2020 • PYTHON FOR DATA ANALYSIS

PPG-DaLiA

30 Janvier 2020

Elise Cabrerizo

Pré-requis d'utilisation

Jupyter Notebook : Attention il faut changer le chemin d'accès de l'entrée 2 et de l'entrée 13 afin de pointer sur les Pickels

API : Il faut changer le chemin d'accès du model_simple.sav et du scaler.sav. Pour lancer l'API :
python 'chemin daccès jusqu'au dossier'/manage.py runserver

N.B. : Je n'ai pas pu mettre mon modèle sur GitHub car il est trop lourd. Il faut donc soit relancer le JupiterNotebook ou télécharger puis dézipper le fichier 'PourAPI' qui contient le scaler et le modele_simple au lien suivant :

https://drive.google.com/file/d/1Ugzx4e1Zcyj9tBF9hfUAiSLMO6a2Pgav/view?usp=sharing

Introduction

L'étude travaille avec 15 sujets , sept sujets masculins et huit sujets féminins, âgés en moyenne de 30,60. On connait aussi leur âge, sexe, taille, poids, type de peau et niveau de forme physique.
Chaque sujet a suivi un protocole de collecte de données défini, comprenant huit activités différentes. Ce protocole dure au total 2h30.
Les donnéesont été enregistrées avec deux appareils: un appareil porté sur la poitrine et un appareil porté au poignet.

Il s'agit donc de prédire l'activité qu'exécute le sujet en fonction de ses informations personnelles et son activité cardiaque.

Importation du dataset

Le dataset est découpé en 15 parties, ce qui correspond au 15 sujets d'étude.
On y retrouve 2 csv, un zip, les données du Respiban et enfin ce qui nous intéresse le pickel.

En effet, le pickel regroupe toutes les informations.
Ce fichier est un dictionnaire, avec les clés activity, label(Fréquence cardiaque), questionnaire (informations sur le sujet), rpeaks, signal (toutes les données brutes synchronisées du RespiBAN et de Empatica E4 et enfin l'id du sujet.

Etant en mineure CyberSécurité et n'ayant jamais eu de cours de python, j'ai d'abord voulu me familiariser avec le dataset.
Pour cela j'ai d'abord importé les données d'un sujet afin de reproduire la datavisualisation, qui montre le lien entre l'activité et la fréquence cardiaque, du schéma présent sur la dernière page du ReadMe.

Une fois la prise en main faite, il a donc fallu regrouper les différents Pickles. Ceci a été une étape important car ne faisant pas tous la même taille il a d'abord fallu les mettre dans un tableau puis traiter les informations pour les mettre dans le même dataframe. La première étape a été de les mettre tous à la même échelle (celle de l'activity). Afin que l'accélération soit prise en compte dans les futurs modèles j'ai divisé la colonne en 3 (x,y,z). J'ai aussi dû renommer les colonnes qui avaient perdu leur étiquettes (0 apparaissait).

Pour les modèles, j'ai suivi pas à pas le TD sklearn et reproduit la même chose. Je me suis alors rendu compte qu'il y avait des valeurs Nan dans la colonne label, j'ai alors trouvé une fonction qui répétait toutes les valeurs x fois afin de garder la même moyenne. Cette fonction prenant le x inférieur, il a donc aussi fallu rajouter la valeur moyenne de toutes les données afin de compléter les dernières lignes vides.

MA REFLEXION

LES DIFFÉRENTS MODELES

Comme le modèle induit une classification j'ai utilisé les modèles suivants :

- DecisionTreeClassifier
- RandomForestClassifier
- ExtraTreesClassifier
- AdaBoostClassifier

ExtraTreesClassifier étant le meilleur modèle (meilleur résultat) je l'ai donc choisi afin de l'implémenter dans l'API.

Mise en place de l'API

MISE EN PLACE

J'ai suivi le PDF Django_tutoriel de Luc Bertin.

MODELE

Il a fallu reprendre les colonnes choisies du X et Y afin de créer model et serializers.

PROBLEME RENCONTRÉ

Anaconda en 64 bits et python en 32 bits. Il m'a donc fallu réinstaller totalement python en 64 bits ainsi que les packages.

Un exemple d'appel avec l'API. On voit que le modèle retourne 5 comme activité (ce qui est valide car les données correspondent à la ligne 15100 du dfalldataframes).

Un exemple d'appel avec l'API. On voit que le modèle retourne 3 comme activité (ce qui est valide car les données correspondent à la ligne 4869 du dfalldataframes).

Conclusion

Je ne pensais pas réussir à aller jusqu'au bout du projet par manque de connaissances mais je suis contente d'avoir réussi à sortir un projet qui, je l'espère, est plutôt propre.