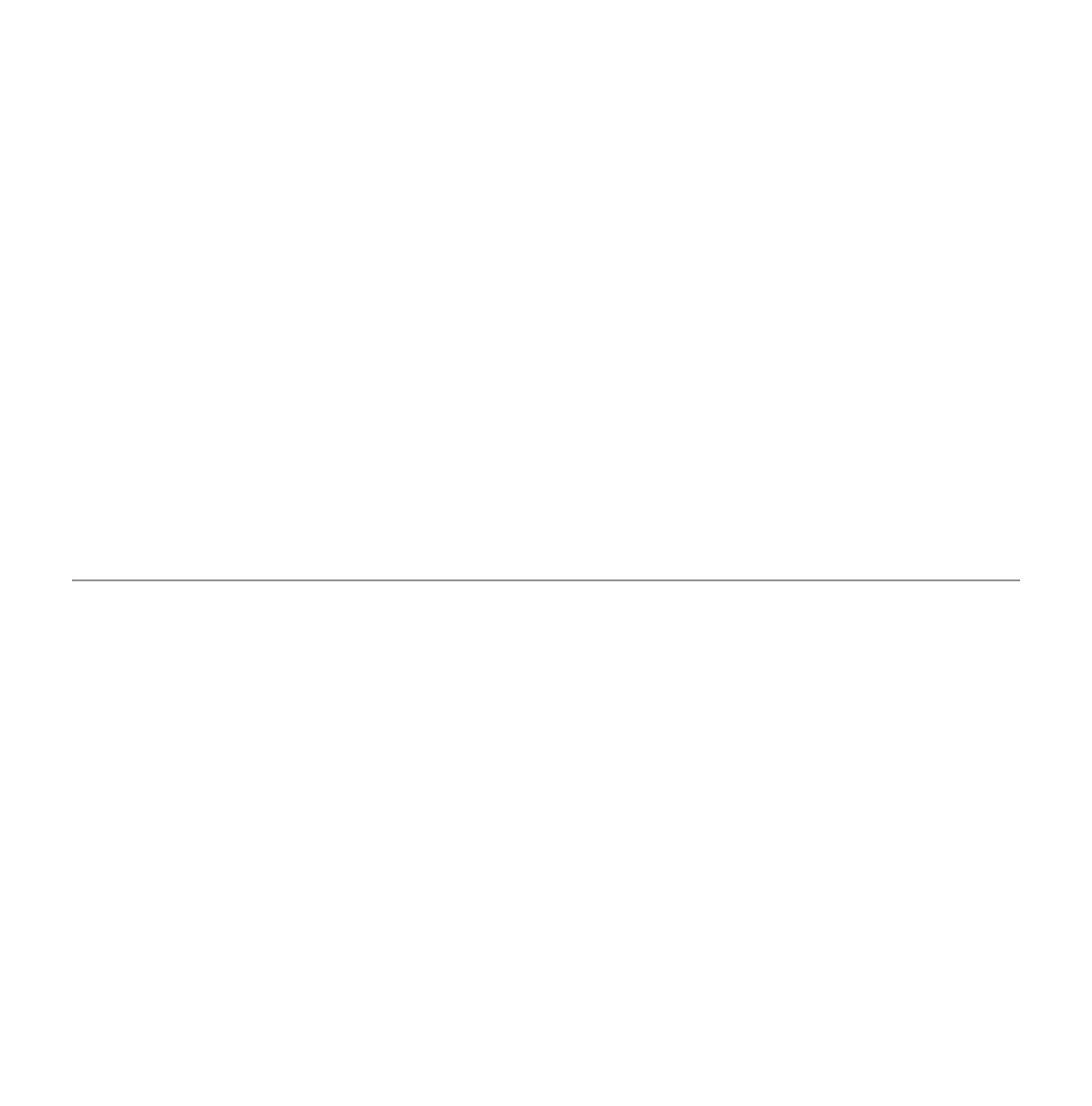
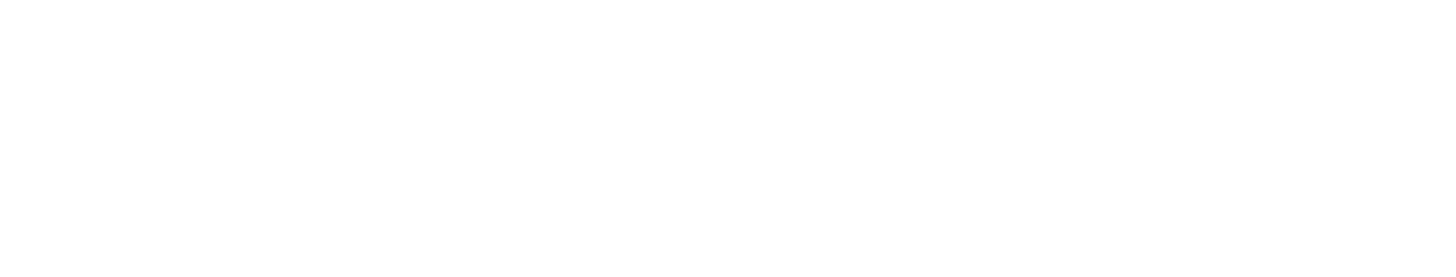
# Une image contenant Police, horloge, symbole, logo Description générée automatiquementEfrei Paris



Projet Python

My first chatBot



Maxime Avocat

Pratheep Parthepan

GROUPE D.

|

ANNEE 2023

-

2024

Table des matières

1. Page de garde
2. Table des matières
3. Introduction
4. Présentation fonctionnelle du projet
5. Présentation technique du projet
6. Présentation des résultats
7. Conclusion

Introduction

Le projet que nous vous proposons porte sur l'analyse de texte. Ce projet vous permettra de comprendre quelques concepts de base utilisés dans le traitement de texte et de vous aider à comprendre l’une des méthodes utilisées pour développer des chatbot et/ou des intelligences artificielles génératives telles que : chatGpt.

Il est évident, qu’il ne s’agit pas ici de manipuler des réseaux de neurones, mais nous allons dans ce projet nous focaliser sur la méthode basée sur le nombre d’occurrences des mots permettant de générer des réponses intelligentes à partir d'un corpus de textes. L'objectif est de concevoir un système qui peut répondre à des questions en se basant sur la fréquence des mots dans le corpus.

Présentation fonctionnelle

Cette application repose sur l’algorithme de base suivant :

Prétraitement des données : votre programme commence par collecter et prétraiter un ensemble de documents afin de comprendre la nature de leurs contenus avant de s’en servir lors de l’élaboration des réponses. Cette phase permet de nettoyer le texte en supprimant la ponctuation, en convertissant les lettres en minuscules, et en divisant le texte en mots (ou "tokens").

Création d'une matrice TF-IDF : Pour chaque mot unique dans les documents, vous serez amenés à calculer un vecteur TF-IDF en utilisant la méthode TF-IDF. Chaque mot est associé à un vecteur de dimension égale au nombre de documents dans le corpus. Cela crée une matrice TF-IDF où chaque ligne représente un mot et chaque colonne représente un document.

Représentation des questions: Lorsqu'une question est posée, le chatbot effectue le même prétraitement sur la question. Ensuite, il calcule un vecteur TF-IDF pour cette question en utilisant le même vocabulaire que les documents. Le vecteur de la question est de la même dimension que les vecteurs associés aux mots du corpus.

Calcul de la similarité : Le chatbot calcule la similarité entre le vecteur de la question et les vecteurs des mots du corpus à l'aide de la similarité cosinus ou d'une autre mesure de similarité. Cela lui permet de déterminer quels mots du corpus sont les plus similaires à la question.

Sélection de la meilleure réponse : Le chatbot identifie les mots du corpus les plus similaires à la question en fonction de leur score de similarité TF-IDF. Ensuite, il sélectionne la réponse qui contient le plus grand nombre de ces mots similaires.

Fournir la réponse : Le chatbot renvoie la réponse sélectionnée comme réponse à la question posée.

Présentation technique

Principaux algorithmes : fonction tf, fonction idf, fonction tfidf

Présentation des résultats

Nous n’avons pas obtenu de résultat avec pour les 2 dernières questions après des problèmes rencontrés à la 5ème question.

Conclusion

Le projet était intéressant et nous aurions aimé avoir plus de temps pour en venir au bout.