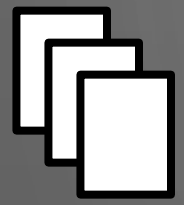
SCHNEIDER ELECTRIC HACKATHON

DATA SCIENCE CHALLENGE

Lydia Petridi

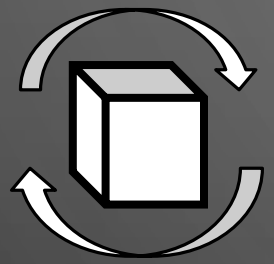
https://www.linkedin.com/in/lydia-petridi-65319222b/

PROCESS HIGHLIGHTS



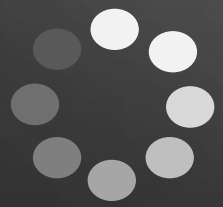
Loading and joining the datasets

- Joining the .csv and .json datasets together
- Intented to join the pdfs also, but was unsuccessful *(bad configuration of java in my PC made the task very time consuming)*



Data Transformation & Creation of the Model

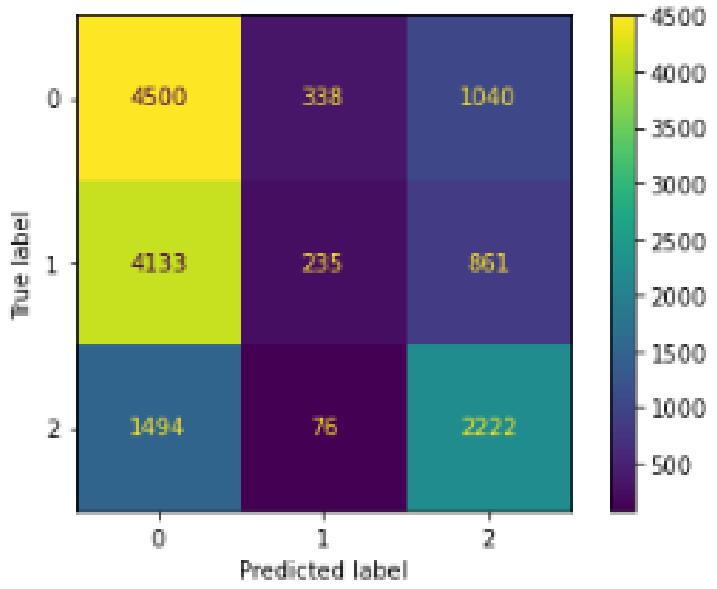
- Replacing strings and datatypes, checking for duplicates and correlations
- Deciding which atributes have more sence in remaining and dropping the rest
- Creating a logistic regression model and preparing the data for its training, testing and aplication



Evaluation & Predictions

- Evaluating the effectiveness of the model both in its initial form and after alterations
- Application of the final model on the data to predict
- Saving the results

MODEL EVALUATION



Classification Report: Logistic Regession

	precision	recall	f1-score	support
0	0.47	0.24	0.31	5878
1	0.44	0.48	0.46	5229
2	0.49	0.82	0.61	3792
accuracy			0.47	14899
macro avg	0.47	0.51	0.46	14899
weighted avg	0.47	0.47	0.44	14899

Class 0: Nitrogen oxides (NOX)
Class 1: Carbon dioxide (CO2)
Class 2: Methane (CH4)

CONCLUSIONS

- Model with poor efficiency and goodness scores.
- More alterations and better reconsideration about the relevant atributes is needed.

Final Thoughts

- Lost way too much time with the pdfs.
- I had a lot of fun doing it!!

Thank you!