



INTRODUCTION

Pour ce projet, deux notebooks ont été réalisés. Le premier consiste d'un modèle supervisé pour prédire la catégorie d'une application. Le deuxième consiste d'un modèle non-supervisé pour trouver des clusters de commentaires.

PROBLÈME N°1

Le premier problème sur lequel j'ai travaillé était la prédiction de la catégorie d'une application en se basant sur le nom de l'application. Pour cela, j'ai utilisé un **classifieur multi-class OAA avec des Perceptrons**.

PROBLÈME N°2

Le deuxième problème que j'ai essayé d'adresser était la prédiction du sentiment d'un commentaire sans supervision. Pour cela, j'ai utilisé **l'algorithme des k-moyennes**.

TRAITEMENT DE TEXTE

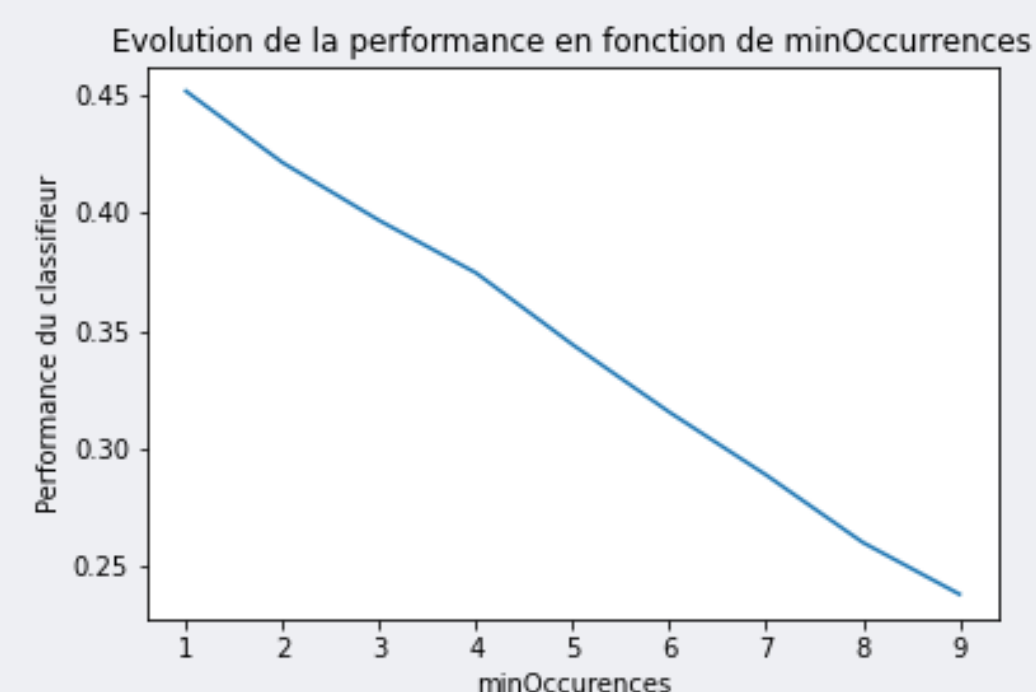
Les deux problèmes traités ont nécessité un traitement de texte. Pour faire cela, j'ai écrit une fonction pour extraire les mots les plus fréquents pour créer un dictionnaire. Ensuite, ce dictionnaire est utilisé pour créer une description pour chaque ligne de données (un tableau de bits de présence pour chaque mot du dictionnaire).

RESULTATS

Les résultats pour le premier problème étaient satisfaisants. La validation croisée a été utilisée pour évaluer la performance du classifieur.

Nous obtenons une performance d'environ **45%** dans la prédiction de la catégorie d'une application, contre 3% pour une prédiction aléatoire.

Nous avons aussi observé que la performance du classifieur dépendait de la taille du dictionnaire.



Les résultats pour le deuxième problème étaient moins intéressants. L'algorithme avait tendance à trouver des clusterings basés sur la présence ou non d'un seul mot. Néanmoins, ces mots étaient des mots clés populaires qui exprimaient un sentiment particulier.

Pour conclure, la solution proposée pour le premier problème semble adéquate. Pour le deuxième problème, j'aimerais savoir comment je pourrais améliorer les résultats.