

1. Présentation du Projet

1.1 Titre du Projet :

Détection des articles Clickbait

1.2 Description du Projet :

Le projet de Détection de Clickbait vise à développer un système robuste et précis pour identifier un article clickbait dans les articles en ligne, les titres et les publications sur les réseaux sociaux. Le clickbait fait référence au contenu conçu pour attirer l'attention et encourager les utilisateurs à cliquer sur un lien, souvent en utilisant des informations trompeuses ou exagérées.

2. Objectifs

2.1 Objectifs Principaux

- Développer un modèle d'apprentissage profond capable de détecter avec précision le contenu de clickbait.
- Implémenter une solution évolutive capable de traiter de grands volumes de données textuelles en temps réel.
- Intégrer le modèle de détection de clickbait dans les plateformes ou systèmes existants pour une utilisation plus étendue.

2.2 Objectifs Spécifiques

- Atteindre un taux de précision minimum de 90 % dans la détection de clickbait.
- Réduire au minimum les taux de faux positifs et de faux négatifs.
- Assurer que le modèle s'adapte à différents types de contenu, y compris les articles, les titres et les publications sur les réseaux sociaux.

3. Portée

3.1 Inclusions :

- Entraînement d'un modèle d'apprentissage profond sur un ensemble de données diversifié de contenu de clickbait et non-clickbait.
- Ajustement fin du modèle pour améliorer les performances sur des types de contenu spécifiques.
- Développement d'une API pour l'intégration avec des systèmes externes.

3.2 Exclusions :

- Détection de clickbait dans des contenus non textuels (par exemple, images, vidéos).
- Analyse du comportement de l'utilisateur ou des métriques d'engagement.

4. Spécifications Techniques

4.1 Stack Technologique :

- Cadre d'apprentissage profond : TensorFlow.
- Langage : Python.
- Développement d'une interface web : Streamlit, FastAPI ou Flask.

4.2 Architecture du Modèle :

- Considération de réseaux neuronaux récurrents (RNN) ou d'architectures basées sur les transformateurs pour le traitement de données séquentielles.
- Utilisation d'incorporations de mots pré-entraînées (par exemple, Word2Vec, GloVe) pour la compréhension sémantique.

5. Besoins en Données

5.1 Données d'Entraînement :

- Ensemble de données annotées de contenu de clickbait et non-clickbait.
- Représentation diversifiée des sources de contenu, des genres et des styles d'écriture.

5.2 Données de Validation :

- Un ensemble de données distinct pour la validation du modèle afin d'assurer la généralisation.

6. Métriques de Performance

6.1 Critères d'Évaluation :

- Précision, Rappel, et F1 Score.
- Courbe caractéristique de fonctionnement du récepteur (ROC) et Aire sous la Courbe (AUC).

6.2 Critères d'Acceptation :

- Le modèle doit atteindre la précision minimale spécifiée et un équilibre précision/rappel.