MSPR 2

Plan

1. **Introduction : Analyse des Besoins**

* Contexte général : Origine et objectif du projet (prévision des pandémies à l'aide de l'IA).
* Description des enjeux : Prévention, sécurité internationale, accessibilité pour les chercheurs non techniques.
* Présentation des besoins fonctionnels :
  + API IA prédictive (taux de transmission, mortalité, propagation géographique).
  + Interface utilisateur interactive et ergonomique.
* Contraintes identifiées :
  + Intégration fluide avec le backend existant.
  + Accessibilité (standards WCAG).
  + Scalabilité et sécurité.

2. **Méthodologie**

* **Collecte et Préparation des Données** :
  + Sources des données historiques sur les pandémies.
  + Approches de nettoyage et structuration.
* **Paramétrage de l'Environnement Technique** :
  + Choix du Framework et des outils (TensorFlow, PyTorch, ou autres).
  + Préparation des environnements de développement collaboratifs.
* **Modélisation et Codage** :
  + Justification du modèle choisi (réseaux de neurones, forêts aléatoires, etc.).
  + Description des architectures implémentées.
* **Entraînement du Modèle** :
  + Définition des stratégies d’entraînement (supervisé, semi-supervisé, non supervisé).
  + Sélection et traitement des données d'apprentissage.
* **Tests et Évaluation** :
  + Mise en place de la procédure de test.
  + Analyse des performances des prédictions avec métriques clés.

3. **Développement de l'Interface Utilisateur**

* Conception des éléments visuels interactifs et responsifs.
* Explication des bibliothèques utilisées (D3.js, Chart.js, Plotly).
* Stratégies pour l'accessibilité et ergonomie.

4. **Conclusion : Résultats, Discussion, Perspectives**

* Présentation des résultats obtenus (précision du modèle, performance).
* Points forts et limitations identifiés.
* Bilan du projet (objectifs atteints).
* Perspectives pour améliorer ou étendre les fonctionnalités (ex : nouveaux modèles de prédiction, compatibilité API).

5. **Annexes**

* Documentation technique (paramètres du modèle, extraits de code clés, etc.).
* Références bibliographiques ou outils utilisés.

INTRODUCTION ET ANALYSE DES BESOINS

Ce projet s'inscrit dans une démarche visant à développer une solution technologique avancée basée sur l'Intelligence Artificielle (IA) pour analyser et prévoir des phénomènes complexes à partir de données structurées. L'objectif principal est de concevoir une API IA capable de prédire des indicateurs clés tels que le taux de transmission, la mortalité et la propagation géographique à partir de données historiques préalablement nettoyées et normalisées.

Le développement technique repose sur plusieurs axes stratégiques. Tout d'abord, l'API doit être conçue pour être facilement intégrée dans l'infrastructure backend existante, en respectant les exigences de scalabilité et de sécurité. Une documentation exhaustive accompagnera cette API pour faciliter son utilisation et son déploiement. L'interface utilisateur, développée en parallèle, devra permettre une interaction fluide avec le modèle prédictif via des visualisations de données dynamiques et des fonctionnalités de filtrage, en exploitant des bibliothèques modernes. Le respect des normes d'ergonomie et d'accessibilité, notamment WCAG, est également primordial.

Sur le plan du traitement des données, le projet nécessitera la mise en œuvre d'un pipeline robuste, incluant la collecte, le prétraitement, et l'adaptation des données au modèle d'apprentissage. La sélection des architectures de Machine Learning les plus adaptées (réseaux de neurones, forêts aléatoires, etc.) sera cruciale pour garantir la précision et la pertinence des prédictions. Enfin, des stratégies d'entraînement supervisé, semi-supervisé et non supervisé seront définies, accompagnées d'une procédure de test rigoureuse pour analyser et optimiser les performances du modèle.

Ce projet constitue une opportunité unique de combiner des solutions algorithmiques et des interfaces modernes dans le cadre d'une application concrète et impactante. Chaque étape, du choix des outils au développement des fonctionnalités, contribuera à la réalisation d'une plateforme IA performante et évolutive.

I Méthodologie

Le développement de ce projet repose sur une démarche méthodique, structurée en étapes clés, qui couvrent l’ensemble du cycle de vie du développement du modèle d’apprentissage et de l’application associée.

1. **Collecte et Préparation des Données**

Dans un premier temps nous reprenons les données que nous avons nettoyé et utilisé lors de la première partie du projet, pour rappel ces données concernent le VIH à travers le monde. Lors du premier nettoyage nous avons choisi de supprimer les données les moins fiables ou manquantes, afin de garantir une utilisation optimale des sources dont nous disposons. Cette première organisation nous a permis d’avoir à disposition des données facilement utilisables en sélectionnant uniquement les informations dont nous avons besoin.

2. **Configuration de l’Environnement Technique**

Nous avons choisi de continuer à utiliser les technologies que nous avons déjà employé dans la première partie du projet, à savoir Python afin d’implémenter la partie API IA Prédictive. En plus de l’usage de FastAPI, que nous avions utilisé pour mettre en place l’API de l’ETL, nous avons utilisé joblib, sklearn et matplotlib afin d’entrainer le modèle et de créer une visualisation. Nous utilisons aussi github afin de transmettre et partager les avancées et nouveautés concernant le projet au sein de l’équipe.

3. **Conception et Développement du Modèle d’Apprentissage**

La phase de conception et de développement du modèle d’apprentissage repose sur une sélection des algorithmes, en tenant compte des besoins liés à la prédiction. Pour aatteindre nos objectifs, nous avons envisagés différents modèles de Machine Learning, comme les réseaux de neurones, les forêts aléatoires ou les approches de boosting .   
Nous essayons de garder une approche qui respecte les meilleurs pratiques du domaine. Notre objectif en optimisant des hyperparamètres clés, comme la profondeur des arbres, le taux d’apprentissage, ou encore la structure des couches d’un réseau de neurones est d’atteindre un équilibre optimal entre biais et variance.

La gestion des données d’apprentissage est également un aspect fondamental de notre travail. Nos données seront utilisées en deux jeux distincts, un jeu d’entrainement et un jeu de test afin d’obtenir des mesures objectives et évaluer notre modèle afin de le guider au mieux vers un apprentissage qui prévienne les problèmes de surapprentissage.

4. **Entraînement et Évaluation du Modèle**

L'entraînement du modèle repose entièrement sur une procédure d'apprentissage supervisé, adaptée à la nature des données entièrement étiquetées disponibles. Cette approche permet de maximiser la pertinence des prédictions en exploitant des ensembles de données clairement définis pour chaque classe ou catégorie cible.

(métriques de performances : classes prédictives, AUC-ROC, F1 score ) Quelqu’un sait ce que c’est ? peut rajouter un mot à ce propos ? ou on en parle juste pas ?

Des ajustements progressifs des hyperparamètres seront effectués en fonction des résultats obtenus, avec pour objectif d’améliorer le taux d’apprentissage et de minimiser l’erreur globale. Ces ajustements s’appuieront sur des techniques comme la validation croisée, qui permettront de sélectionner les paramètres les plus appropriés tout en réduisant le risque de surapprentissage. Ce processus itératif garantit un modèle performant, fiable et capable de généraliser correctement sur des données non vues.

5. **Développement de l’Interface Utilisateur**

Le développement de l'interface utilisateur repose sur une approche moderne et interactive, utilisant Vue.js comme framework principal pour le front-end. Vue.js a été choisi pour sa flexibilité, sa capacité à gérer efficacement les composants dynamiques, et sa facilité d'intégration avec les bibliothèques de visualisation de données comme D3.js, Plotly et Chart.js. Ces outils nous permettent de créer des visualisations interactives, offrant une analyse intuitive des prédictions et des données.

Nous avons mis en place une interface responsive afin qu’elle puisse être utilisées sur le plus grand nombre de support afin de garantir une accessibilité maximale. Une attention particulière a été portée à l’ergonomie et au respect des standards WCAG (Web Content Accessibility Guidelines) afin de fournir une interface inclusive, ce qui inclut par exemple le choix de contrastes et typographies optimisés.

Cette combinaison de technologies modernes et de bonnes pratiques garantit une expérience utilisateur fluide et intuitive, tout en répondant aux exigences techniques et fonctionnelles du projet.

6. **Phase de Test et Validation**

La phase de test et validation est essentielle pour garantir la qualité et la fiabilité de l'API et de l'interface développées. Les tests unitaires seront réalisés sur chaque composant de l'API afin de vérifier leur bon fonctionnement individuel, tandis que les tests d’intégration se concentreront sur la cohérence globale et l’interaction entre les différents modules du système. Cette approche permet d'identifier rapidement les éventuelles incohérences ou dysfonctionnements.

Parallèlement, le recueil de retours utilisateurs sera effectué en simulant des scénarios d’utilisation représentatifs. Ces tests sur le terrain offriront une vision pratique de l’efficacité du système et permettront de relever les points à améliorer, notamment en matière de performances ou d’expérience utilisateur. Une attention particulière sera portée à la facilité d’accès et d’utilisation, en cohérence avec les attentes des utilisateurs finaux.

Enfin, des ajustements finaux seront réalisés en fonction des résultats des tests et des retours utilisateurs. Ces optimisations permettront de répondre aux besoins identifiés, tout en renforçant la robustesse et la pertinence du système pour son utilisation à grande échelle.

II **Développement de l'Interface Utilisateur**

* 1 Conception des éléments visuels interactifs et responsifs.
* 2 Explication des bibliothèques utilisées (D3.js, Chart.js, Plotly).
* 3 Stratégies pour l'accessibilité et ergonomie.

III. **Conclusion : Résultats, Discussion, Perspectives**

* Présentation des résultats obtenus (précision du modèle, performance).
* Points forts et limitations identifiés.
* Bilan du projet (objectifs atteints).
* Perspectives pour améliorer ou étendre les fonctionnalités (ex : nouveaux modèles de prédiction, compatibilité API).

Cette méthodologie garantit une approche rigoureuse et itérative, permettant de développer une solution IA performante, évolutive et adaptée à ses usages finaux