

PROJET 06 : Classificateur de Fake News

OBJECTIF DU PROJET

La désinformation se propage plus vite que la vérité. L'IA peut aider à signaler le contenu suspect pour examen humain, protégeant la démocratie et la santé publique.

FICHIER DE DONNÉES

- fake_news.csv
- 800 lignes
- Inclut des valeurs manquantes, valeurs aberrantes et déséquilibres pour un apprentissage réaliste

TÂCHES À RÉALISER

■ TÂCHE PRINCIPALE

Classer l'article comme Vrai ou Faux

■ TÂCHES SECONDAIRES (OPTIONNELLES)

1. Extraire les "Mots-clés Principaux" trouvés dans les fake news
2. Détecter les modèles de partage "Bot-like" (Partages élevés, lectures faibles)
3. Regrouper les articles par Sujet (Politique, Santé, Célébrités)
4. Prédire la "Viralité" (Nombre de partages)

LIVRABLES ATTENDUS

1. ****Exploration des Données (EDA)****
 - Analyse statistique descriptive
 - Visualisations des distributions
 - Analyse des corrélations
 - Identification des valeurs manquantes et aberrantes
2. ****Ingénierie des Fonctionnalités****
 - Création de nouvelles variables pertinentes

- Transformation des variables existantes
- Encodage des variables catégorielles
- Normalisation/Standardisation si nécessaire

3. ****Modélisation****

- Choix et justification du/des modèle(s)
- Entraînement et validation
- Optimisation des hyperparamètres
- Évaluation avec métriques appropriées

4. ****Résultats et Interprétation****

- Performance du modèle principal
- Résultats des tâches secondaires (si réalisées)
- Insights et recommandations
- Visualisations des résultats

CONSEILS

- Commencez par une exploration approfondie des données
- L'ingénierie des fonctionnalités est CRUCIALE pour ce projet
- Documentez votre démarche et vos choix
- Comparez plusieurs approches si possible
- Pensez à l'aspect pratique et utilisable de vos résultats

Bon travail ! ■