

Projet Machine Learning



Martin SANCHEZ, Gaétan ROMERO, Matthieu TRINQUART,
Léa GARCIA, Marie-Lou DESBOS

24 mai 2022

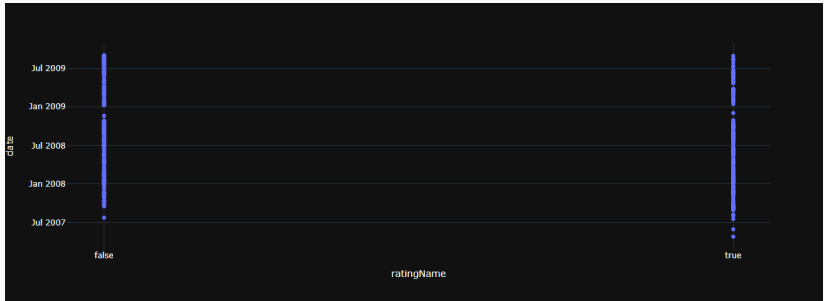
Faculté des sciences de Montpellier

SOMMAIRE

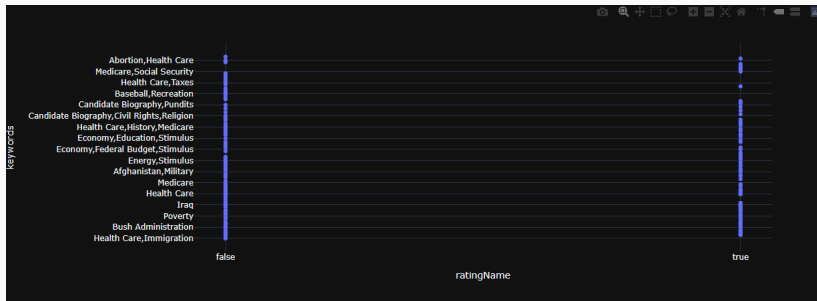
1. Visualisation des données
2. Première classification
3. Meilleurs résultats
4. Amélioration avec les hyper paramètres
5. Conclusion

Visualisation des données

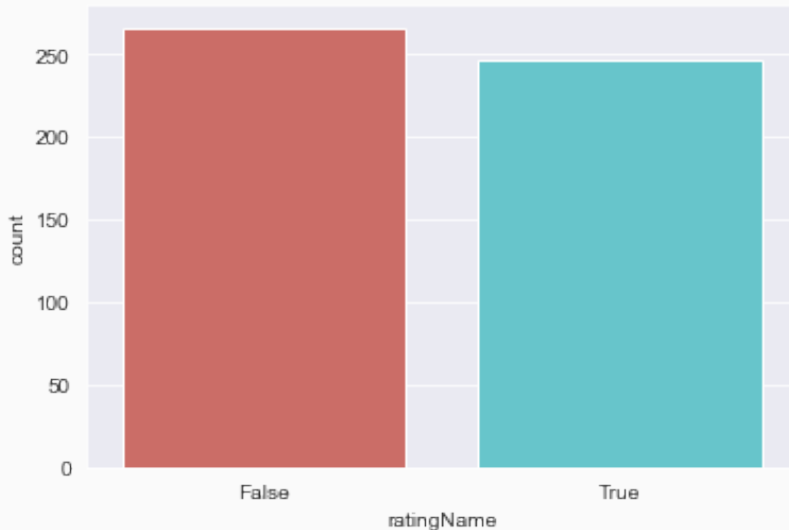
VISUALISATION DES DONNÉES



VISUALISATION DES DONNÉES



VISUALISATION DES DONNÉES

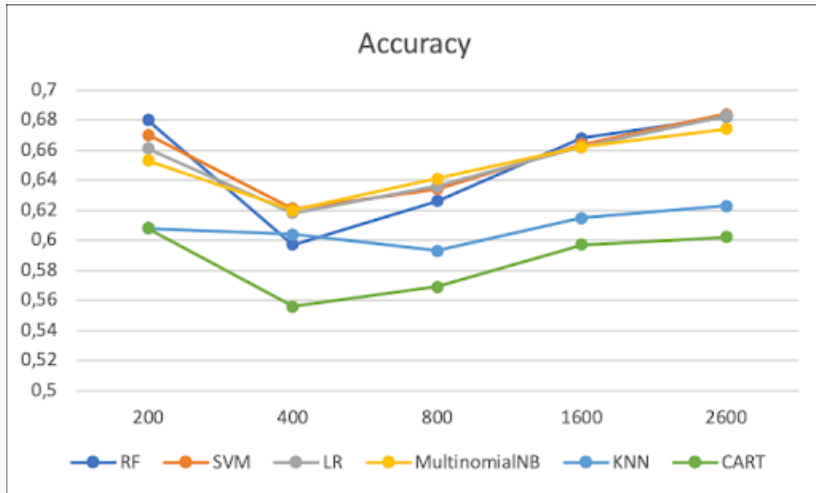


Première classification

PREMIÈRE CLASSIFICATION

```
9  
10 models = []  
11 models.append(('MultinomialNB', MultinomialNB()))  
12 models.append(('LR', LogisticRegression(solver='lbfgs')))  
13 models.append(('KNN', KNeighborsClassifier()))  
14 models.append(('CART', DecisionTreeClassifier()))  
15 models.append(('RF', RandomForestClassifier()))  
16 models.append(('SVM', SVC()))
```


PREMIÈRE CLASSIFICATION



TRAITEMENT DES DONNÉES

```
#Définition variables apprentissage et de prédiction

# Extraction des variables
Xsource = df.source
Xtext = df.text
Xheadline = df.headline
Xauthor = df.author
Xentities = df.named_entities_claim
Xkeywords = df.keywords

#Concaténation du texte à traiter
X = Xtext + Xheadline + Xsource + Xauthor + Xentities + Xkeywords

#Variable de prédiction
y = df['ratingName']
```

TRAITEMENT DES DONNÉES

```
98
99 ↳ # création d'un objet de la classe TextNormalizer
100 text_normalizer = TextNormalizer(removestopwords=True,
101                                   lowercase=True,
102                                   removedigit=False,
103                                   getstemmer=True,
104                                   getlemmatisation=False)
105
```

Meilleurs résultats

MEILLEURS RÉSULTATS

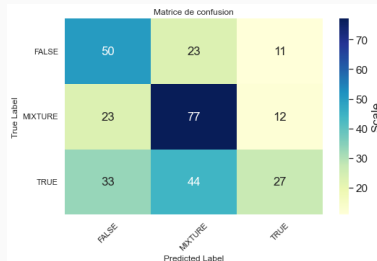
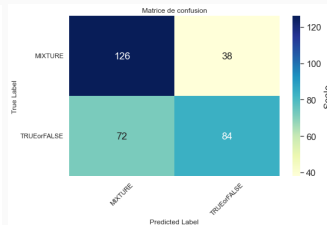
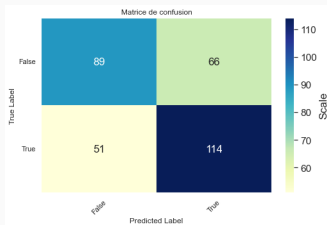
LE TOP 3

- Multinomial Naive Bayes
- Random Forest
- Le cas SVM

Amélioration avec les hyper paramètres

AMÉLIORATION AVEC LES HYPER PARAMÈTRES

Deux classes : Vrai et Faux / Deux classes : Vrai ou Faux et Mixture



Conclusion

CONCLUSION

