Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université de Sousse
Institut Supérieur de Gestion de Sousse

Rapport

5eme année Informatique Groupe 4

Spécialité :
Génie Logiciel

IA Project : Prédiction de la Qualité de l'Air

Réalisé par	*Encadré par*
BERGAOUI Rim RHOUMA Sarra	Mme Mahbouba

Année Universitaire 2024/2025

TABLE DES MATIÈRES

INTRODUCTION GÉNÉRALE

L'air que nous respirons est un élément essentiel à notre bien-être et à notre santé. Cependant, les niveaux croissants de pollution atmosphérique dans les zones urbaines et industrielles posent de graves menaces à la qualité de vie. Les métropoles modernes, bien qu'équipées de capteurs de surveillance, peinent souvent à prévoir les pics de pollution avec précision, ce qui limite leur capacité à mettre en place des mesures préventives efficaces.

Dans ce contexte, le présent projet vise à développer un modèle prédictif capable d'estimer la qualité de l'air en utilisant des données environnementales et météorologiques. Ce projet s'inscrit dans une démarche proactive visant à prévenir les effets néfastes de la pollution, à réduire les risques pour la santé publique et à proposer des solutions durables.

Chapitre

1 Cadre Générale du Projet

Sommaire

1.1 Introduction

La pollution atmosphérique est une problématique urgente dans les grandes métropoles. Bien que les capteurs mesurent la qualité de l'air en temps réel, les villes manquent souvent d'outils prédictifs pour anticiper les variations de la pollution. Ce chapitre explore le contexte du projet, ses enjeux et ses objectifs.

1.2 Contenu

1.2.1 Contexte

Une métropole fictive souffre de graves problèmes de pollution causés par :

- La circulation routière, principale source de particules fines (PM2.5, PM10) et dioxyde d'azote (NO2).
- Les activités industrielles, qui augmentent les concentrations d'ozone (O3) et de monoxyde de carbone (CO).

1.2.2 Problématique

L'incapacité à prévoir les pics de pollution limite les actions préventives. Résultats :

- 35% de journées polluées (AQI > 100).
- Augmentation de 20% des urgences respiratoires.

1.3 Objectifs

- Prédire l'indice de qualité de l'air (AQI) avec 90% de précision.
- Proposer des alertes pour planifier des mesures préventives.
- Réduire la pollution en adaptant les politiques locales.

1.4 Conclusion

Ce projet répond à des besoins critiques en combinant données environnementales et algorithmes avancés pour améliorer la qualité de vie des citoyens.

Chapitre

2 Collecte des Données

Sommaire

2.1 Introduction

Les données jouent un rôle clé dans tout projet de prédiction. Ce chapitre présente les sources, caractéristiques et limitations des données utilisées pour modéliser la qualité de l'air.

2.2 Contenu

2.2.1 Source des données

— **Origine :** Dataset disponible sur Kaggle.
— **Taille :** 42 461 entrées avec 41 colonnes.
— **Domaine couvert :** données météorologiques, pollution et trafic.

2.2.2 Exemples de variables

— **Météo :** température, humidité, vitesse du vent.
— **Pollution :** PM2.5, PM10, NO2, O3, CO.
— **Géographie :** latitude, longitude, localisation.

2.3 Limites des données

— Présence de valeurs manquantes.
— Données bruitées causées par des erreurs de capteurs.

2.4 Conclusion

Les données collectées sont riches et variées, offrant une base solide pour entraîner un modèle prédictif robuste.

Chapitre

3 Prétraitement et Préparation des Données

Sommaire

3.1 Introduction

Avant d'entraîner un modèle, il est crucial de nettoyer et préparer les données. Ce chapitre détaille les étapes de prétraitement pour garantir la qualité des données.

3.2 Contenu

3.2.1 Nettoyage des données

— Suppression des colonnes inutiles.
— Imputation des valeurs manquantes.
— Détection et gestion des valeurs aberrantes (via boxplots).

3.2.2 Encodage des variables catégoriques

Les colonnes textuelles comme **location name** sont transformées en valeurs numériques.

3.2.3 Mise à l'échelle des données

— Normalisation des caractéristiques pour homogénéiser les échelles.
— Division en ensembles d'entraînement (80%) et de test (20%).

3.3 Conclusion

Un prétraitement minutieux garantit la cohérence et la fiabilité des données pour une modélisation efficace.

Chapitre

4 Analyse Exploratoire des Données (EDA)

Sommaire

4.1 Introduction

L'analyse exploratoire permet de comprendre les relations entre les variables et d'identifier les tendances principales. Ce chapitre présente les observations tirées de cette analyse.

4.2 Contenu

4.2.1 Visualisation des distributions

— Histogrammes pour étudier la répartition des variables.
— Cartes thermiques pour visualiser les corrélations.

4.2.2 Relations clés

— La température et l'humidité influencent les concentrations d'ozone.
— Le trafic routier est fortement corrélé aux niveaux de PM2.5 et NO2.

4.2.3 Analyse multivariée

— Pairplots pour détecter les relations complexes.
— Boxplots pour étudier les outliers.

4.3 Conclusion

L'EDA révèle des insights importants, comme l'impact des conditions météorologiques sur la pollution, qui orientent le choix des prédicteurs.

Chapitre

5 Modélisation

Sommaire

5.1 Introduction

La modélisation est l'étape où les algorithmes sont utilisés pour apprendre et prédire la qualité de l'air. Ce chapitre détaille les modèles testés, leurs résultats et leur efficacité.

5.2 Contenu

5.2.1 Modèles testés

— **Régression Linéaire :** utilisé comme référence.
— **Forêt Aléatoire (Random Forest) :** bon équilibre entre précision et interprétabilité.
— **Gradient Boosting :** modèle avancé combinant plusieurs prédicteurs faibles.

5.2.2 Évaluation des modèles

— **Gradient Boosting :** 91% de précision.
— **Forêt Aléatoire :** 89% de précision, excellente alternative rapide.
— **Régression Linéaire :** résultats limités à cause des relations non linéaires.

5.2.3 Interprétation des résultats

— Les modèles non linéaires capturent mieux les interactions complexes entre variables.

5.3 Conclusion

Le modèle Gradient Boosting offre les meilleures performances pour ce projet, tandis que la Forêt Aléatoire reste une option rapide et efficace.

Chapitre

6 Visualisation et Tableau de Bord

Sommaire

6.1 Introduction

Les résultats doivent être compréhensibles et utilisables par les décideurs. Ce chapitre présente les outils interactifs conçus pour analyser et exploiter les prédictions.

6.2 Contenu

6.2.1 Tableau de bord interactif

— Suivi des prévisions de qualité de l'air.
— Analyse des périodes critiques.
— Recommandations basées sur les prédictions.

6.2.2 Visualisation des tendances

— Cartes des zones à haut risque.
— Courbes temporelles pour anticiper les pics de pollution.

6.3 Conclusion

Les visualisations facilitent la prise de décision et renforcent l'impact des prédictions.

CONCLUSION ET PERSPECTIVES

CONCLUSION GÉNÉRALE

Ce projet a démontré la possibilité de prédire la qualité de l'air avec une grande précision. Grâce aux données environnementales et à des modèles avancés, des alertes préventives peuvent être mises en place pour limiter les effets de la pollution.

Impact

1. Amélioration de la santé publique grâce à des actions préventives.
2. Réduction de l'impact environnemental des pics de pollution.

Perspectives

— Intégration de données en temps réel pour un système plus réactif.

— Exploration de modèles encore plus avancés, comme les réseaux neuronaux.