Rapport du stage effectué du 26/05/2019 au 26/07/2019

Dans la société : H2R



à Paris 9ème

Cameron Hochberg

étudiant en informatique concentration Intelligence Artificielle, systèmes informatique, réseaux et sécurité

Bachelor of Science in Computer Science minor in Japanese and Computer Gaming

2019

A l’université

Case Western Reserve University



Plan

Remerciments

Introduction

Les travaux effectués et les apports du stage

Conclusion

Remerciements

Je tiens à remercier Jacob Ouanounou, mon maître de stage qui m’a accompagné tout le long du projet en me conseillant sur les différentes approches sur l’intelligence artificielle ainsi que sur la généralisation de code.

Introduction

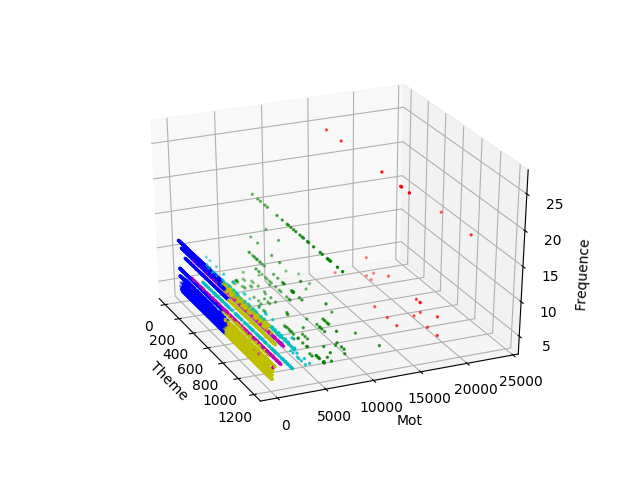
J’ai effectué un stage du 26 Mai au 26 Juillet 2019 au sein de l’entreprise H2R. Au cours de ce stage, j’ai pu m’intéresser au projet SIG, une app qui permet d’identifier concrètement des problèmes d’ordre psychologique. Cet outil est une avancée tant technologique qu’intellectuelle qui permettra aux utilisateurs de mieux comprendre leurs rapports avec les autres et la place qu’ils occupent dans leur environnement.

Plus largement, ce stage a été l’opportunité pour moi d’appréhender le monde de l’intelligence artificielle. J’y ai appris les différentes méthodes de classification ainsi que d’entraînement d’algorithmes.

Mon stage a essentiellement consisté à créer un prototype d’intelligence artificielle permettant d'identifier le thème d’un texte.

I/ Les travaux effectués et les apports du stage

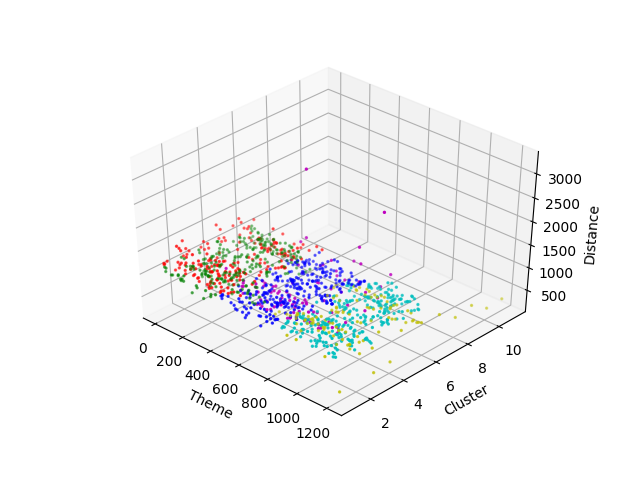
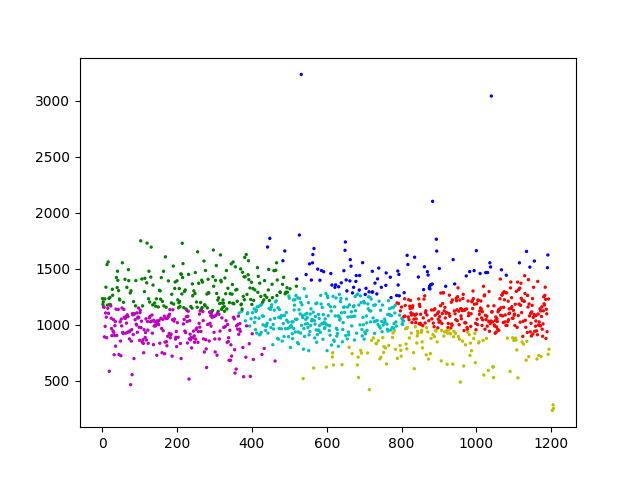
Au cours de ce stage, j’ai eu l’opportunité de découvrir une partie du monde de la recherche dans l’intelligence artificielle et de comprendre de manière globale les difficultés que l’on pouvait rencontrer au cours d’un projet aussi petit soit-il.



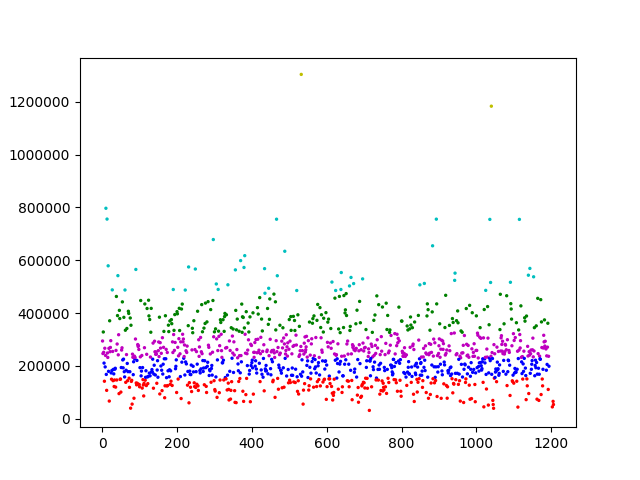
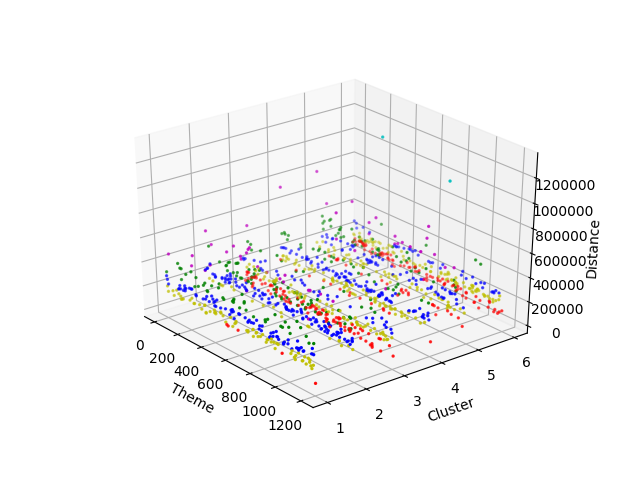
En premier lieu, j’ai essayé de trouvé un moyen d’utiliser la fonction KMeans pour classer des textes selon leur thème. Pour cela, j’ai d’abord utilisé la fréquence d’apparition ainsi que la variance de chaque mot de chaque texte pour essayer de trouver des groupes de thèmes communs sans succès:

Puis nous avons établi plusieurs façon de calculer la “distance” entre un texte et un thème. Cela nous a donné quelques résultats prometteur comme ci dessous.

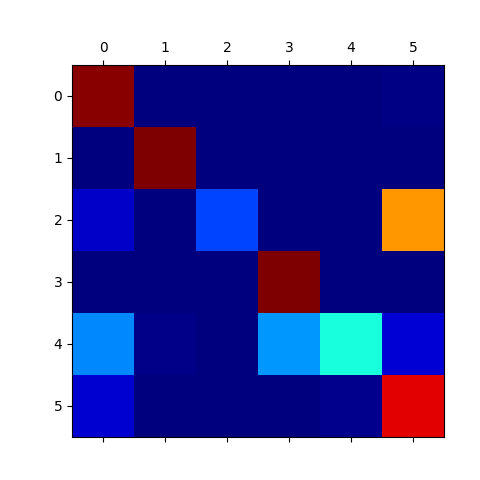
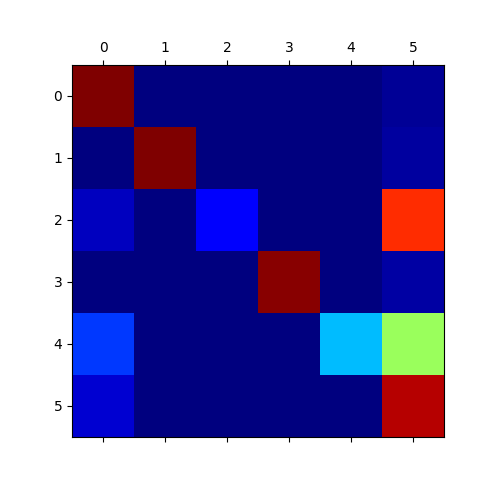
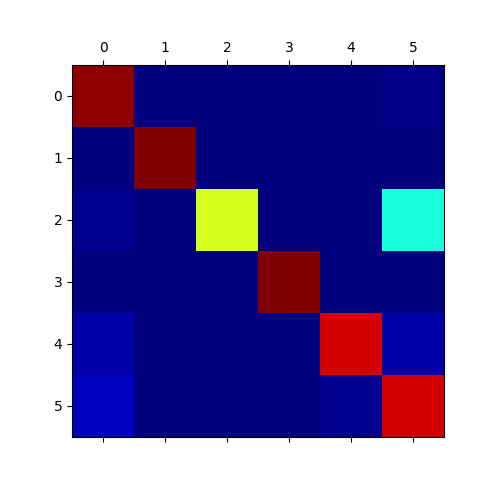
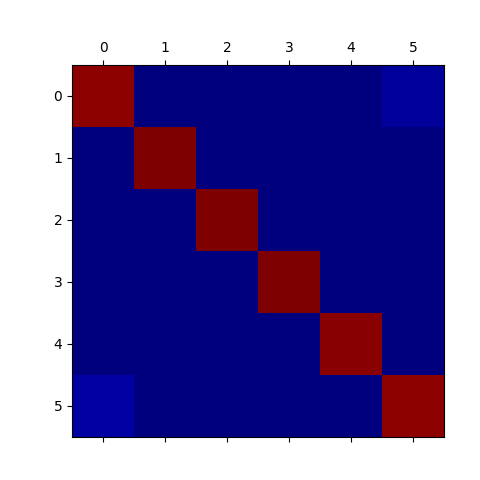
Première fonction de distance: où f(x) est la moyenne du mot x pour le texte analysé et fk est la moyenne de la fréquence de ce mot pour le groupe k.



Seconde fonction de distance



Après j’ai changé d’approche: au lieu d’essayer de regrouper chaque texte en un thème commun, j’ai essayé des algorithmes de classification. Ceux ci ont eu beaucoup plus de succès. En utilisant les algorithmes Support Vector Classification (SVC), Naive Bayes et Stochastic Gradient Descent (SGD), j’ai obtenu des matrices de confusion où notre but est d’obtenir une matrice diagonale:

Finalement, grâce à ces algorithmes, j’ai pu prouver que ce que nous voulions faire en classant des textes en fonction de leur thème était possible. 

Conclusion

Pour conclure, au cours de ce stage, j’ai réussi à prouver que nous pouvions classer des textes en fonction de leur thème en utilisant différents algorithmes de classifications.