

Université Espoir de Calvary Chapel



Sciences Informatiques

Projet

Détecteur de commentaire toxique

Dans le cadre du cours : Introduction Machine Learning

Préparé par :

Jennifer BOURDEAU

Lorvenson CONSTANT

Wendy NESTAND

Blanjina PROSPERE

Luckson SURPRICE

Dispensé Par :

Seige POTEAU

Le : 21/03/2023

I. Introduction

II. Collecte et préparation des données

III. Techniques de traitement du langage naturel

IV. Expérimentation de différents algorithmes de classification

V. Sélection du modèle final

VI. Évaluation du modèle

VII. Conclusion

I. Introduction

Les commentaires toxiques sur les plateformes en ligne sont un problème croissant qui peut causer des dommages émotionnels et psychologiques aux personnes qui en sont victimes. L'utilisation du machine Learning pour détecter et classifier les commentaires toxiques peut aider à lutter contre ce phénomène. Ce document a pour but de fournir des informations sur l'utilisation du machine Learning pour classifier les commentaires toxiques ainsi que les différentes étapes impliquées dans ce processus.

II. Collecte et préparation des données

La première étape dans l'utilisation du machine Learning pour détecter et classifier les commentaires toxiques est la collecte et la préparation des données. Un jeu de données annotées de commentaires toxiques est nécessaire pour entraîner le modèle de classification. Le jeu de données "Toxic Comment Classification Challenge" a été utilisé pour ce projet. Les données ont été préparées en les nettoyants, en les normalisant et en les vectorisant.

III. Techniques de traitement du langage naturel

Les techniques de traitement du langage naturel sont utilisées pour traiter les commentaires textuels avant leur utilisation dans le modèle de classification. Les techniques de traitement du langage naturel comprennent le nettoyage des données, la suppression des mots vides, la lemmatisation, la tokenisation et la vectorisation.

IV. Expérimentation de différents algorithmes de classification

Plusieurs algorithmes de classification ont été expérimentés, tels que la régression logistique, le k NN, le SVM et le réseau de neurones convolutifs. Les modèles ont été entraînés avec différentes combinaisons de techniques de traitement du langage naturel pour trouver le meilleur modèle de classification.

V. Sélection du modèle final

Le modèle de réseau de neurones convolutifs a montré la meilleure performance parmi les différents algorithmes de classification expérimentés. Le modèle a été entraîné sur les données préparées en utilisant différentes combinaisons d'hyperparamètres pour obtenir le meilleur résultat possible.

VI. Évaluation du modèle

Le modèle final a été évalué en utilisant des métriques telles que la précision, le rappel et la F1- score. Le modèle a montré une précision moyenne de 90% et une F1- score moyenne de 0,89.

VII. Conclusion

L'utilisation du machine Learning pour détecter et classifier les commentaires toxiques peut aider à réduire la propagation de discours haineux sur les plateformes en ligne. Le processus implique la collecte et la préparation des données, l'utilisation de techniques de traitement du langage naturel, l'expérimentation de différents algorithmes de

classification et la sélection du modèle final. Les résultats ont montré que le modèle de réseau de neurones convolutifs peut être efficace pour classifier les commentaires toxiques en ligne.