Knn : Contrairement à la méthodes mot-clé, Knn on a besoin d'utiliser notre base d'apprentissage. Dans notre base d'apprentissage, nous avons annoté tous les tweets à la main. Quand on prend un nouveau tweet, on calcule sa distance avec tous les tweets dans notre base, et puis on choisit les 5 tweets les plus proches. Ce nouveau tweet sera de la classe (positif, négatif ou neutre) qui est majoritaire parmi ces 5 tweets de la base.

-----------------------------------------------------------------------------

Bayes : Pour la méthode Bayes, on doit calculer la probabilité qu’un nouveau tweet soit de chaque classes, une fois calculer notre tweet sera de la classe la plus « probable ».

On divise notre base de tweets en trois partie (neutre, négative, positive) dans un premier temps. Dans un second temps pour chaque mot du nouveau tweet on cherche combien de tweets de la base contiennent ce mot et leurs natures. On en déduit la probabilité que ce mot appartient à une classe. Puis on multiple les probabilités de tous les mots par la proportionnalité de tweets de chaque classe de la base de tweets. Cette façon de procéder est la présence en uni-gramme.

On peut aussi mettre le résultat de la présence de chaque mot à la puissance du nombre d'occurrence de chaque mot dans le nouveau tweet.

Une autre méthode consiste à prendre en compte les combinaisons de mots : on utilise la méthode bi-gramme, c’est à dire comparer les mots par groupe de deux. Par exemple : j'aime la pomme. On les divise en "j'aime la" et "la pomme". Finalement, on utilise la combinaison uni-gramme et bi-gramme en même temps pour obtenir le résultat le plus précis possible.

Analyse expérimentale :

-----------------------------------------------------------------------------

Pour comparer le taux d'erreurs des trois algorithmes, on réutilise notre base d’apprentissage. Pour mot-clé, c'est simple, on utilise juste cette méthode pour annoter automatiquement notre base de tweets et les comparer avec les résultats donnés à la main. Pour le Knn et Bayes, on doit diviser notre base de tweets en 10 parties, et utilise les 9 partie pour annoter la partie restante, et ça pour les 10 parties.

----------------------------------------------------------------------------

Conclusion :

Ce semestre, on a appris comment utiliser l'api twitter pour obtenir les tweets récents et swing pour designer l'interface de notre application. Nous nous sommes intéressé à ce projet parce que ça nous a permis d'apprendre trois algorithmes de classification et connaître l'avis générales des gens sur certains sujets. C'est la première fois que nous faisons une application du début jusqu'à la fin. Ça nous a aidé à améliorer notre technique de développement. Pendant ce projet, nous avons rencontré beaucoup de problèmes, c'est aussi important pour nous d'apprendre à trouver et résoudre les problèmes.