

Titre du projet : Création d'un modèle de prédiction basé sur des données historiques

1. Contexte et Objectifs

1.1. Contexte

Le présent projet vise à concevoir et implémenter un modèle prédictif capable de fournir des estimations ou des prévisions en se basant sur des données historiques. L'objectif est de développer un outil efficace et précis pour répondre à des besoins spécifiques d'analyse et d'aide à la décision.

1.2. Objectifs

- Développer un modèle de prédiction utilisant des algorithmes de machine learning, capable de traiter des données historiques et de générer des prédictions fiables.
- Collecter, nettoyer et structurer les données historiques nécessaires.
- Analyser et comprendre les patterns dans les données.
- Sélectionner les algorithmes les plus adaptés à la problématique (exemple : régression, arbres de décision, réseaux neuronaux, etc.).
- Évaluer la performance du modèle à l'aide de métriques adaptées (précision, rappel, F1-score, etc.).
- Intégrer le modèle dans une interface utilisateur ou un système existant pour une exploitation pratique.

2. Cahier des charges fonctionnel

➤ Analyse des besoins :

- Identification des sources de données.
- Définition des cas d'usage spécifiques (prévision de ventes, maintenance prédictive, détection d'anomalies, etc.).

➤ Préparation des données :

- Collecte des données à partir de sources définies.
- Nettoyage des données pour éliminer les doublons, valeurs manquantes ou incohérentes.
- Création de nouvelles variables ou features pertinentes pour la prédiction.

➤ Développement du modèle :

- Étude comparative des algorithmes potentiels.
- Implémentation du modèle avec un langage adapté (Python, R, etc.).
- Optimisation des hyperparamètres pour maximiser les performances.

➤ **Validation et évaluation :**

- Diviser les données en ensembles d'entraînement, de validation et de test.
- Comparer les résultats obtenus avec d'autres approches existantes (benchmarking).

➤ **Déploiement :**

- Intégration dans un outil ou une application (web ou desktop).
- Documentation du fonctionnement pour les utilisateurs finaux.

3. Cahier des charges technique

- **Langage de programmation :** Python (Scikit-learn, TensorFlow, Keras, etc.) ou R.
- **Outils d'analyse des données :** Pandas, NumPy, Matplotlib, Seaborn.
- **Environnement de développement :** Jupyter Notebook, PyCharm, ou autres IDE.
- **Sources de données :** Fichiers CSV, bases de données relationnelles (MySQL, PostgreSQL), ou APIs.
- **Plateforme de déploiement :** Application web (Flask/Django) ou interface desktop.