

Analyse de sentiments

HMIN232M – Méthodes de la science des données

B. Rima E. Youssef T. Shaqura

M1 Informatique AIGLE

25 avril 2019

Sommaire

Pré-traitements

Visualisation des données

Cross-validation

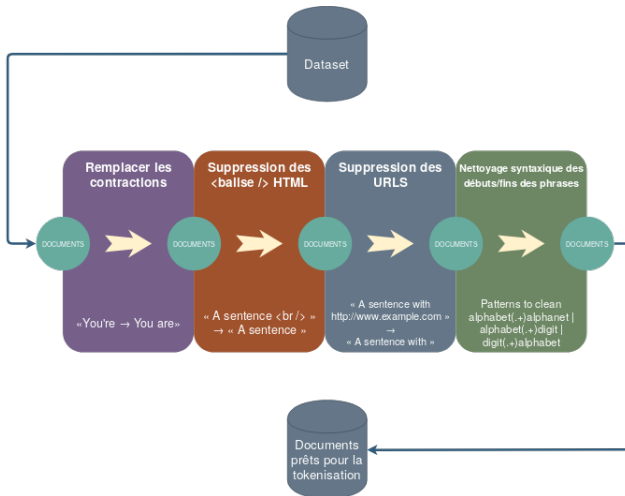
Calibrage des hyperparamètres

Création des pipelines

Conclusion

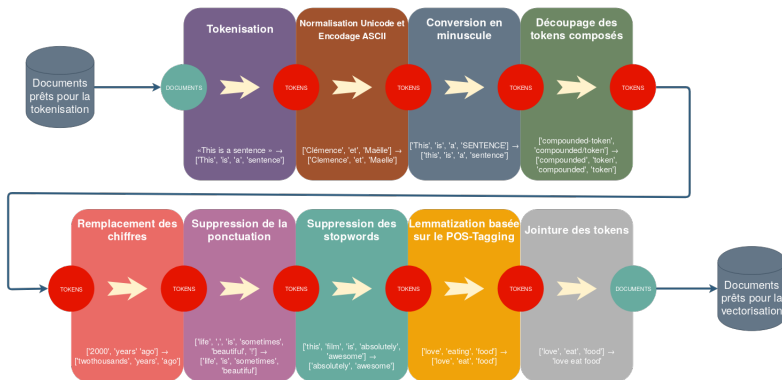
Préparation à la tokenization

Pré-traitements



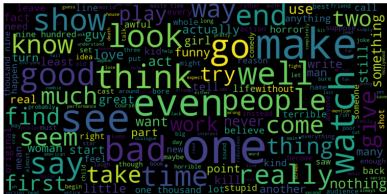
Tokenisation et normalisation

Pré-traitements



WordCloud

Visualisation des données

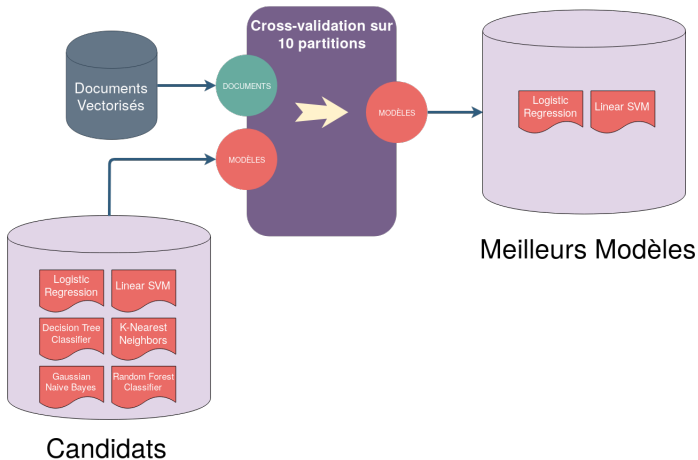


On peut s'attendre à...

- Beaucoup d'ironies
- Phrases à polarités différentes dans les avis

Figure – Les mots les plus fréquents dans les avis négatifs

Principe Cross-validation



Résultats de la cross-validation

Cross-validation

Modèle	\overline{score}	σ
LinearSVC	92%	1%
SGDClassifier	92%	1%
LogisticRegression	91%	0.8%
GaussianNB	84%	1%
RandomForestClassifier	81%	1%
KNeighborsClassifier	79%	1%
DecisionTreeClassifier	75%	0.8%

Principe

Calibrage des hyperparamètres

Pipeline pour Logistic Regression

Création des pipelines

Pipeline pour Gaussian Naive Bayes

Création des pipelines



Résultats

Création des pipelines

Schéma globale de nos traitements

Conclusion

