

# Rapport de Projet :

## Création d'un Ensemble de Données pour la Détection de Fake News en Arabe

membre de groupe:

ikram djait

nfidsa halima

## Introduction :

L'objectif de ce projet était de développer un ensemble de données complet pour faciliter la détection de fake news en langue arabe. Cet ensemble de données permet non seulement de distinguer les articles de presse authentiques des faux, mais aussi de les catégoriser par thème, offrant ainsi des perspectives sur la propagation de la désinformation dans différents domaines. Le projet impliquait la collecte de données à partir de sites Web de vérification des faits en arabe réputés

## objectifs de projet :

- Création d'un Ensemble de Données pour la Détection de Fake - - - News en Arabe
- Collecte de Données Fiables et Diversifiées
- Classification des Articles en Authentiques et Faux
- Catégorisation par Thème pour une Vue d'Ensemble
- Utilisation de Sources Vérifiées pour une Fiabilité Maximale
- Transparence et Accessibilité via un Dépôt GitHub

## méthodologie:

- Planification et Recherche : Définir les objectifs et rechercher les techniques de collecte de données.
- Sélection des Sources : Choisir des sites arabes réputés pour collecter les articles.
- Collecte de Données : Utiliser BeautifulSoup pour extraire les articles de ces sources.
- Annotation : Classer manuellement les articles comme vrais ou faux et les catégoriser par thème.
- Validation : Vérifier la précision des annotations par une révision croisée.
- Création de l'Ensemble de Données : Compiler les annotations dans un fichier CSV.
- Analyse et Rapport : Analyser les statistiques de l'ensemble de données et rédiger un rapport détaillé.
- Publication : Mettre à disposition l'ensemble de données, le code et le rapport sur GitHub pour une accessibilité transparente.

## categorization des articles:

**Politique** : Affaires gouvernementales, élections, politiciens et décisions politiques.

**Religion** : Pratiques religieuses, croyances, institutions religieuses et événements religieux.

**Santé** : Santé publique, maladies, traitements médicaux, découvertes médicales et bien-être.

**Économie** : Marchés financiers, emploi, entreprises, investissements et politiques économiques.

**Sports** : Événements sportifs, équipes, athlètes, compétitions et développements sportifs.

**Justice** : Questions légales, affaires judiciaires, droits de l'homme, justice sociale et systèmes juridiques.

**Environnement** : Changement climatique, conservation, pollution et énergies renouvelables.

**Éducation** : Système éducatif, réformes éducatives, institutions d'enseignement et questions éducatives.

**Technologie** : Avancées technologiques, innovations, entreprises technologiques, gadgets et applications.

**Culture** : Arts, musique, cinéma, littérature, traditions culturelles et événements culturels.

## Ensemble de Données d'Articles Arabes:

- Topic d'article : Le sujet principal de l'article.
- classification : La classification de l'article en tant que vrai ou faux.
- Type d'article : Le type d'article, tel que défini dans le contenu HTML de la page.
- Arther d'article : L'auteur de l'article.
- Description d'article : La description ou le résumé de l'article.
- Date de l'article : La date de publication de l'article.
- Fake or Real : Une indication si l'article est classé comme faux ou réel en fonction du type d'article.

## Conclusion :

La création d'un ensemble de données robuste pour la détection de fake news en arabe est essentielle pour lutter contre la désinformation dans la région. Ce projet a réussi à atteindre ses objectifs en collectant et en annotant des articles de presse en arabe à partir de sites Web de vérification des faits réputés. L'ensemble de données, ainsi que le code de grattage et le rapport complet, constituent une ressource précieuse pour les chercheurs et les praticiens travaillant dans le domaine de la détection et de la lutte contre la désinformation.

**Conclusion :**

La création d'un ensemble de données robuste pour la détection de fake news en arabe est essentielle pour lutter contre la désinformation dans la région. Ce projet a réussi à atteindre ses objectifs en collectant et en annotant des articles de presse en arabe à partir de sites Web de vérification des faits réputés. L'ensemble de données, ainsi que le code de grattage et le rapport complet, constituent une ressource précieuse pour les chercheurs et les praticiens travaillant dans le domaine de la détection et de la lutte contre la désinformation.

Answer

La création d'un ensemble de données robuste pour la détection de fake news en arabe est essentielle pour lutter contre la désinformation dans la région. Ce projet a réussi à atteindre ses objectifs en collectant et en annotant des articles de presse en arabe à partir de sites Web de vérification des faits réputés. L'ensemble de données, ainsi que le code de grattage et le rapport complet, constituent une ressource précieuse pour les chercheurs et les praticiens travaillant dans le domaine de la détection et de la lutte contre la désinformation.