Projet chef d’œuvre

Détection des troubles de la santé mentale et du besoin de traitement avec l'IA chez les gens qui travaillent dans le domaine de la technologie.

Geoffroy Daumer

**Introduction du projet**

Un résumé de la demande exprimée par le client.

● Le contexte.

● la formulation de la problématique.

● La solution souhaitée/envisagée.

Résumé de la réponse finale apportée.

**Analyse de la demande**

Les enjeux réglementaires pour le traitement des données de l'application.

Les utilisateurs du projet.

Le schéma fonctionnel utilisé pour le projet - *screenshot ou via un applicatif.*

Un planning prévisionnel. Une présentation du macro-planning initial.

**Développer la gestion des données analytiques d’un projet d’application**

La documentation des sources de données identifiées nécessaires au projet, en fonction des résultats de l’analyse exploratoire.

Les visualisations synthétisant les caractéristiques du ou des jeux de données - *screenshot ou via un applicatif.*

La stratégie de nettoyage des données.

Les éléments de modélisation des données - *schémas, requêtes, etc.*

La justification du ou des systèmes de gestion de bases de données analytique adaptés à la modélisation des données.

La procédure de mise en place du système de gestion de base de données analytique.

**Développer un programme d'intelligence artificielle**

Un exemple de résultat obtenu à l’issue de l’exécution du script de récupération et de préparation des données - *screenshot ou via un applicatif.*

La justification du choix de l’algorithme (par exemple de machine learning ou de deep learning) au regard des objectifs du projet et des données

disponibles,

La justification des outils adaptés à l’implémentation de l’intelligence artificielle.

La formalisation, dans le cas d’un algorithme de machine learning ou de deep learning, de la méthode d'entraînement et/ou de surentraînement.

La méthode d’alimentation de l’intelligence artificielle avec les données disponibles. Justifier, dans le cas du machine learning, la méthodologie utilisée

pour le partitionnement des données de test et d’entraînement.

L'explication du fonctionnement général du programme final fonctionnel de l’intelligence artificielle.

Les résultats des tests de performance.

La justification de la méthodologie de mise à disposition et d’interaction avec le programme d’intelligence artificielle - *Une API REST dans la plupart des cas.*

L’explication du fonctionnement général du programme permettant l’interaction entre le programme d’intelligence artificielle et l’application visée.

**Développer une application d’intelligence artificielle**

Les éléments (documents, schémas…) de formalisation des fonctionnalités de l’application.

Le schéma de l’architecture de l’application - *screenshot ou via un applicatif.*

La modélisation de la ou des structures des données sous forme de schémas standards pour les données opérationnelles - *MCD, MPD* - *screenshot ou via un*

*applicatif.*

L’identification et la justification du ou des systèmes de stockage adaptés à la modélisation des données.

La procédure de mise en place du système de gestion de base de données.

Présentation de la bonne gestion des droits et des accès à l’application - *ACL, sécurisation, etc.*

L’identification des éléments et parties critiques de l’application à monitorer.

La procédure de débogage de l’application - *messages de débogage, usage des tests automatisés, analyse et compréhension des messages d’erreur, etc.*

**Gérer un projet de développement d’application**

La présentation du planning - *screenshot ou via un applicatif.*

La présentation d’un ou des compte rendu d’avancement.

Un retour d’expérience sur la gestion du projet et de la méthode des gestion adoptée.

Table des matières

[1. Introduction du projet 5](#_Toc161698032)

[Compréhension du besoin client 6](#_Toc161698033)

[Présentation des données 6](#_Toc161698034)

[Etat de l'art 6](#_Toc161698035)

[5. Etat de l’art 7](#_Toc161698036)

[Références 17](#_Toc161698037)

[Annexe 18](#_Toc161698038)

# 1. Introduction du projet

Nous vivons dans une société où les cas de problèmes de santé mentale sont en forte croissance, et ils restent à ce jour mal soignés.

Un client m'a contacté pour développer une application qui intégrerait l'IA, pour permettre de détecter le besoin de traitement relatif aux problèmes de santé mentale chez les travailleurs du domaine de la technologie.

Ce client a mené une enquête auprès des travailleurs, leur posant un certain nombre de questions à choix multiples, et recueillant les résultats sous forme d'un jeu de données.

Parmi les questions posées, une question "Avez-vous recherché un traitement pour des problèmes de santé mentale ?", c'est cette question qui servira de cible à l'algorithme d'IA.

La solution finale apportée est une application sur laquelle on peut répondre au questionnaire du client, et où un algorithme d'IA prédit en conséquence si nous devrions considérer le fait de prendre un traitement pour la santé mentale.

# Compréhension du besoin client

Avec l'IA, on pourrait détecter plus facilement les troubles de la santé mentale, qui sont en permanente croissance.

De plus, les travailleurs du domaine de la technologie sont souvent assis, derrière des écrans. Ce manque d'activité physique, voir sociale, peut être un facteur d'un bien être diminuant.

# Présentation des données

Le jeu comprend 1259 lignes et 27 colonnes, il y a donc 1259 questionnaires remplis.

Une image contenant capture d’écran, texte, noir

Description générée automatiquementIl y a une colonne "Timestamp", une colonne "Age" (valeur continue), "Genre", "Country", "state" (pour les États américains), le reste des colonnes c'est des questions à choix multiples, allant de 2 à 5 environ. Il y a une colonne "Comments" pour les commentaires.

# Etat de l'art

(1)Pour réaliser la veille sur les algorithmes à utiliser, il faut chercher du côté des modèles de classification, la prédiction est une catégorie : Besoin d'un traitement ou non.

De plus, la grande majorité du jeu de

# 5. Etat de l’art

L’extraction de données sur du texte se fait majoritairement avec des expressions régulières dans les articles scientifiques. C’est une méthode relativement simple à déployer. Cependant, elle se limite rapidement avec la diversité du texte et la complexité des éléments à extraire. Quant aux algorithmes de deep learning, ils nécessitent d’annoter une grande quantité de données.

Jusqu’à maintenant, il n’y a pas vraiment de méthodologie développée pour ces applications. On remarque une absence de tâches communes qui permettraient de comparer différentes méthodes de développement, ainsi qu’une absence de métriques communes.

De nouvelles pipelines basées sur des modèles pré-entraînés comme GPT-3 pourraient résoudre en partie le problème du manque de données d’entrainement

Expressions régulières :

Exemple d’application :

Recherche du groupe de chiffres après « volume télé-systolique » dans un compte rendu. Pour connaitre ce volume, qui sera classé dans une colonne appropriée.

Ou, recherche de plusieurs motifs : exemple « cmd / cardiopathie dilatée / cardiomyopathie dilatée », si l’un d’eux est présent : mettre « oui » dans la colonne « cmd » (cardiomyopathie dilatée)

* Les expressions régulières sont très pratiques et obtiennent de bonnes performances, elles peuvent notamment extraire des entités textuelles, là où des modèles complexes doivent intervenir dans l’IA (réseaux de neurones, LLMs).

# Références

1. ***https://datascientest.com/algorithme-de-classification-definition-et-principaux-modeles.***

# Annexe