

Analyse et Prédiction de la COVID-19

Projet Machine Learning - Santé Publique

Aperçu du projet

- * Ce projet analyse les données officielles de la COVID-19 fournies par l'OMS.
- * Objectif : Explorer les données, construire des modèles prédictifs et classifier les zones à haut risque.
- * Outils : Python, Pandas, Seaborn, Scikit-learn.

Compréhension du problème

- * Problème : Anticiper les hausses de cas COVID-19.
- * Objectif : Classifier si une région est à haut risque d'une flambée.
- * Critère de succès : Précision $\geq 75\%$.

Compréhension des données

- * Source : OMS - bases officielles.
- * Format : CSV mis à jour régulièrement.
- * Variables : Cas, Décès, Vaccinations, Politiques sanitaires.
- * Observations : Données manquantes, hétérogènes.

Préparation des données

- * 1. Nettoyage et traitement des valeurs manquantes.
- * 2. Conversion des dates, création de variables temporelles.
- * 3. Normalisation par population.
- * 4. Création variable cible Haut_Risque (0/1).

Modélisation

- * Algorithmes utilisés :
 - * • Régression logistique (classification binaire).
 - * • Forêt aléatoire (meilleures performances).
- * Évaluation : accuracy, precision, recall, f1, AUC.

Résultats et Insights

- * La forêt aléatoire a donné les meilleurs résultats.
- * Précision : >75%.
- * Variables importantes : nouveaux cas, décès, vaccination.
- * ROC-AUC élevé -> bonne capacité discriminante.

Recommandations commerciales

- * 1. Renforcer la surveillance dans les zones à forte hausse des cas.
- * 2. Planification proactive : vaccination, communication publique.
- * 3. Allocation optimale des ressources médicales.

Conclusion & Next Steps

- * Les modèles ML permettent de prévoir les zones à haut risque COVID-19.
- * Next Steps : ajouter vaccination, densité population, tester d'autres modèles.
- * Créer un dashboard interactif pour suivi en temps réel.