

Compte rendu personnel de :  
**BUZURA Darius**

Manager actuel : **CAILLAUD Maxime**

## Projet Finance GPT - Compte rendu n°1

du 06/11/2023 au 17/11/2023

Tâches effectuées	Durée
README.MD	45min
Rendu 4 : Récupération du jeu de données (et tokenizer les données)	2h30
Création du modèle pour le compte rendu personnel	1h

### Objectif sur cette période :

- Réaliser un Readme.MD de projet de qualité : J'ai souhaité rendre le readme.md agréable à la lecture car c'est la description de projet et c'est important pour une bonne compréhension
- Récupération de jeu de données : Pendant le cours j'ai eu un peu de mal avec cette partie (surtout la partie normalisation et tokenizer les données). J'ai compris le concept mais j'ai eu du mal avec la mise en place de code. En souhaitant m'améliorer j'ai souhaité m'en occuper de cette partie

### Tâches accomplies

- **Tâche 1 : Réalisation de Readme.MD**

J'ai rédigé le fichier [Readme.MD](#) sur github. En effet ce fichier est la couverture de notre projet sur github et il est important qu'il contienne des informations claires et il sera mis à jour avec l'avancée de projet. Pour le rédacteur j'ai repris des parties de la consigne fournie, j'ai rédigé le document au format markdown, avec des liens hypertexte et des logos.

- **Tâche 2 : Récupération du jeu de données**

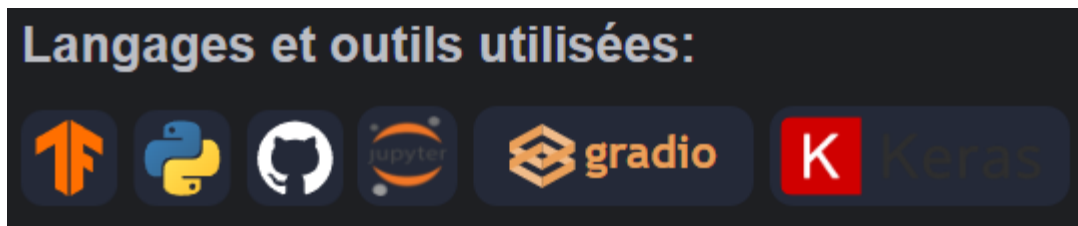
J'ai récupéré le jeu de données [financial\\_phrasebank\\_dataset](#). Le modèle GPT 2 ne peut pas prendre ce jeu de données tel qu'il est car il est au format texte, et le modèle ne peut pas prendre du texte en entrée, il prend des vecteurs. Donc j'ai tokeniser les données pour les rendre au format adapté pour être prises en charge par le modèle.

- **Tâche 3 : Création du modèle pour le compte rendu personnel**

Au sein de l'équipe nous avons décidé de créer ce modèle de rendu personnel pour avoir tous le même support d'appui, et le même document à compléter et cela allait nous être très utile tout au long de ce projet. Je n'ai pas rencontré de difficultés sur cette tâche.

## Difficultés rencontrées :

- Difficulté 1 : Tokenizer les données
  - Cela à été la plus grande difficulté que j'ai rencontré durant cette période. En effet j'ai eu du mal à trouver le bon tokenizer et la bonne manière de s'y prendre. J'ai eu plusieurs tentatives, je me suis aussi aidé des notes prises durant la dernière séance de révisions, qui m'ont permis de comprendre ce que je dois faire. J'ai
  - J'ai beaucoup consulté la doc de hugging face [Open AI GPT2](#), mais aussi celle de [Keras GPT2 Tokenizer](#). J'ai essayé de comprendre à quoi sert le tokenizer, à quoi ressemblent les données que je dois traiter, en j'ai repris les cours et les explications vues en cours, j'ai repris mes notes.
- Difficulté 2 : Les logos dans le Readme.MD
  - J'ai utilisé une bibliothèque de logo, mais cette bibliothèque ne contenait pas les logos de keras, gradio, jupyter. Comme j'ai souhaité les intégrer, j'ai cherché une alternative.
  - Pour résoudre le problème j'ai cherché les logos sur internet. Ensuite j'ai donné un style grâce au css pour mettre le même arrière plan et la même forme que les autres logos (comme dans la photo ci-dessous)



## Conclusion :

Je trouve que durant cette période j'ai atteint mon bon niveau de productivité. Le fait de plonger dans un nouveau concept qui est l'ia, cela à boosté ma motivation, et m'as motivé à chercher et à me documenter sur ce que je ne savais pas, et surtout d'exercer pour mieux comprendre.