Final Capstone - Coursera

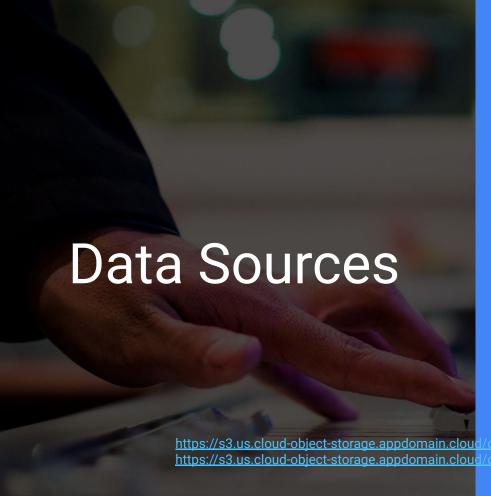
Data Science specialization

Presentation of the **Capstone Project Study collisions severity**

The problem

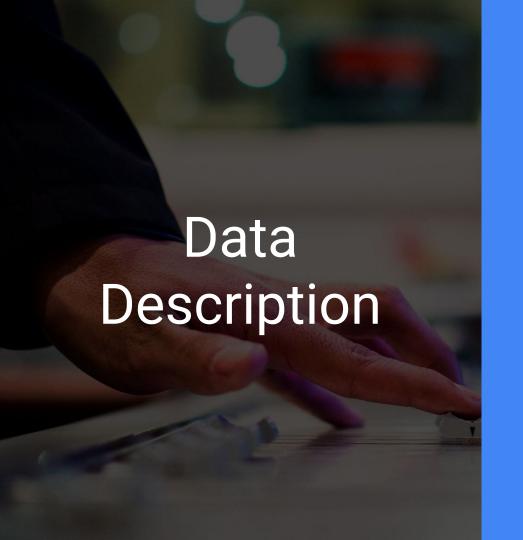
Car accidents are one of biggest causes of death among humans nowadays.

We want to automatically estimate the severity of every recorded car collision given some parameters



We will use the data provided by the course on coursera. it can be, along with the metadata downloaded from the links below.

https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0/01EN/version-2/Data-Collisions.csv https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf



This is a weekly updated data of all collisions provided by SPD and recorded by Traffic Records that includes all types of collisions displayed at intersection, mid-block or segment since 2004 until present

Project main steps

Expliquez à quelle étape du processus vous êtes arrivé et les défis que vous avez encore à relever

Business understanding Insérez votre texte ici

Insérez votre texte ici

Data

Preparation Insérez votre texte ici

Insérez votre texte ici

Evaluation

Insérez votre texte ici Insérez votre texte ici

Data

Requirement
Insérez votre texte ici

Insérez votre texte ici

Modeling

Insérez votre texte ici Insérez votre texte ici

Data Preparation

Our data preparation has 5 main steps:

- Feature Selection (Delete dependencies)
- Feature Encoding (Cat -> Num)
- Feature Formatting (Casting)
- Missing values cleaning (Mainly Dropping)
- Feature engineering (Adding meaningful feature from existing data)

Modeling and Evaluation

We used 4 well known Scikitlearn Models

- Random Forest (score : 0.719)
- KNN (score: 0.734)
- Decision Trees (score : 0.738)
- Logistic Regression (score: 0.672)

And then, we combined the results of these 4 models.

Final Score: 0.73

Conclusion

Some of our models overfitted, that's why we found an actual bad F1 score when we combined the 4 models.

We believe that model tuning is very important and preventing overfitting is even more important

Thank You

I would thank coursera and IBM for making this rich specialization available.

I would thank you dear peer reviewer for your time and your effort in making it until the end.

Congratulations for finishing your specialization

You may reach me on linkedin:

https://www.linkedin.com/in/samir-boua
ziz-9b36b6196/?originalSubdomain=dz

