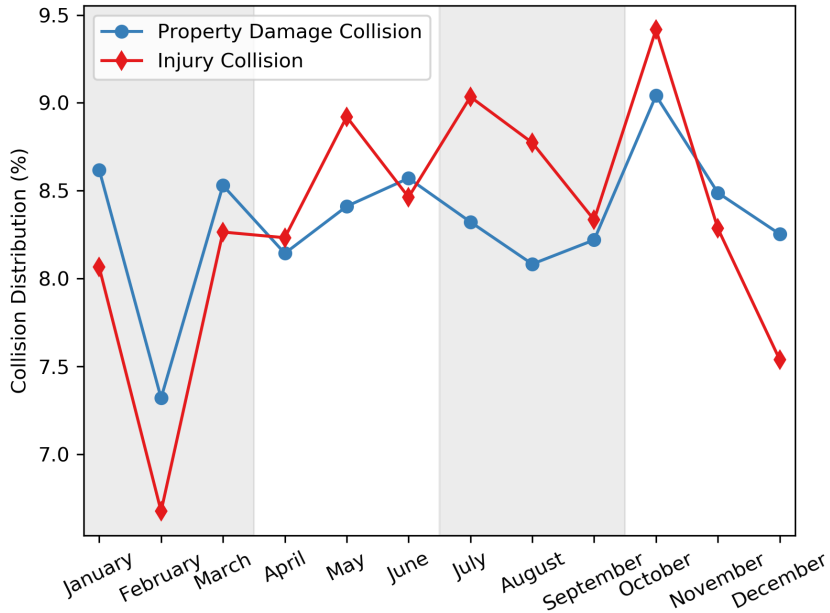
TIME PERIOD (1)

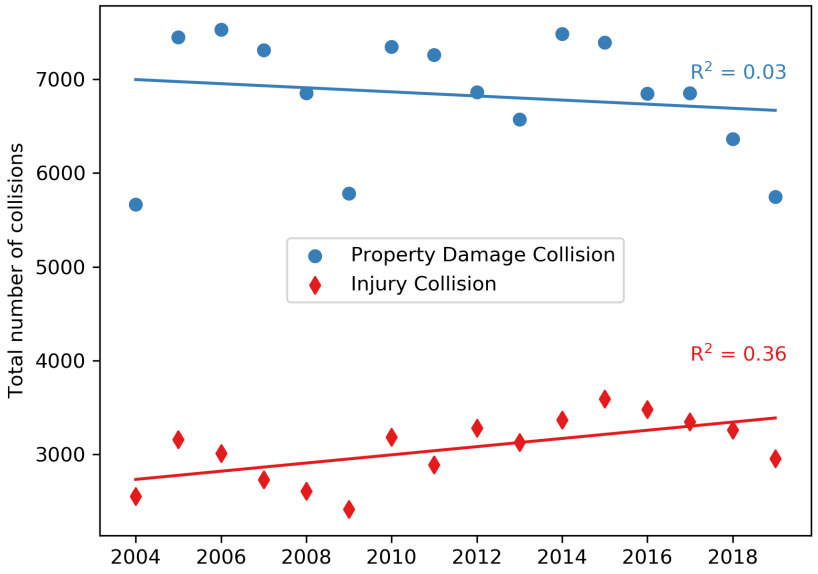




- Slightly more collisions of Class II happen during rush hour and less during the night.
- In proportion more injury collisions happen during the week than during the weekend.

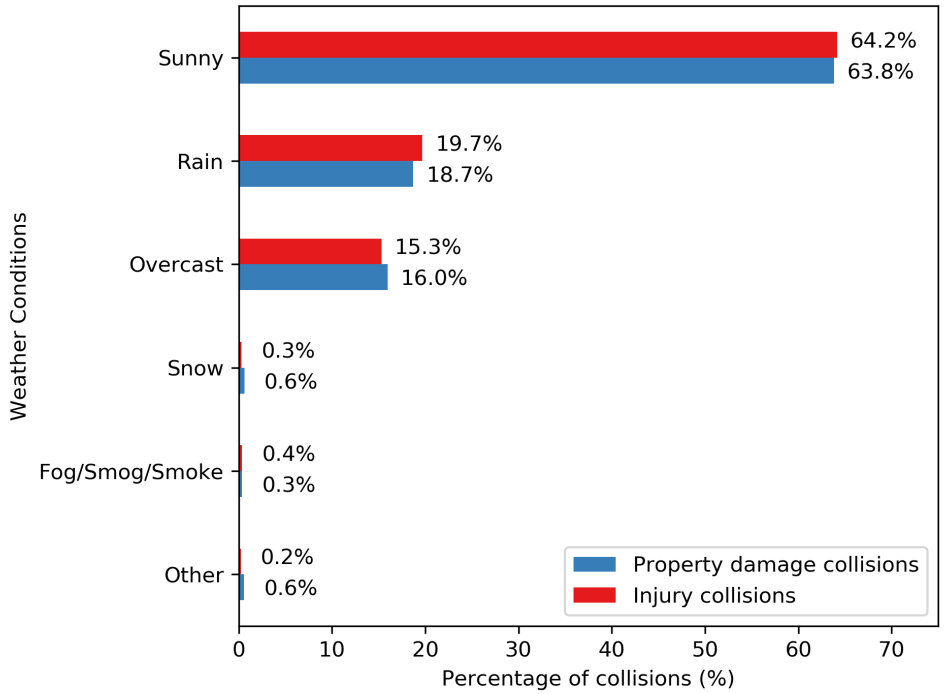
TIME PERIOD (11)

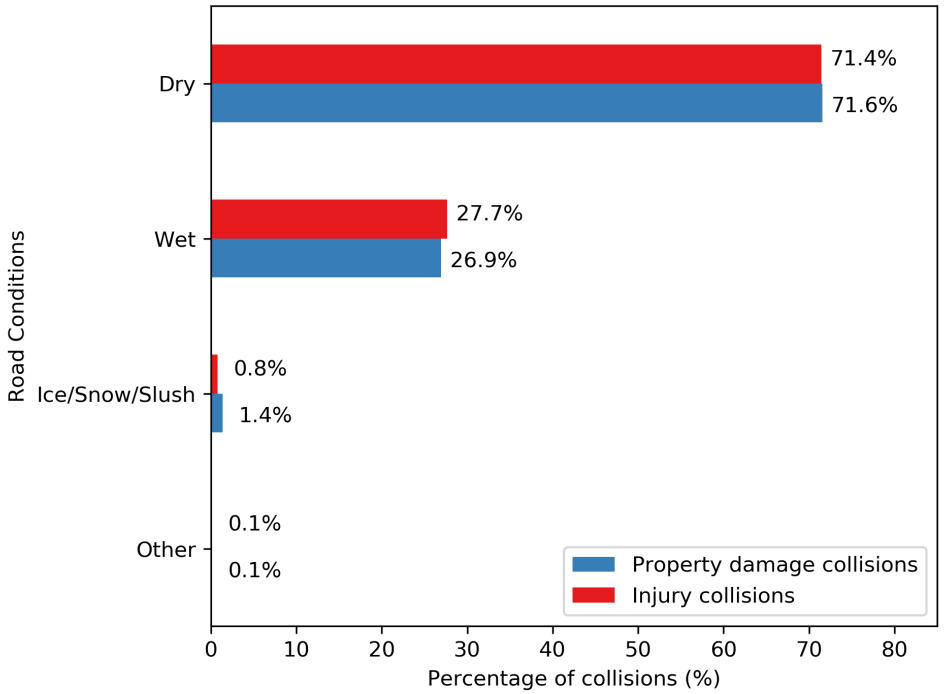




- In proportion more injury collisions happen during summer months.
- The total number of collisions per year does not follow a linear progression, but has a more complex pattern.

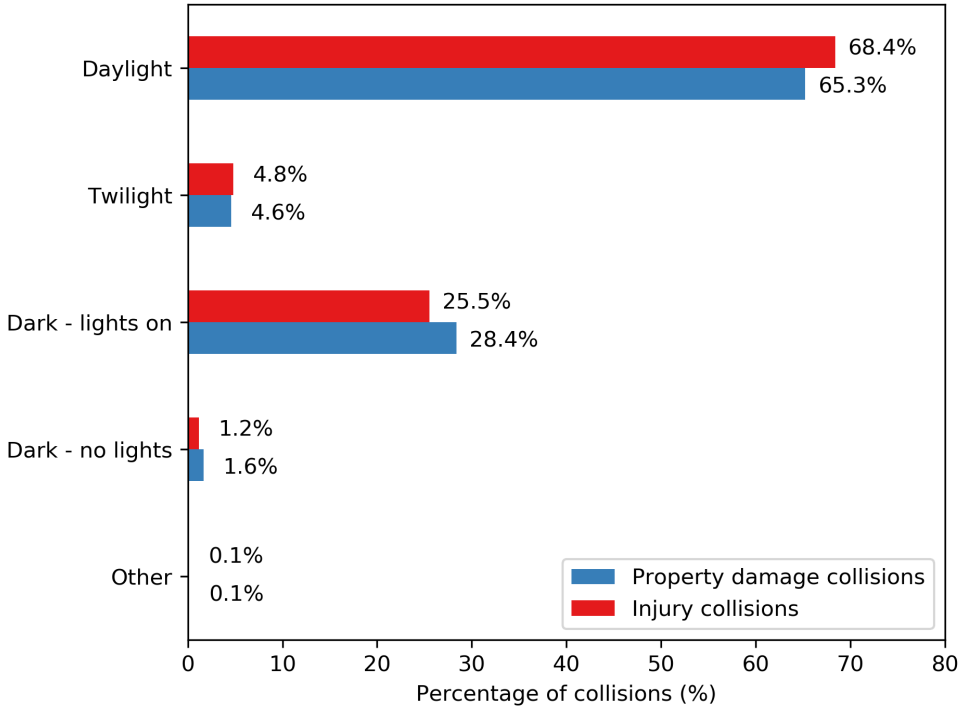
WEATHER AND ROAD CONDITIONS





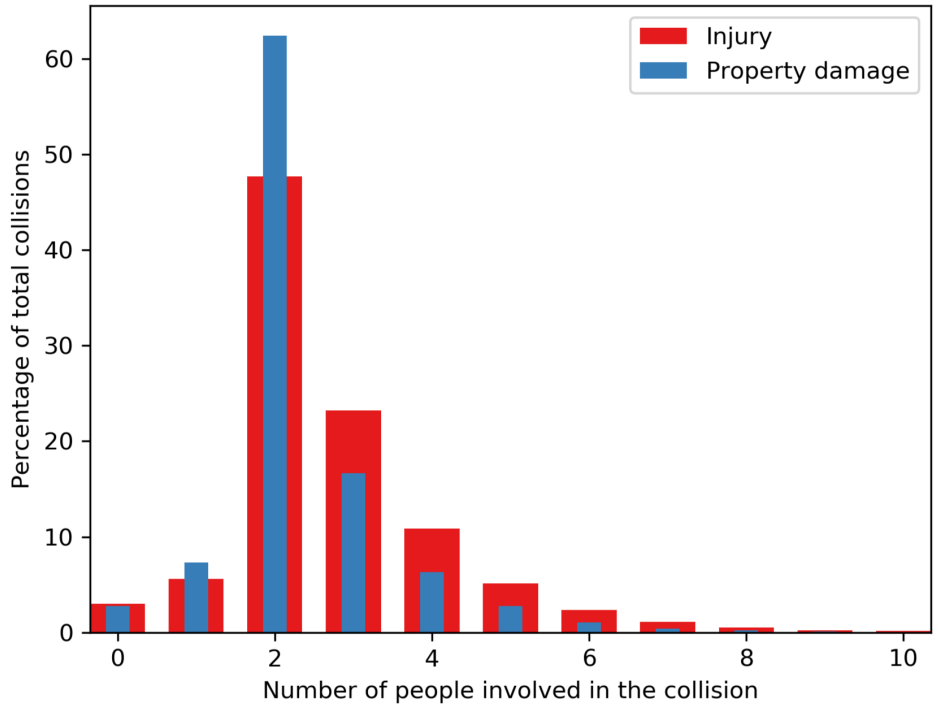
- Most collisions happen at good weather and road conditions.
- No condition is found to predominantly cause collisions of Class I or II.

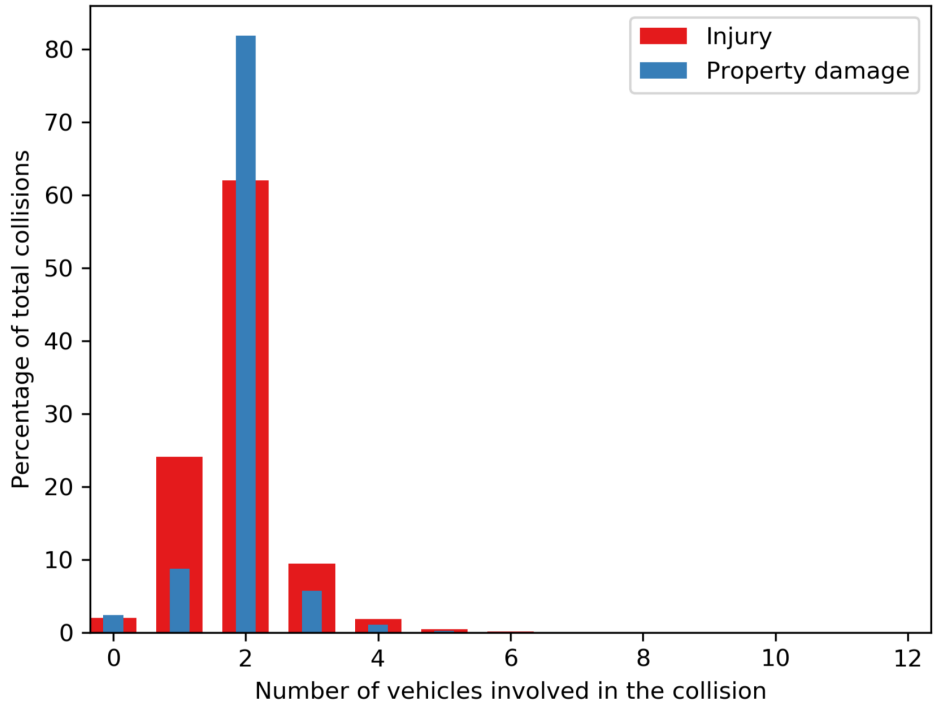
ROAD SEGMENT AND STREET ILLUMINATION



- 2/3 of Class I collisions happen on the block; for Class II it is only 1/2 of them.
- Most accidents happen during the day with small differences between the two classes.

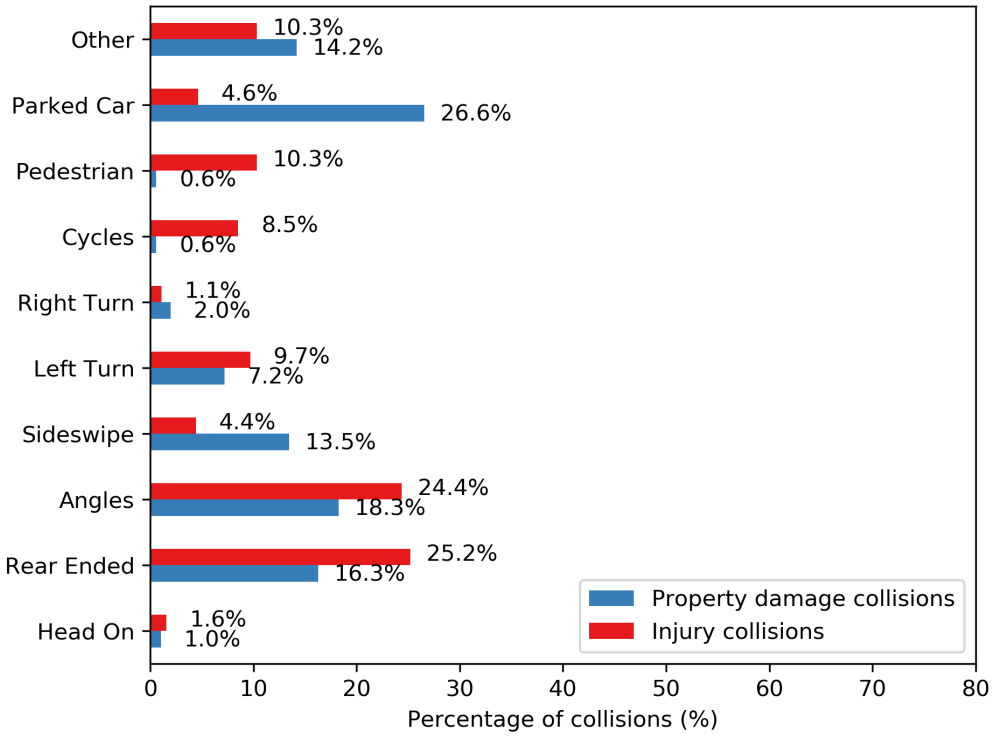
PARTICIPANTS IN THE ACCIDENT





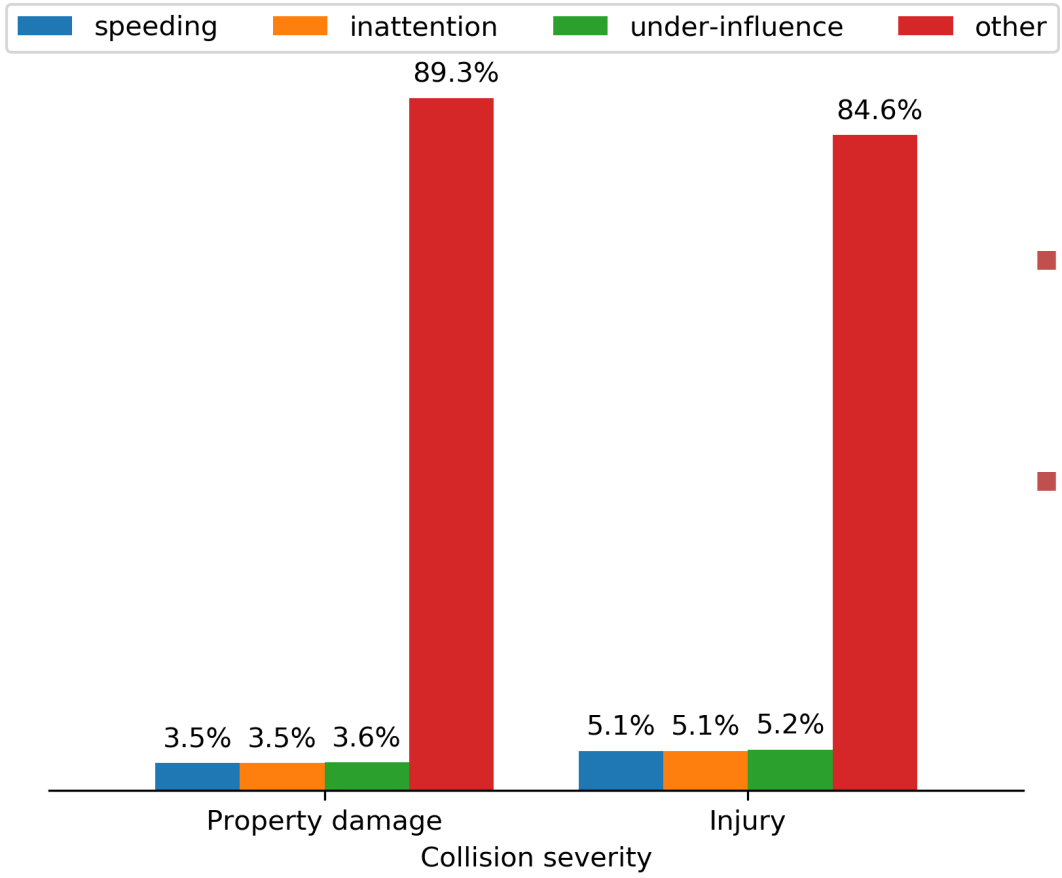
- The injury collisions are higher in proportion with:
  - more than 3 people involved;
  - 1 or more than 2 cars involved.

COLLISION TYPE



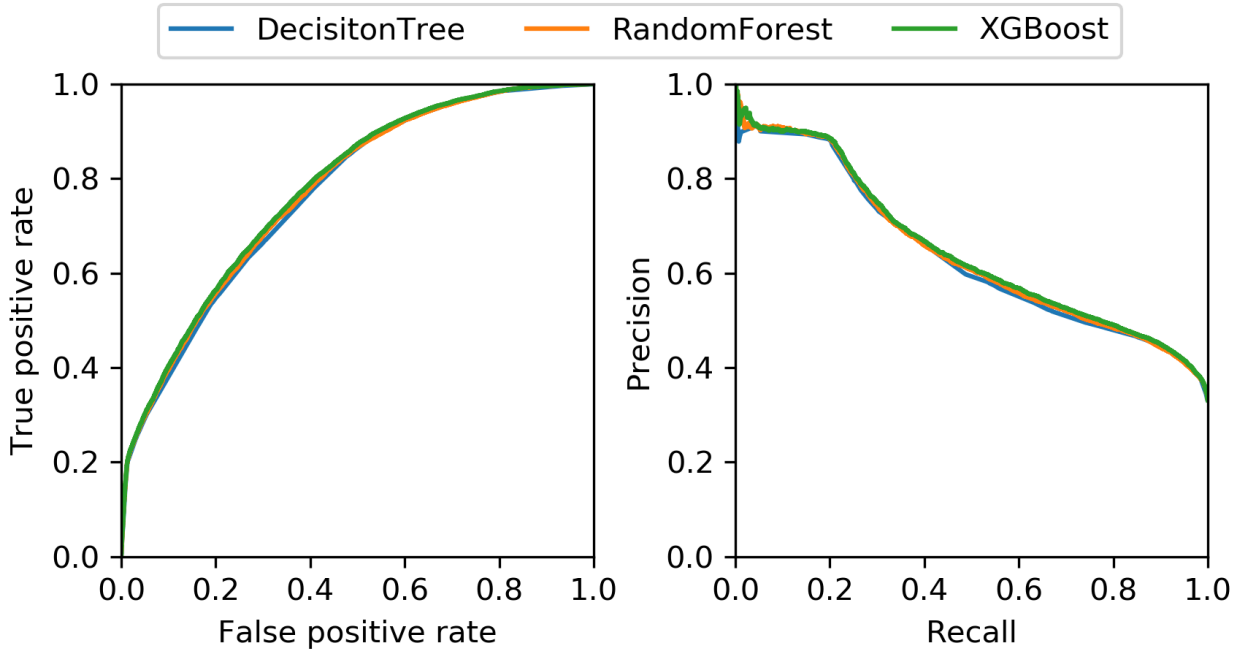
- Most Class I collisions happen:
  - with parked cars;
  - at angle/rear-end/sideswipe.
- Most Class II collisions happen:
  - at angles/rear-end;
  - with pedestrians/cyclists.

RECKLESS DRIVING



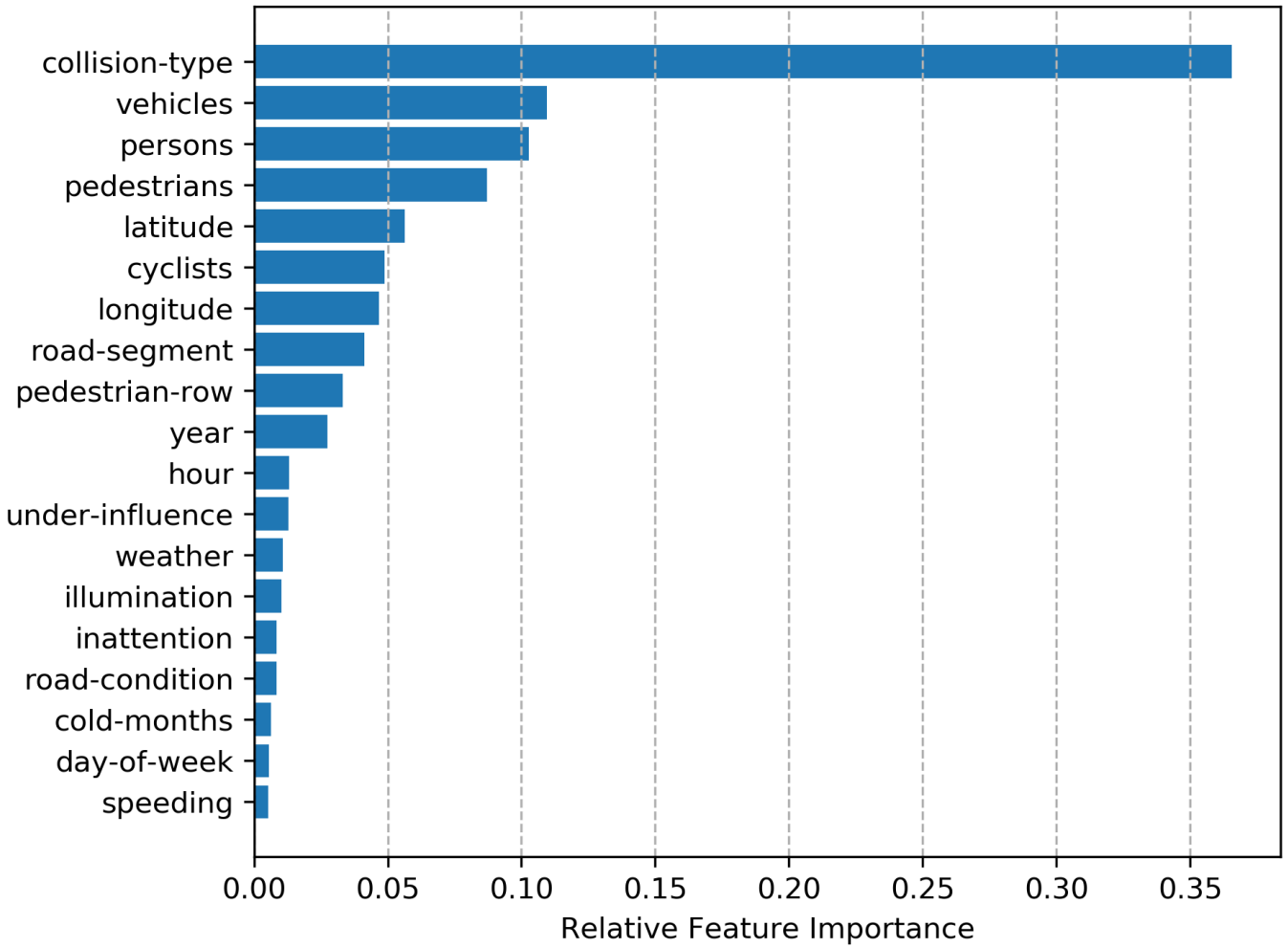
- Speeding, inattention while driving and being under the influence of drugs or alcohol are equally represented within each class.
- Collisions with injury have slightly more cases of reckless driving.

CLASSIFICATION MODELS



- Accuracy estimated as the area under the ROC curve.
  - Best value 0.70 for XGBoost.
- All three classification models give similar results.

FEATURE IMPORTANCE



- Accuracy estimated as the area under the ROC curve.
  - Best value 0.70 for XGBoost.
- All three classification models give similar results.

CONCLUSIONS, RECOMMENDATIONS, FUTURE WORK

- The model has a good accuracy, but may need to be improved.
- The most important feature is the collision type, which provides a classification in hindsight.

- Recommended preventive measures following the data analysis:
  - Supervise the observance of safety distance on the most affected road segments ;
  - Control the visibility of right-of-way road signs ;
  - Increase the safety of pedestrians and cyclists ;
  - Control/Improve the state, illumination and visibility of the zebra crossings and cycling paths.

- Future work includes improving the predictibility of the model by adding feaures on:
  - Traffic congestion and two-vehicle collisions ;
  - Availability of cycling paths and visibility of cyclists.