**Compte rendu des meilleurs scores dans plusieurs modèles**

**SANS LE TEXT PREPROCESSING 670 de tweets sur 470 features**

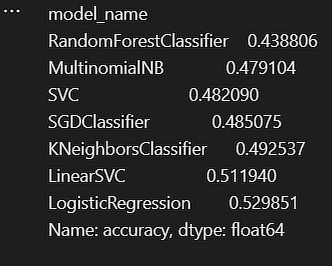
RandomForestClassifier(n\_estimators=200, max\_depth=3, random\_state=0)

KNeighborsClassifier(30)

Une image contenant texte

Description générée automatiquement**CV = 5**

**CV = 10**



Une image contenant texte

Description générée automatiquement**CV = 40**

Une image contenant texte

Description générée automatiquement**CV = 20**

**CV = 50**

RandomForestClassifier(n\_estimators=400, random\_state=0)

Une image contenant texte

Description générée automatiquement

**CV = 40**

RandomForestClassifier(n\_estimators=400, random\_state=0)

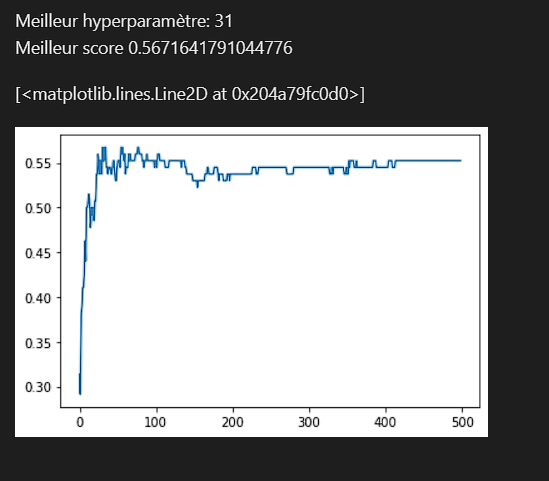
Une image contenant texte

Description générée automatiquement

Une image contenant texte

Description générée automatiquement**CV = 60**

RandomForestClassifier(n\_estimators=k, random\_state=0)



Transformer pour n-iteration = 50

*TFIDVECTORISER*

**SANS LE TEXT PREPROCESSING (1232 tweets,889 features) VERSION 0 TWEETS**

LinearSVC() : 0.515695067264574

LogisticRegression(random\_state=0) : 0.44170403587443946

**RandomForestClassifier(n\_estimators=400, random\_state=0) : 0.566798**

**AVEC LE TEXT PREPROCESSING**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Nombre de tweets** | **Nombre de features** | **% test split** | **Suppressions / transformations** | **RandomForest n\_estimators= 400** | **LinearSVC** | **LogisticRegression** |
| **2276** | **889** | **0.2** | **numérique, justice juridique** | **0.599** | **0.587** | **0.591** |
| **2296** | **6732** | **0.2** | **numérique, justice juridique** | **0.626** | **0.647** | **0.604** |
| **2760** | **7117** | **0.2** | **numérique, droit politique** | **0.635** | **0.641** | **0.592** |
| **3240** | **7904** | **0.2** | **numérique droit politique** | **0.608** | **0.651** | **0.634** |
| **3738** | **8443** | **0.2** | **numérique démocratie** | **0.6591** | **0.652** | **0.659** |
| **3738** | **8443** | **0.27** | **numérique démocratie** | **0.638** | **0.67** | **0.638** |

Non, l'idée est d'avoir le modèle avec les tweets nettoyés, mais ce qui devrait normalement être fait, c'est une fonction qui prend en entrée une liste de tweets bruts mais la fonction doit nettoyer de la même manière ces tweets bruts comme pour l'entraînement que tu as déjà pu faire