

Compte Rendu

Projet d'Analyse des Résultats COVID

Réalisé par:

Groupe 4

BENABBAS Ikram

BEN AISSA Sabri

CHENNA Manel

MBARKI Yousra

Introduction

L'objectif de ce projet est de collecter, nettoyer, analyser et visualiser les données liées à la pandémie de COVID-19 à partir du site web "www.worldometers.info/coronavirus". Les résultats de cette analyse sont ensuite enregistrés dans un fichier CSV pour une utilisation ultérieure. Ce compte rendu couvre les principales étapes du projet et les résultats obtenus.

Collecte de Données

La première étape consiste à collecter les données à partir du site web en utilisant des bibliothèques Python telles que *Beautifulsoup*, *NavigableString*, *Pandas*, et *Requests*. Le code extrait les informations relatives au COVID-19, notamment le nombre de cas, de décès, de guérisons, etc., pour chaque pays ou région.

Nettoyage et Prétraitement des Données

Après la collecte des données, un processus de nettoyage et de prétraitement est appliqué pour garantir que les données sont de haute qualité. Les étapes comprennent la suppression de doublons, le remplacement des valeurs manquantes par zéro, et la conversion des colonnes numériques en types de données appropriés.

Analyse des Données

L'analyse des données est axée sur deux principaux objectifs :

Trier par Total Deaths : Les données sont triées en fonction du nombre total de décès, ce qui permet d'identifier les pays ou régions les plus touchés par la pandémie.

Visualisation des Données : Plusieurs graphiques sont générés pour visualiser les données, notamment des graphiques de barres pour les 10 pays les plus touchés et des diagrammes circulaires pour la répartition des cas à l'échelle mondiale.

Résultats

Le projet génère les résultats suivants :

- ❖ Un fichier CSV contenant les données COVID-19 nettoyées.
- ❖ Un graphique des 10 pays les plus touchés par le COVID-19 en termes de décès.
- ❖ Un graphique en barres montrant la progression du nombre de cas dans les 10 pays les plus touchés.

- ❖ Un diagramme circulaire illustrant la répartition des cas dans le monde.

Analyse de Sentiment sur les Tweets COVID-19 :

En parallèle, ce projet inclut également une analyse de sentiment sur les tweets liés au COVID-19 en utilisant divers algorithmes de classification. Voici ce que nous avons fait dans ce code :

- Importation des bibliothèques nécessaires pour l'analyse de sentiment et la classification.
- Prétraitement des données textuelles des tweets, y compris la conversion en lettres minuscules, la suppression de la ponctuation et des chiffres, ainsi que la suppression des espaces inutiles.
- Classification des sentiments des tweets en trois catégories : positif, neutre et négatif, en utilisant la bibliothèque TextBlob.
- Utilisation de la classification par texte en utilisant des algorithmes tels que KNeighborsClassifier, DecisionTreeClassifier, RandomForestClassifier, LogisticRegression et MultinomialNB.
- Création de visualisations, notamment des graphiques de barres pour la distribution des labels de sentiment et des diagrammes circulaires pour la répartition des sentiments.
- De plus, nous avons entraîné un modèle de régression linéaire basé sur la longueur des tweets pour prédire les sentiments. Ce modèle a été évalué et utilisé pour prédire le sentiment d'un nouveau tweet.

Conclusion

Ce projet d'analyse des résultats COVID-19 permet de collecter, nettoyer et visualiser les données relatives à la pandémie. Il fournit des informations utiles sur la situation durant la période COVID-19 dans différents pays et régions, tout en incluant une analyse de sentiment des tweets liés au COVID-19 pour une perspective complète.