**Cahier des Charges Détaillé : Prédiction des Pandémies et Gestion Proactive des Épidémies**

**1. Contexte et Objectifs**

**Contexte :**  
Les pandémies récentes, telles que la COVID-19, ont mis en évidence l'importance cruciale d'une réponse rapide et prédictive pour réduire les impacts sanitaires, économiques et sociaux. Le projet vise à développer un modèle prédictif basé sur des données historiques et en temps réel afin de fournir des estimations précises sur l'évolution des pandémies.

**Objectifs généraux :**

* Développer un modèle de prédiction des pandémies.
* Identifier les zones à haut risque de propagation.
* Proposer des stratégies proactives pour limiter les impacts.

**Objectifs spécifiques :**

* Analyser les données historiques des pandémies et des maladies infectieuses.
* Prévoir les tendances futures basées sur des facteurs climatiques, démographiques et comportementaux.
* Développer une plateforme interactive pour visualiser les prédictions et recommander des actions.

**2. Analyse des Besoins**

**Cas d’usage principaux :**

1. **Systèmes de santé publique** : Prévoir les besoins en ressources médicales.
2. **Prise de décision gouvernementale** : Planification des mesures de confinement ou vaccination.
3. **Alerte précoce** : Identifier les premiers signes de propagation.

**Sources de données :**

* **Données historiques** : Cas confirmés, taux de mortalité, durée des pandémies précédentes.
* **Données en temps réel** : Rapports journaliers des infections, mobilité des populations.
* **Facteurs environnementaux** : Température, humidité, pollution.
* **Facteurs démographiques** : Densité de population, âge moyen, accès aux soins.

**3. Cahier des Charges Fonctionnel**

**Fonctionnalités principales :**

1. **Collecte et nettoyage des données** :
   * Collecter des données à partir de bases publiques (OMS, CDC, etc.).
   * Gérer les données manquantes et éliminer les anomalies.
2. **Modélisation prédictive :**
   * Utiliser des algorithmes avancés (Réseaux Neuronaux Récurrents, Modèles ARIMA).
   * Identifier les corrélations entre les différents facteurs.
3. **Validation et évaluation :**
   * Diviser les données en ensembles d’entraînement, validation et test.
   * Évaluer la précision avec des métriques telles que RMSE, F1-Score.
4. **Visualisation et interaction :**
   * Développer une interface utilisateur intuitive pour visualiser les prédictions.
   * Intégrer des alertes en temps réel basées sur les seuils de risque.

**4. Cahier des Charges Technique**

**Technologies recommandées :**

* **Langages de programmation** : Python (Scikit-learn, TensorFlow, PyTorch) ou R.
* **Outils d’analyse** : Pandas, NumPy, Matplotlib, Seaborn.
* **Bases de données** : PostgreSQL, MongoDB, ou API publiques (ex : Johns Hopkins).
* **Environnement de développement** : Jupyter Notebook, PyCharm.
* **Plateformes de déploiement** : Flask/Django pour une application web.

**Spécifications techniques :**

* **Capacité de traitement** : Manipuler de grands ensembles de données.
* **Performances** : Optimisation des modèles pour un traitement rapide des prédictions.
* **Sécurité** : Assurer la confidentialité des données sensibles.

**5. Livrables**

1. **Modèle de prédiction** :
   * Algorithmes prédictifs optimisés et validés.
2. **Rapport d’analyse** :
   * Documentation sur les données, les algorithmes utilisés et les performances.
3. **Interface utilisateur** :
   * Application pour visualiser les prédictions et recevoir des alertes.
4. **Guide utilisateur** :
   * Documentation pour l’utilisation et la maintenance du système.

**6. Planning et Échéancier**

**Durée estimée :** 45 jours  
**Phases principales :**

1. **Semaine 1 :** Collecte et préparation des données.
   * Identifier les sources de données fiables.
   * Nettoyer les données et créer des variables pertinentes.
2. **Semaine 2 :** Analyse exploratoire des données.
   * Identifier les patterns et corrélations dans les données.
   * Sélectionner les algorithmes préliminaires.
3. **Semaine 3-4 :** Développement du modèle prédictif.
   * Implémenter et entraîner les modèles.
   * Optimiser les hyperparamètres pour améliorer la précision.
4. **Semaine 5 :** Validation et évaluation.
   * Tester le modèle avec des données de validation.
   * Comparer les performances avec des benchmarks.
5. **Semaine 6 :** Déploiement et documentation.
   * Intégrer le modèle dans une application web.
   * Rédiger la documentation technique et utilisateur.

**7. Budget estimatif**

* **Ressources humaines** : Data scientists, développeurs.
* **Outils et logiciels** : Licences pour outils premium (si nécessaire).
* **Infrastructure** : Serveurs cloud pour le traitement des données.

**8. Conclusion**

Ce projet permettra d’améliorer la préparation et la réponse face aux pandémies, tout en utilisant des technologies de pointe pour un impact global. Il s’agit d’une opportunité unique de contribuer à la santé publique mondiale et à la prévention des crises sanitaires.