

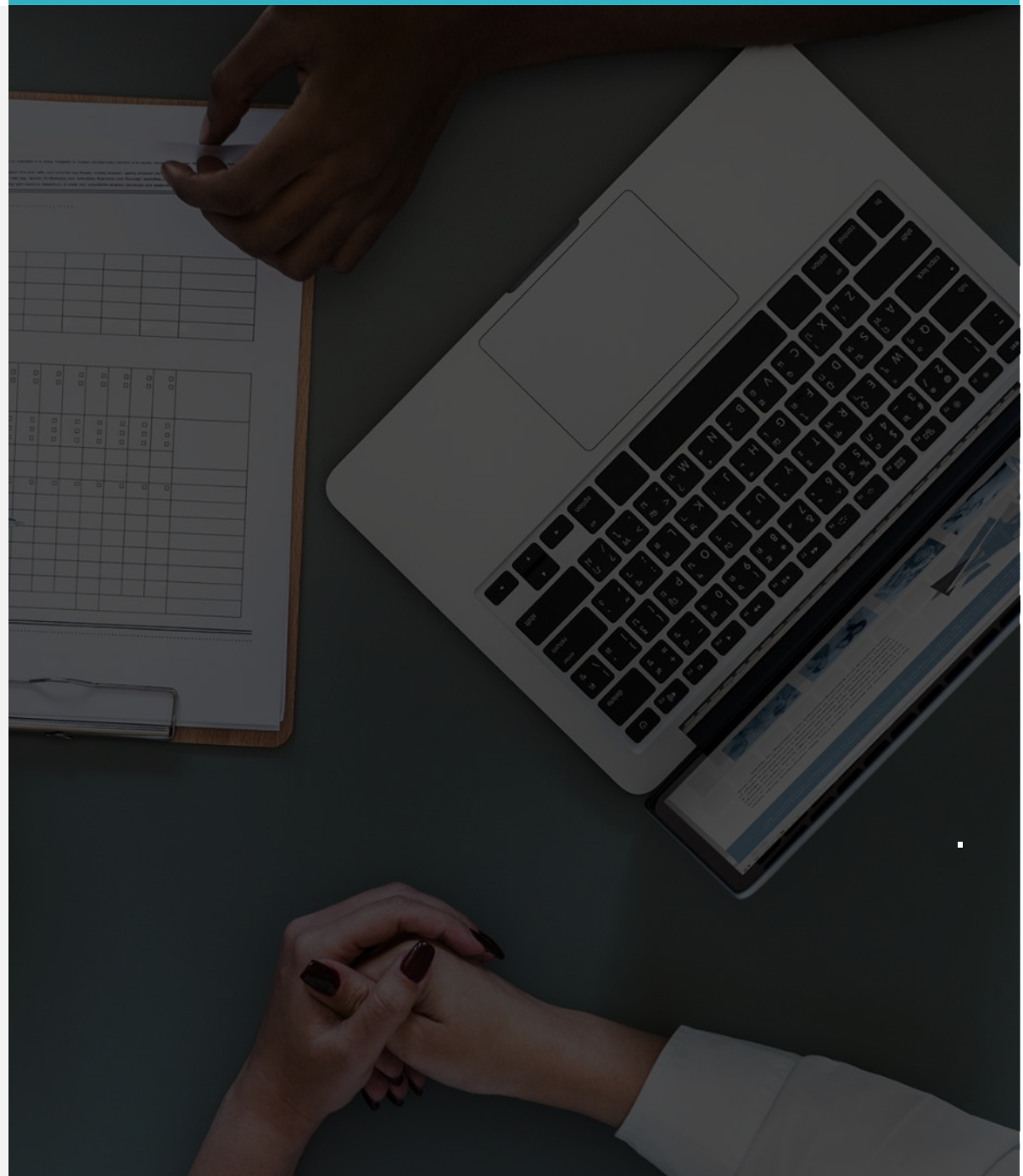


Luis Terán

Will you get injured
when driving accident
occurs?

Luis Terán.

- Motor vehicle accidents continue to be one of the leading causes of accidental deaths and injuries in the United States.
- More than 38,000 people die every year in crashes on U.S. roadways.
- The U.S. traffic fatality rate is 12.4 deaths per 100,000 inhabitants.
- An additional 4.4 million are injured seriously enough to require medical attention.
- Road crashes are the leading cause of death in the U.S. for people aged 1-54 (ASIRT, 2020).





Seattle, USA

The aim of this Project is to predict the probability of the severity of the accident even before a person starts the car.

How can this be solved?



- The original dataset was obtained from the Seattle's government page at: <https://data.seattle.gov/Land-Base/Collisions/gkas-rb8d>
- The dataset is available for public access. It was created in April 8, 2020 and last update register is from August 27, 2020. Consulted on August 28, 2020. Further information of the dataset is available in: https://www.seattle.gov/Documents/Departments/SDOT/GIS/Collisions_OD.pdf
- The dataset consists on a 220,937 car accident's observations in Seattle, USA with 40 different columns of different information.

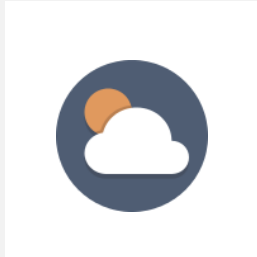
What data should we use for prediction?

For the prediction the related variables were:



Influence of alcohol
or drugs

Whether or not a driver involved was under the influence of drugs or alcohol



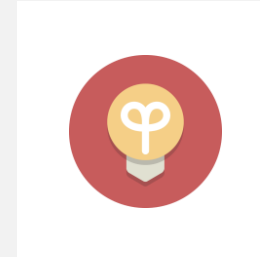
Weather

A description of the weather conditions during the time of the collision.



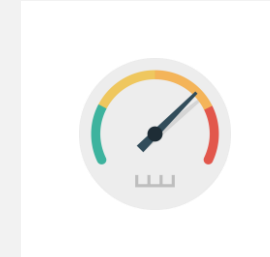
Road Conditions

The condition of the road during the collisions



Lighting
conditions

The light conditions during the collisions



Speed limit

Whether or not speeding was a factor in the collisions

SEVERITY

For measuring the severity of the data the number of involved vehicles, injuries, serious injuries and fatalities was added up with weights in order to create a variable measurement of the severity of the incident. Then, these measures were classified into different categories:

1

Type 1

No incident or
a few car
crashes with
no injuries or
fatalities

2

Type 2

One injured
person and
multiple car
crashes

3

Type 3

Multiple
Injuries or a
serious injured
person, with
multiple car
crashes

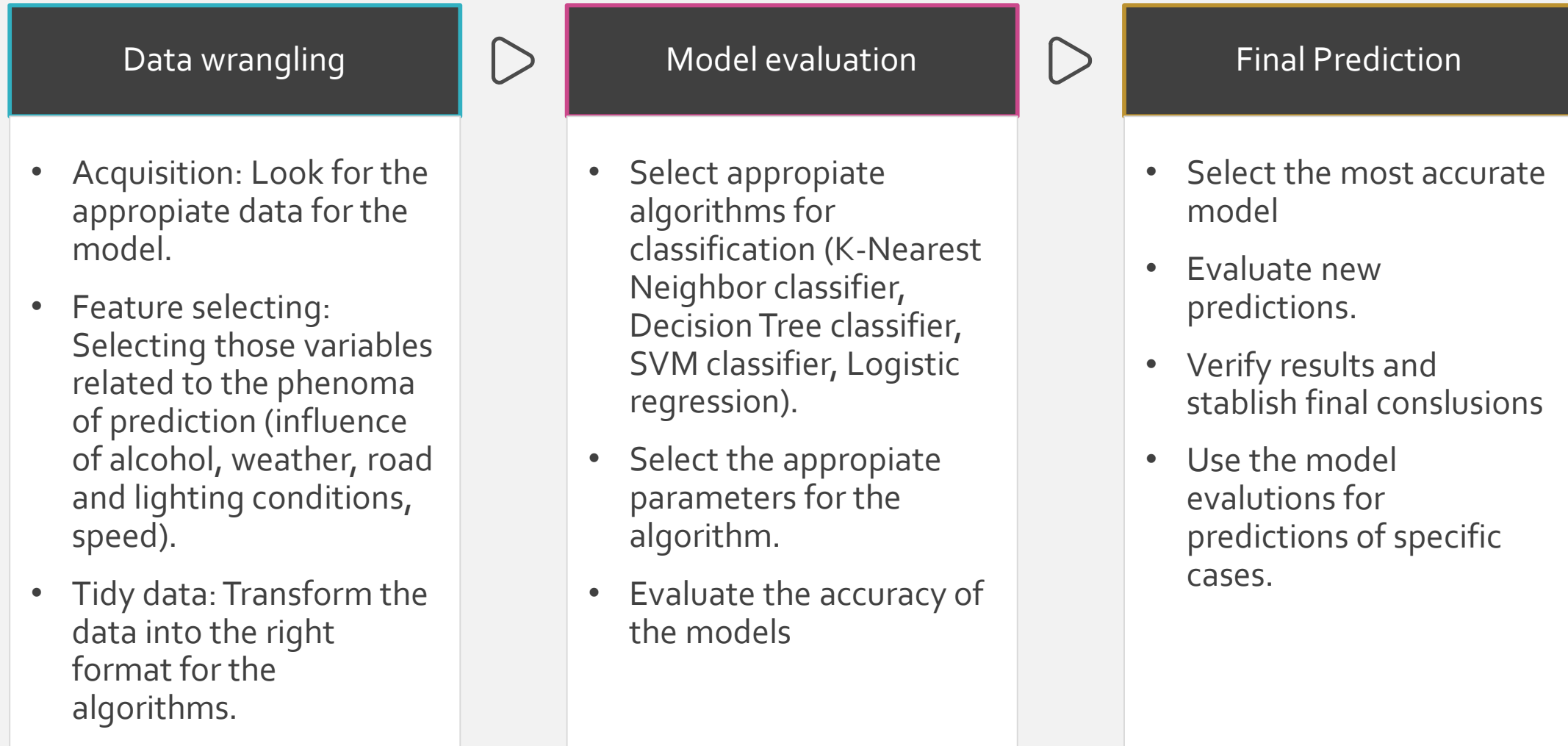
4

Type 4

A lot of injured
persons, more
than one
serious injured
person or a
fatality
occurred

Data processing

The project was developed as follows:



Conclusions

Decision Tree Classifier

From all the analyzed model, the decision tree was the best option for its accuracy, speed and interpretability

Features

The most significant features for classification, in order are:

1. Road conditions
2. Light conditions

3. Influence of alcohol or drugs
4. Weather

The input for prediction can be created from answering these answers

a.- ¿Are you drunk or under the influence of drugs?

- 1. Yes
- 2. No

b.- What's the weather like?

- 1. Clear
- 2. Raining
- 3. Overcast
- 4. Snowing
- 5. Fog/Smog/Smoke
- 6. Sleet/Hail/Freezing Rain
- 7. Blowing Sand/Dirt
- 8. Severe Crosswind
- 9. Partly Cloudy
- 10. Other/Unknown

c.- How would you define the road conditions?

- 1. Dry
- 2. Wet
- 3. Ice
- 4. Snow/Slush
- 5. Standing water
- 6. Sand/Mud/Dirt
- 7. Oil
- 8. Other/Unknown

d. What are the light conditions?

- 1. Daylight
- 2. Dark – Street Lights On
- 3. Unknown
- 4. Dusk
- 5. Dawn
- 6. Dark – No Street Lights
- 7. Dark – Street Lights On
- 8. Dark – Unknown Lighting
- 9. Other

e.- Are you in a hurry? Is it possible that you exceed the speed limit during your travel?

- 1. Yes
- 2. No

Predicting posible scenarios

Two posible scenarios were predicted. One for an optimistic one, and the other for a pessimistic one. So, in case of an accident, these are the probabilities:

Optimistic

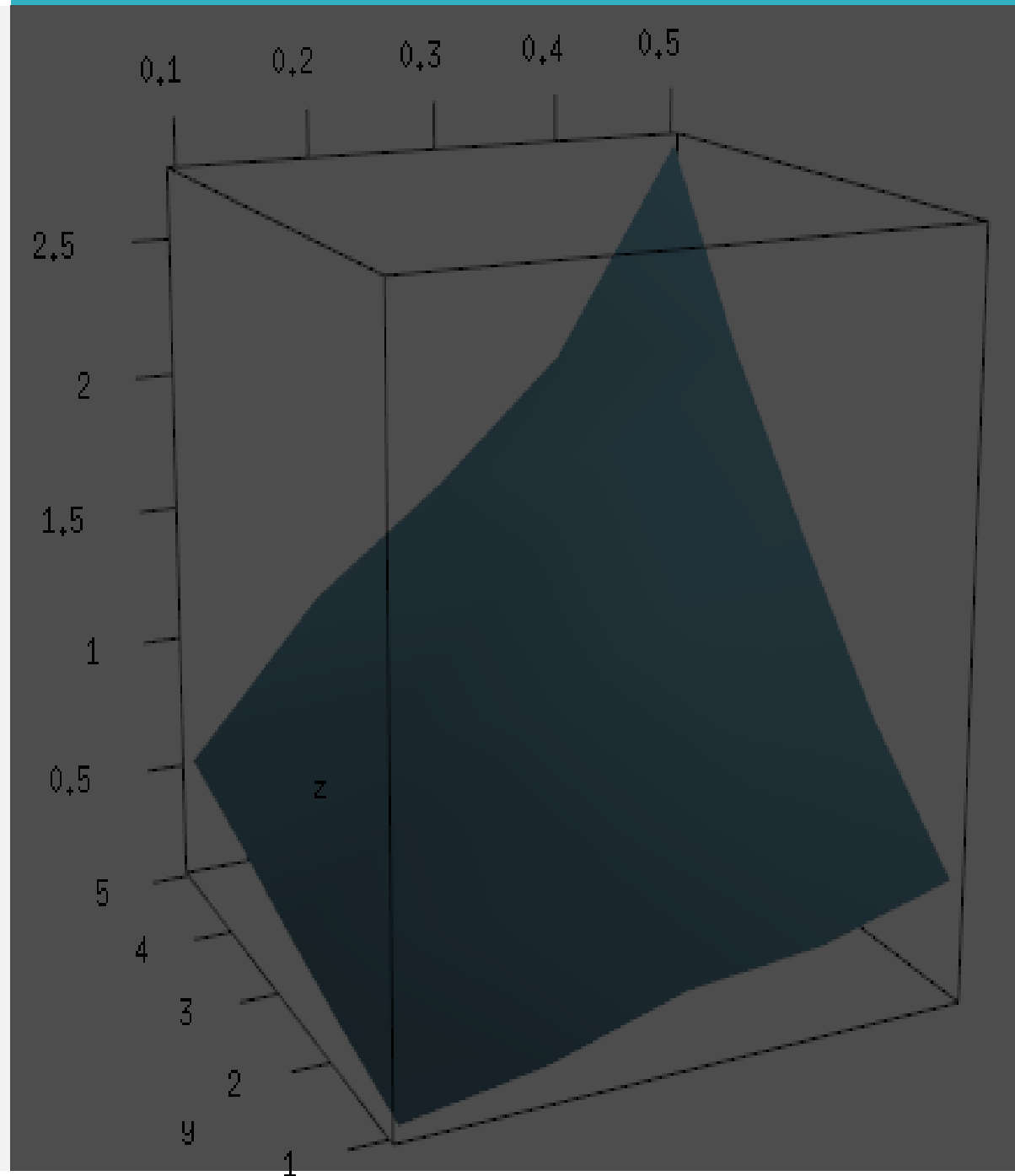
- Input: No alcohol or drug influence, Partly cloudy weather, Dry road conditions, Daylight and no speed limit exceeded.
- Output prediction:
- Type 1= 66.79%,
- Type 2= 24.21%
- Type 3=7.96%
- Type 4=1.03%

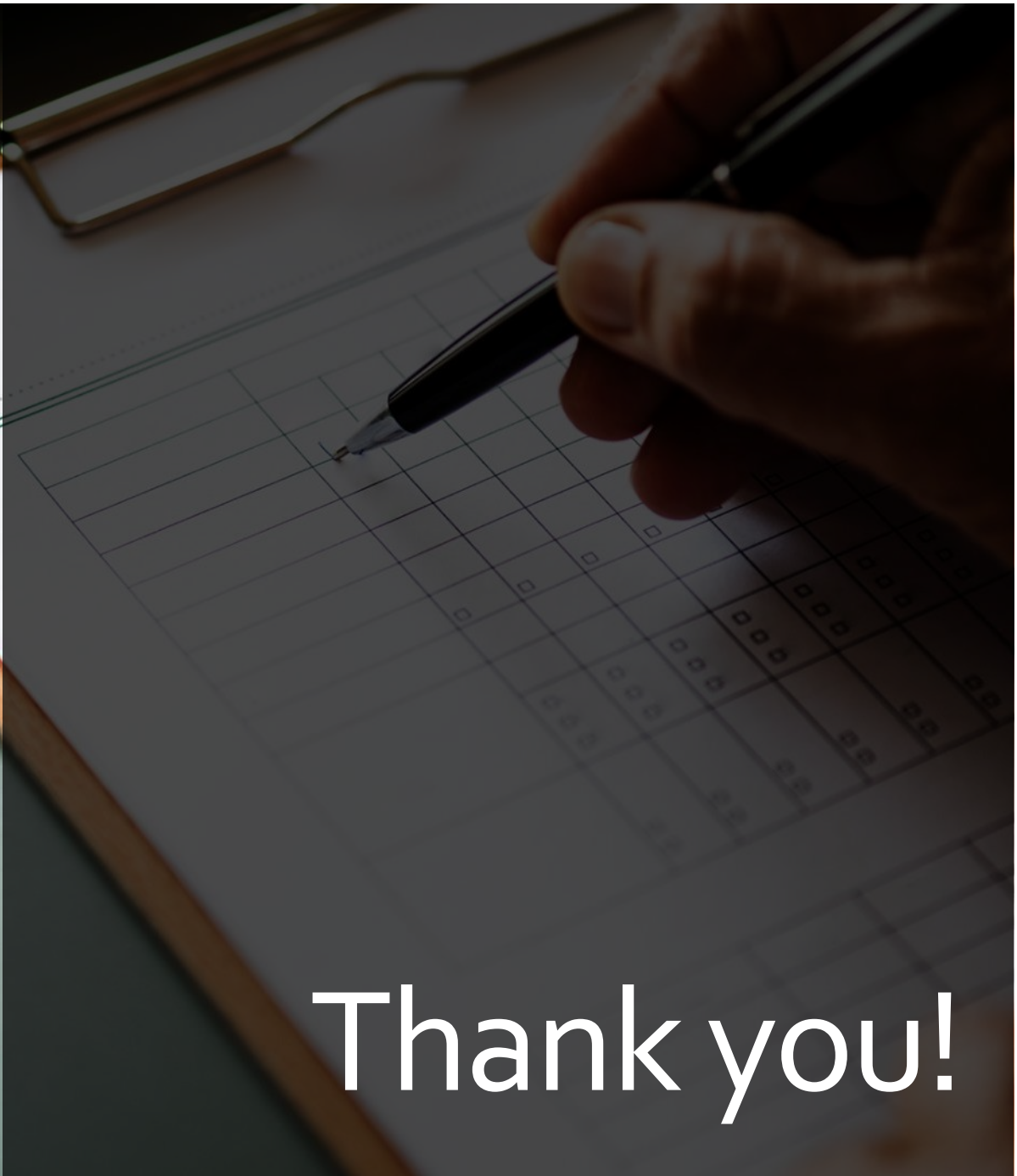
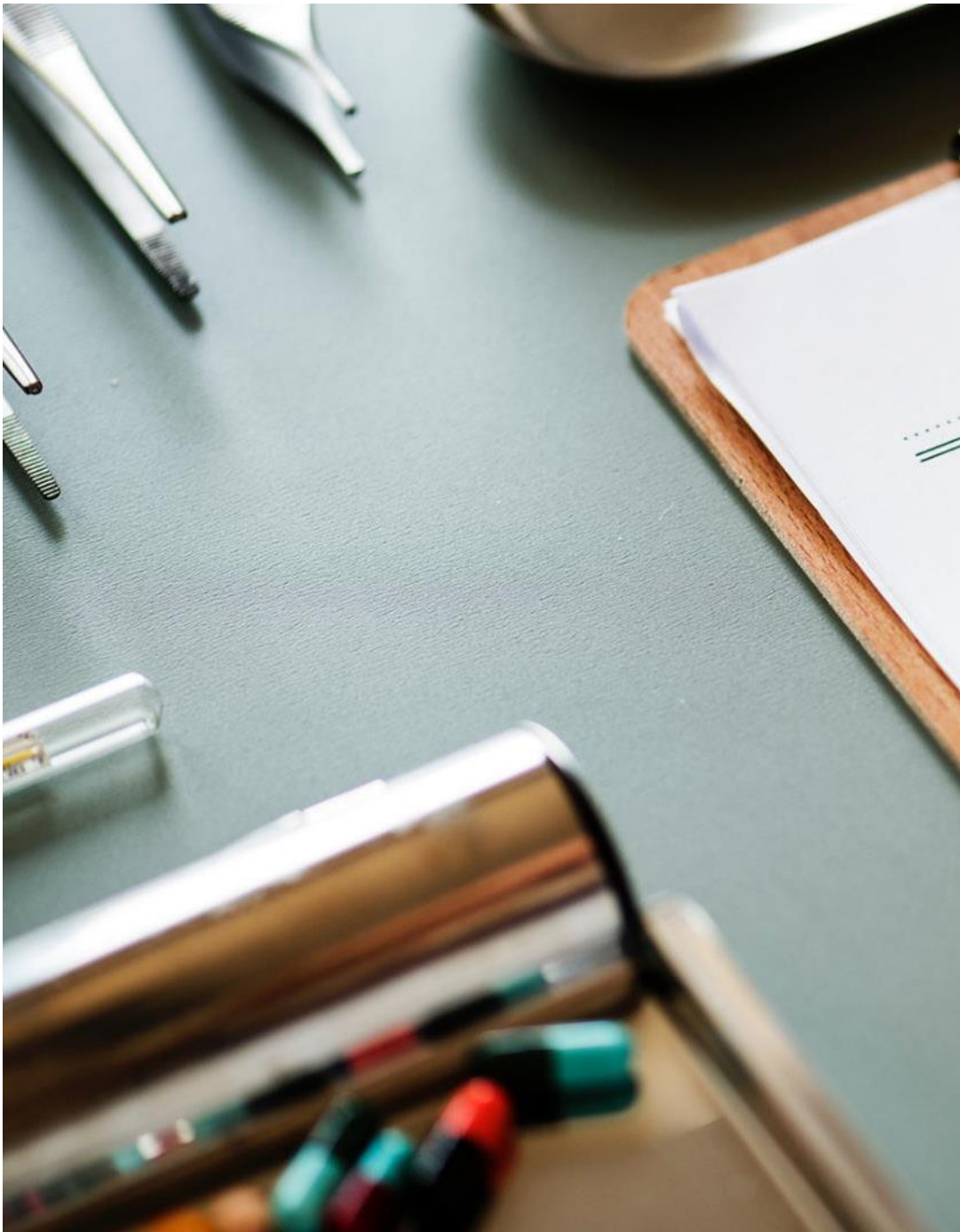
Pessimistic

- Input: Alcohol or drug influence, standing water road conditions, dark - street lights off, speed limit exceeded.
- Output prediction:
- Type 1= 45.67%,
- Type 2= 31.18%
- Type 3=15.84%
- Type 4=7.28%

Discussion

- Even though the prediction model gives reasonable results, the accuracy can be improved by changing different forms of measuring the severity of an accident since it's really hard to find a way of predicting the severity of a road accident.
- A better prediction can be made by getting more data. Also getting more data of no incident drives.
- Using a different scale for the severity of a possible accident, different to the created here, could help for the accuracy of the predictions and a better understanding of the data.





Thank you!