

Détection de phonèmes par analyse du signal vocal

DIJOUX Romain, GHARBI Mohamed

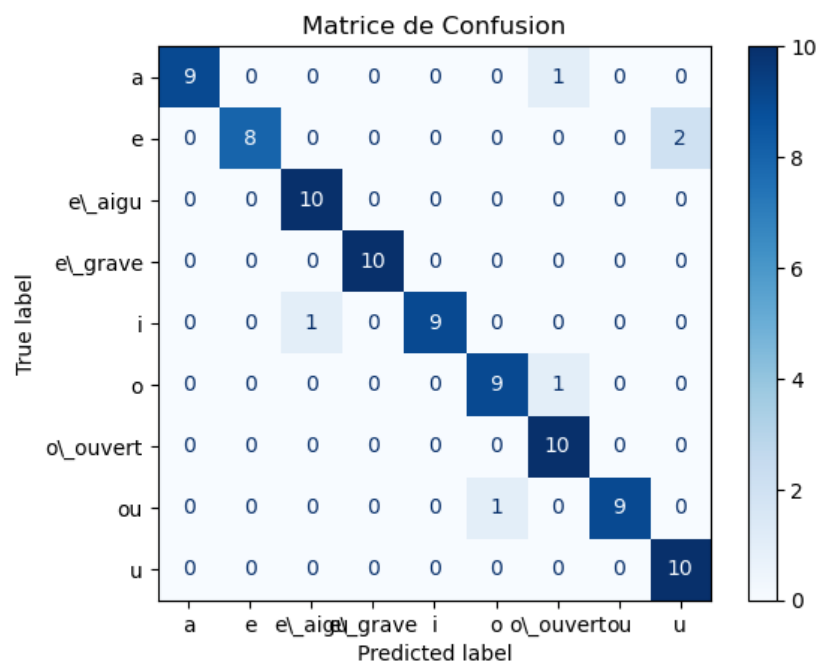
Partie I : Classification des phonèmes sans prétraitement

Apprentissage supervisé sans prétraitement :

On a choisi KNN (k-nearest neighbors ou k plus proche voisins) comme algorithme à appliquer.

KNN est un algorithme d'apprentissage supervisé utilisé pour la classification et la régression. Il appartient à la famille des algorithmes d'apprentissage basés sur les instances, ce qui signifie qu'il utilise les données d'entraînement directement pour effectuer des prédictions sans passer par une étape d'apprentissage explicite.

Après entraînement sur des données de test, en l'appliquant sur la totalité du jeu de test on obtient cette matrice de confusion.



Les éléments sur la diagonale sont bien classés.

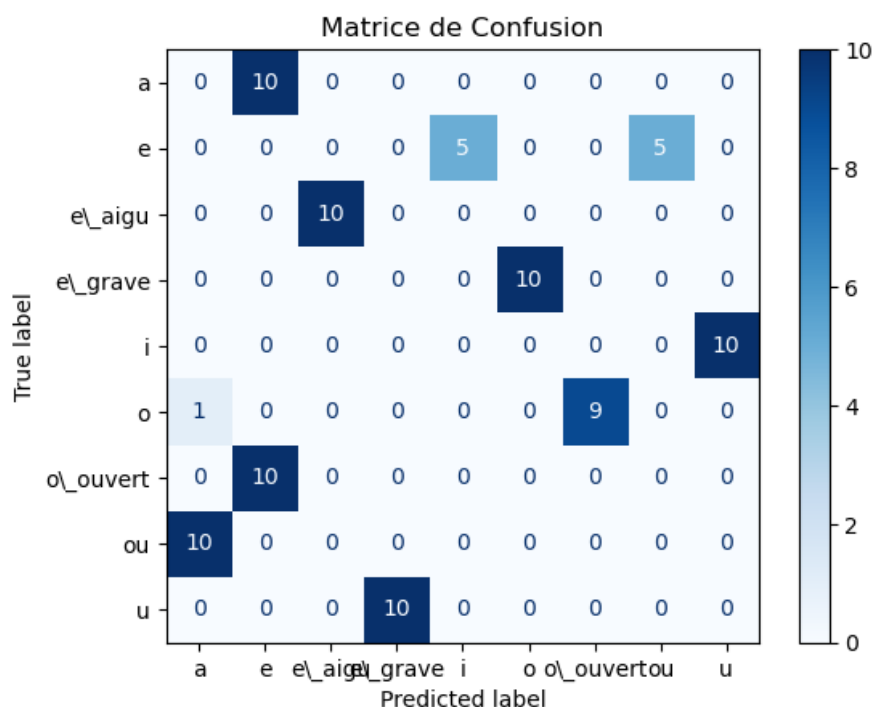
Apprentissage non supervisé sans prétraitement :

On a choisi K-Means comme algorithme à appliquer.

K-means est un algorithme de clustering non supervisé utilisé pour regrouper des éléments en différents groupes en fonction de leurs similitudes. L'algorithme utilise une approche itérative pour assigner chaque élément à un groupe en fonction du centroïde.

Nous obtenons les 9 clusters visualisés sur la matrice de confusion ci-dessous.

Apprentissage non supervisé avec prétraitement : Normalisation standard + ACP + k-means:



Partie III : Analyse des résultats

Les résultats obtenus via l'apprentissage supervisé via KNN sont assez satisfaisants, Légères confusion d'un 'a' comme un 'o ouverts' et de 'e' avec des 'u' , d'un 'i' avec un 'e aigus', d'un 'o' avec un 'o ouvert' et d'un 'ou' avec un o. Dans tous ces cas, on peut voir que ce sont des voyelles proches entre elles dans le triangle voyellique.

Lorsque l'on applique un pré-traitement, la majorité des confusions est corrigée sauf un 'o' qui est classé comme 'o ouvert'. Ces voyelles sont proches sur le triangle voyellique mais on peut aussi supposer que cette anomalie provient de l'accent d'une partie des personnes de notre jeu de données.

La lecture des résultats obtenus via la méthode non supervisée avec le K-means sont plus difficile à lire en raison de des différences de label, néanmoins on peut observer que :

- 'a' a été séparé en deux cluster avec la majorité des échantillons dans le même cluster que le 'o ouvert'
- 'e' a été séparé en trois cluster, confondu avec 'ou' et 'o ouvert'
- 'e aigus' séparé en 3 cluster, confondu en majorité avec des avec des 'i' et en minorité avec des 'e grave'
- 'o', 'o ouvert', et 'ou' sont relativement bien classées, avec 9/10 des échantillons dans réunis dans un cluster
- Les voyelles 'e grave' et 'u' sont bien classées chacune.

On remarque que la majorité des confusions est faite sur des voyelles assez proches sur le triangle voyellique.

En appliquant les pré-traitements à la KNN, on peut voir une amélioration des résultats mais :

- Les 'e' ont été séparés en deux clusters.
- Les 'o ouverts' sont totalement confondus avec des 'a'. Toutes deux sont des voyelles graves.