Introduction

Discovery session

Data Science & Machine Learning
ITS4 & ISE2

Mously DIAW
ENSAE - 2021

Présentation

Mously DIAW

Experiences

Data Scientist / ML Engineer,
RCI Bank & Services

Mentor-évaluateur, OpenClassrooms

Data Scientist, JobTeaser

Educations

Master in Data Science
Polytechnique x Paris Saclay

ITS 2015, Stat & Info décisionnelle
ENSAE - Dakar

Présentez-vous

Course objectives

1. Translate a business need into a data science problem
2. Collect and prepare data for analysis
3. Build Machine Learning algorithms with Python
4. Model structured/complex data, such as images and text
5. Communicate results to specialists and neophyte
6. Create dashboards with relevant graphs and indicators
7. Conduct a research project and perform a proof of concept (POC)

Course outline

Machine Learning Overview

Python crash course

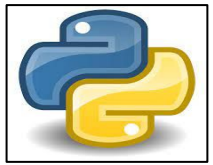
Data analysis & visualization

Data pre-processing

Problems categories

NLP / Deep Learning

General technical information



Python

Programming language

Jupyter lab / Google Colab

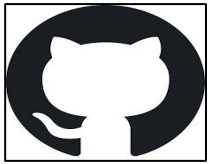
Web-based interactive computing platform

Anaconda / virtualEnv

Find and install packages

Manage Environments

Create isolated Python virtual environments



Github

Code hosting platform for version control and collaboration

Best Python libraries for ML

01

Data Visualisation

Matplotlib
Missingno
Seaborn
Pandas-profiling
Plotly
Bokeh
Streamlit
Dash
Flask
Folium
Geopandas
Geoplot

02

Data scraping / crawling

Scrapy
Selenium
BeautifulSoup
Request
LXML

03

Data preprocessing / Modeling

Pandas
Numpy
Scipy
Theano
Statsmodels
Scikit-learn
Yellow-brick
MLExtend
Xgboost
Lightgbm
CatBoost

04

Deep Learning / Transfer Learning

Keras
Pytroch
TensorFlow
MxNet
VGG
ResNet50
InceptionV3
Xception
Bert

05

NLP

NLTK
Gensim
SpaCy
TextBlob
Polyglot
FastText
GloVe
Pattern
CoreNLP

How to **validate the course** ?

- **Project** to be done
- 3 à 5 persons per **group**
- **Presentation** of the exploratory analysis & modeling done with me

*Each group will have the opportunity to have **mentoring sessions** with me, **every 2 weeks***

Resources

- Python distribution platform: anaconda
- Web Application: Python version: >= 3.6 or <= 3.9
- jupyter notebook or **jupyter lab**
- Code versioning: github
- IDE: **pycharm**, visual studio, sublime text

Contacts

- mouslydiaw@gmail.com
- linkedin
- Twitter
- WhatsApp (exchange group)
- Discord ??

Projects

	Description	Comments
1	Segmentation of the customers of an E-commerce site	
2	House prices prediction	
3	Digit recognizer	
4	Predict future sales	
5	Outbrain click prediction	
6	Prediction of Building energy	
7	Home credit default risk	
8	Predict Closed Questions on Stack Overflow	
9	Dogs vs cats recognition	
10	Movie recommendation	
11	Natural Language Processing with Disaster Tweets	

Reference #1: Data analysis / visualization

https://sunscrapers.com/blog/8-best-python-natural-language-processing-nlp-libraries/

https://www.datarevenue.com/en-blog/data-dashboarding-streamlit-vs-dash-vs-shiny-vs-voila

https://mode.com/blog/python-data-visualization-libraries/

https://medium.com/spatial-data-science/the-best-tools-for-dashboarding-in-python-b22975cb4b83

https://towardsdatascience.com/best-libraries-for-geospatial-data-visualisation-in-python-d23834173b35

https://towardsdatascience.com/top-6-python-libraries-for-visualization-which-one-to-use-fe43381cd658

https://mode.com/blog/python-data-visualization-libraries/

https://betterprogramming.pub/7-must-try-data-visualization-libraries-in-python-fd0fe76e08a0

https://analyticsindiamag.com/top-5-python-libraries-for-data-visualization/

https://geopandas.org/en/stable/gallery/index.html

Reference #2: Data scraping / crawling

https://it-s.com/top-5-python-web-scraping-libraries/

https://analyticsindiamag.com/top-7-python-web-scraping-tools-for-data-scientists/

https://elitedatascience.com/python-web-scraping-libraries

https://towardsdatascience.com/choose-the-best-python-web-scraping-library-for-your-application-91a68bc81c4f

https://medium.com/analytics-vidhya/best-python-libraries-to-perform-web-scraping-9b147bdc44f4

https://scrapingant.com/blog/top-5-popular-python-libraries-for-web-scraping-in-2020

https://www.analyticsvidhya.com/blog/2020/04/5-popular-python-libraries-web-scraping/

Reference #3: Data processing / modelling

https://towardsdatascience.com/how-to-build-your-first-machine-learning-model-in-python-e70fd1907cdd

https://machinelearningmastery.com/machine-learning-in-python-step-by-step/

https://towardsdatascience.com/top-10-python-libraries-for-data-science-cd82294ec266

https://scikit-learn.org/stable/

https://www.scikit-yb.org/en/latest/

https://ealizadeh.com/blog/mlxtend-library-for-data-science

https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/

https://openclassrooms.com/fr/courses/4452741-decouvrez-les-librairies-python-pour-la-data-science

https://openclassrooms.com/fr/courses/6389626-train-a-supervised-machine-learning-model/6398756-build-a-model-with-python

https://intellipaat.com/blog/top-python-libraries-for-machine-learning/

https://hackernoon.com/8-best-python-libraries-for-machine-learning-in-2021-mh48316l

Reference #3: Deep Learning

https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/

https://keras.io/api/applications/

https://pythonrepo.com/tag/transfer-learning

https://www.geeksforgeeks.org/best-python-libraries-for-machine-learning/

https://towardsdatascience.com/best-python-libraries-for-machine-learning-and-deep-learning-b0bd40c7e8c

https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

https://www.zenesys.com/blog/top-10-python-libraries-for-machine-learning

https://www.kdnuggets.com/2018/11/top-python-deep-learning-libraries.html

https://elitedatascience.com/python-deep-learning-libraries

Reference #3: NLP

https://analyticsindiamag.com/top-resources-to-learn-nlp-for-free/

https://realpython.com/nltk-nlp-python/

https://labsonmars.com/python-nlp/

https://labsonmars.com/python-nlp/

https://fasttext.cc/docs/en/supervised-tutorial.html

https://www.machinelearningplus.com/nlp/gensim-tutorial/

https://www.nltk.org/book/

https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html

https://www.kdnuggets.com/2021/09/best-resources-learn-natural-language-processing-2021.html

https://towardsdatascience.com/how-to-use-nlp-in-python-a-practical-step-by-step-example-bd82ca2d2e1e

https://towardsdatascience.com/10-free-resources-for-learning-natural-language-processing-81a380540cd1

https://www.analyticsvidhya.com/blog/2021/09/sentiment-classification-using-nlp-with-text-analytics/

https://www.analyticsvidhya.com/blog/2021/11/pattern-library-for-natural-language-processing-in-python/

https://www.analyticsvidhya.com/blog/2021/08/text-preprocessing-techniques-for-performing-sentiment-analysis/

Questions ???