

Une plateforme d'analyse d'opinions en temps réel sur Twitter avec recommandation

Noureddine Azzouza*, Karima Akli-Astouati*
Samy Ait- Bachir*, Amira Oussalah*

*FEI- Département Informatique, Laboratoire RIIMA
Université de Sciences et de Technologie Houari Boumedién (USTHB)
Bab ezzouar, alger, Algérie
n_azzouza@esi.dz, kakli@usthb.dz, {samyab2, oussalah.amira}@gmail.com

1 Introduction

Le suivi des opinions propose beaucoup de défis dans le domaine du "Data Analytics ", qui est utilisée dans plusieurs domaines tel que le marketing, le "Sentiments Analysis" (Di Capua et al., 2015), ou même la détection des événements (Gaglio et al., 2016).

Dans notre projet, il est question de suivre la propagation des opinions en temps-réel sur le réseau social Twitter, et ce en utilisant l'outil Storm de la fondation Apache, puis d'utiliser des méthodes de visualisation de données afin d'obtenir une représentation dynamique des résultats obtenus.

2 Plateforme de suivi des opinions en Temps réel

Pour la conception de notre plateforme, nous avons opté pour l'utilisation d'une approche basée dictionnaire afin de déterminer la polarité et l'intensité d'une opinion dans un tweet et pour les pré-traitements sur les textes.

2.1 Module d'acquisition des tweets

Ce module est chargé de récupérer des tweets selon des mots-clés, en continu, puis de les transmettre au reste du système. Les tweets doivent être filtrés selon la langue.

2.2 Module de pré-traitement des tweets

Ce module est précédé par une étape de tokenisation qui va séparer les tokens des émojis et émoticônes. Ensuite ce module va effectuer trois étapes dans son processus : La correction lexicale, syntaxique et la normalisation des tokens. La Correction lexicale permet de faire une correction de l'argot, remplacer les acronymes et permettre aussi de reconnaître les expressions de langue. Dans ce module, les statuts sont étiquetés morpho-syntaxiquement. L'étiquetage permet de dissiper l'ambiguïté autour d'un terme pouvant avoir plusieurs étiquettes.