**Corpus**

**Choix du sujet**

Le sujet choisi pour ce projet concerne l’analyse de sentiment. De manière plus précise, on va traiter le degré de positivité et de négativité des textes. Cette analyse consiste à déterminer si un texte contient des opinions positives, négatives ou neutres.

Dans notre projet, le corpus (textes) choisi sera des commentaires de produits maquillages. Le maquillage est un des de nos jours qui incite le croissant économique de tout le monde. Afin d’être plus pointu, nous avons choisi la sous-catégorie de “teint” dans la catégories de maquillages car selon le rapport de l’industrie comestique, les produits les plus vendus en 2023 et la tendance de consommation au domaine. En outre, il existe plus de commentaires qui rendent la cohérence du degré de sentiment plus précise.

**Processus d’analyse**

Le premier étape est recueillir le corpus. Ensuite, le deuxième étape concerne le processus d’analyse du corpus, dont de sentiment, qui implique l’utilisation de l’outil d’algorithme d’apprentissage automatique “weka” pour atteindre à notre objectif d’analyse. Dans cet étape, nous avons entraîner le modèle sur “weka” en fonction de corpus que nous avons collecté manuellement en ligne.

**Choix du corpus**

Nous avons choisi des commentaires de marques différents en lignes qui inclut déjà le degré de positivité et de négativité par 5 étoiles. Les marques : Armani, BobbiBrown, Dior, Estée lauder, Lancôme, MAC, Nars.

Langues choisie : français.

**Recueilli du corpus**

Nous n’avons recueilli que la partie du texte, les emojis et des autres icônes ne sont pas pris en compte.