**Python for Data Sciences and Machine Learning – Day 1**

**Instructor: Jean-Phillippe Forestier**

Zoom link to the course (recurring): <https://us06web.zoom.us/j/4757256633>

Jupyter Notebook to the course material: <http://localhost:8888/tree/Desktop/DSML_Day1>

Github to the raw datasets and the Jupyter material: <https://github.com/jpfnice/ds-ml-day1>

**Things we will cover/ Overview of the course topics:** Graphical user interface, text, application, email

Description automatically generated

1) Numpy array: used to represent matrix

2) Pandas Dataframe: two-dimensional collection; when you have to deal with excel or csv files. Represent to help in memory and will allow

3) Matplotlib

4) EDA (Exploratory Data Analysis) with Seaborn, Pandas, and Matplotlib

5) Feature engineering with numpy, scipy, scikit\_learn and Pandas

- Ways to convert string into numeric values (numpy)

- Ensure the data is well distributed

- normalizing the data

- Data has to be transformed to numpy array first and then have to be corrected for distribution of the data

**Different set of machine learning algorithms – models will predict depending on the data we are working with**

6) Machine Learning with scikit\_learn

7) Linear Regression

8) Kmeans

9) Scores

10) Decision Tree

11) Random Forest

12) SVM

13) Neural Network

14) Cross Validation

- Obtain a more accurate score

15) Feature Importance

- The scores will allow to improve the accuracy

16) Feature Selection

17) Pipeline

- It is fairly difficult to find a model that will initially give the best score/prediction

- Tool called Pipeline that will help in model assessment

- Test the model with different parameters and input data

18) Tensor Flow & tf.Keras

- Tensorflow is different from scikit-learn

- PyTorch is another distribution set up by Facebook

- TensorFlow is better suited for deep-learning whereas sci-kit learn is not that suited (requires more computational power which TF is better for)

- Neural Networks are DL algorithms that are better implemented by TensorFlow