

Nom et Prénom : CHINY MOHAMED

Date de soutenance : 14/01/2023

Directeur de Thèse : CHIHAB YOUNES

Sujet de Thèse :

Évaluation de l'e-réputation à travers les réseaux sociaux : une approche Machine Learning et NLP

Résumé:

L'objectif principal de ce travail consiste à proposer un système d'analyse de sentiments capable de jauger l'e-réputation et l'image de marque d'une entité sur les canaux du microblogging, en l'occurrence, les réseaux sociaux et les plates formes de partage d'avis. En effet, dans la littérature existante, de nombreux modèles sont proposés dans ce sens. Cependant, la grande majorité de ces modèles-là reposent sur les algorithmes d'apprentissage supervisé qui requièrent une phase d'entraînement. Lors de cette phase, une masse colossale de données étiquetées sont consommées afin d'atteindre des résultats de prédiction satisfaisantes. Or, la préparation des données d'entraînement étiquetées de qualité requièrent beaucoup de temps de main d'œuvre. Dans notre travail, nous avons proposé un modèle d'analyse de sentiments hybride qui repose sur les réseaux de neurones récurrents Long Short-Term Memory (LSTM), le lexique Valence Aware Dictionary and sEntiment Reasoner (VADER) et l'algorithme de pondération Term Frequency-Inverse Document Frequency (TF-IDF). Ces trois modèles ont été combinés dans un modèle de classification où nous avons implémenté chacun des 5 algorithmes suivants : Logistic Regression, k-Nearest Neighbors, Random Forest, Support Vector Machine et Naive Bayes. Le système proposé a été entraîné sur un petit volume de données étiquetées et a permis d'atteindre des métriques de performances satisfaisantes, en l'occurrence, l'exactitude et le score F1. Outre les bonnes performances enregistrées, le système a dévoilé sa capacité de transférer les connaissances apprises afin de traiter des données provenant d'un champ autre que celui qui a délivré les données d'entraînement. Nous avons également étudié l'impact des influenceurs sur la réputation des entreprises dans les pays en voie de développement. Comme cas d'étude, nous avons suivi un influenceur célèbre au milieu des jeunes et qui a été recruté par une grande entreprise marocaine afin de promouvoir l'un de ses événements phares sur les réseaux sociaux. En d'autre part, nous avons suivi une approche axée sur l'évaluation de la réputation vis-à-vis de la satisfaction des clients. En effet, nous avons contribué à déterminer les identifiants de satisfaction dans le domaine du tourisme et plus particulièrement dans le secteur d'hébergement P2P, en nous servant des techniques de fouille de données, l'apprentissage automatique et le traitement du langage naturel. A la fin, nous avons proposé un système simplifié de recommandation des programmes dans une plateforme de streaming et mené une étude qui vise à analyser les sentiments des usagers des réseaux sociaux à l'égard de la pandémie COVID-19.

Abstract :

The main objective of this work is to propose a sentiment analysis system capable of gauging the e-reputation and the brand image of an entity on microblogging channels, in this case, social networks and opinion sharing platforms. Indeed, in the existing literature, many models are proposed in this sense. However, the vast majority of these models are based on supervised learning algorithms that require a training phase. During this phase, a huge amount of labeled data is consumed in order to reach satisfactory prediction results. However, the preparation of quality labeled training data requires a lot of manpower. In our work, we proposed a hybrid sentiment analysis model based on Long Short-Term Memory (LSTM) recurrent neural networks, Valence Aware Dictionary and sEntiment Reasoner (VADER) lexicon and Term Frequency-Inverse Document Frequency (TF-IDF) weighting algorithm. These three models were combined into a classification model where we implemented each of the following 5 algorithms: Logistic Regression, k-Nearest Neighbors, Random Forest, Support Vector Machine and Naive Bayes. The proposed system was trained on a small volume of labeled data and achieved satisfactory performance metrics, in this case, accuracy and F1 score. In addition to the good performance recorded, the system revealed its ability to transfer the knowledge learned to process data from a field other than the one that delivered the training data. We also studied the impact of influencers on the reputation of companies in developing countries. As a case study, we followed a famous influencer among young people who was recruited by a large Moroccan company to promote one of its flagship events on social networks. On the other hand, we followed an approach focused on the evaluation of reputation towards customer satisfaction. Indeed, we contributed to determine the satisfaction identifiers in the tourism domain and more specifically in the P2P accommodation sector, using data mining techniques, machine learning and natural language processing. At the end, we proposed a simplified system of program recommendation in a streaming platform and conducted a study that aims to analyze the feelings of social network users towards the COVID-19 pandemic.