**Projet P8 : Formation DATA SCIENTIST**

**Plan de travail**

Dans le cadre du projet 8 de la formation de Data Scientis il m’a été demandé de travailler sur un projet déjà traité mais en apportant une approche plus moderne.  
En effet le métier de Data Scientist évolue très rapidement et une écoute particulière des nouvelles méthodes doit être accordée.

Ceci dans le but d’obtenir toujours des algorithmes performants.

**J’ai choisi de reprendre le Projet 6 : Catégoriser automatiquement des questions**. C’est un projet qui m’avait particulièrement plu pour les raisons suivantes :

* Traitement du langage naturel
* Traitement de pré processing
* Pertinence des résultats obtenus

Dans cette seconde approche je souhaite m’orienter vers des modèles en réseaux de neurones.

* Dans un premier temps je vais développer un modèle simple qui servira de Baseline à l’aide de la libraire de Keras Tensorflow. Le turoriel se trouve à cette adresse <https://keras.io/examples/nlp/multi_label_classification/>
* Dans un second temps je vais mettre en œuvre les Transformers. Ce réseau de neurones est particulièrement adapté pour les traitements du langage car il travaille de séquence à séquence. Pour rappel, un modèle seq2seq est un modèle qui prend en entrée une séquence (une suite d’éléments du même type) et renvoie une séquence en sortie. <https://ledatascientist.com/a-la-decouverte-du-transformer/>