**Titre du Projet** : **La détection du sarcasme en arabe**

**Objectifs du Projet :**

La détection du sarcasme est le processus visant à déterminer si un morceau de texte est sarcastique ou non. Le sarcasme représente l'un des principaux défis pour les systèmes d'analyse de sentiment. La raison en est qu'une phrase sarcastique porte généralement un sentiment implicite négatif, bien qu'il soit exprimé à l'aide d'expressions positives. Cette contradiction entre le sentiment apparent et celui intentionnel crée un défi complexe pour les systèmes d'analyse de sentiment.

La détection du sarcasme a attiré l'attention dans d'autres langues, mais l'arabe est encore en retard à cet égard. Il y a eu peu d'efforts pour la détection du sarcasme en arabe, tels que les travaux de (Karoui et al., 2017 ; Ghanem et al., 2020) et la tâche partagée organisée par (Ghanem et al., 2019). Des efforts récents ont été déployés pour construire des ensembles de données standard pour cette tâche, tels que (Abbes et al., 2020 ; Abu Farha et Magdy, 2020). La tâche partagée sur la détection du sarcasme et du sentiment en arabe se tiendra avec WANLP@EACL2021. Cette tâche partagée se concentrera sur l'analyse des tweets et l'identification de leur sentiment ainsi que la détermination de leur caractère sarcastique ou non.

**Méthodologie :**

1. Collecte de Données : Collecter des données à partir de différentes plateformes de médias sociaux, en se concentrant sur des sujets spécifiques pertinents.

2. Prétraitement des Données : Nettoyer et prétraiter les données textuelles pour les rendre compatibles avec les modèles de Machine Learning.

3. Construction d'un Modèle de Machine Learning : Mettre en œuvre un modèle de Machine Learning adapté à l'analyse de sentiments, tel qu'un modèle de classification.

4. Entraînement du Modèle : Entraîner le modèle sur un ensemble de données annoté pour la classification des sentiments.

5. Évaluation du Modèle : Évaluer les performances du modèle sur un ensemble de test et ajuster les hyper paramètres si nécessaire.

**Livrables Attendus :**

1. Rapport Technique : Document détaillé décrivant chaque étape du projet, y compris les choix de conception et les résultats obtenus.

2. Code Source : Ensemble de scripts, modules, et fichiers nécessaires à la reproduction du projet.

**Calendrier Prévisionnel :**

1. Semaines 0 : Téléchargement des données.
2. Semaines 1-2 : Uploader les données Téléchargées puis effectuer les prétraitements nécessaires suivant l’application

2. Semaines 3-4 : Construction du modèle et entraînement initial.

3. Semaines 3-4 : Évaluation du modèle et ajustements.

4. Semaines 5-6 : Développement de l'interface utilisateur.

5. Semaines 9-10 : Finalisation, rédaction du rapport, et le dépôt de rapport.

**Critères d'Évaluation :**

Le projet sera évalué en fonction de la qualité du modèle de Machine Learning, de la précision de l'analyse de sentiments, de la convivialité de l'interface utilisateur, et de la clarté du rapport final.

Baseline\_score=33.71