

# Modèles prédictifs pour la détection précoce des canulars diffusés sur Twitter

Didier Henry\*, Erick Stattner\*

\* LAMIA, Université des Antilles, Guadeloupe, France  
didier.henry ; erick.stattner@univ-antilles.fr

## 1 Introduction

De nos jours, les médias sociaux sont largement utilisés quotidiennement pour accéder aux nouvelles. Malheureusement, ces plateformes utilisées par des millions de personnes ne sont pas à l'abri de la désinformation. Les rumeurs de la mort des célébrités sur les médias sociaux se propagent très largement dans un court laps de temps et sont difficilement vérifiables. Ces types de rumeurs peuvent conduire à des situations inquiétantes ou stressantes. Dans ce travail, nous avons abordé le problème de détection des canulars sur la mort de célébrités dans les médias sociaux.

## 2 Collecte et extraction de données

Notre objectif est d'observer et de comprendre les différences entre les vraies et les fausses rumeurs de mort de célébrités au niveau de la diffusion, des tweets et des utilisateurs. Le principal défi que nous avons rencontré était d'accéder aux tweets sur ces rumeurs. Grâce au moteur de recherche de Twitter<sup>1</sup> nous avons collecté un grand nombre de tweets liés à 25 rumeurs relatives à la mort de célébrités, dont 13 réelles (c'est-à-dire que la célébrité est vraiment morte) et 12 fausses (c'est un canular sur la mort de la célébrité).

Dans un premier temps, nous avons pris en compte plusieurs caractéristiques de diffusion au moment  $t_n$  (le nombre de tweets postés, le nombre de retweets publiés, le pourcentage de tweets publiés, le pourcentage de retweets publiés, le pourcentage de nouveaux messages, un booléen égal à vrai si la rumeur se propage aussi en français et en espagnol en moins d'une heure)

Dans un deuxième temps, nous nous sommes intéressés aux caractéristiques des messages. Tout d'abord, nous effectuons une analyse du contenu des messages au moment  $t_n$  (le nombre moyen de caractères des messages, le pourcentage de messages contenant une URL et une URL distincte (pointant vers des sites autres que Twitter), le pourcentage de messages contenant l'acronyme RIP, le pourcentage de messages avec au moins un smiley triste, etc.). Deuxièmement, nous avons utilisé TextBlob<sup>2</sup> une API en langage Python afin d'extraire les sentiments

1. <https://twitter.com/search-advanced>

2. <http://textblob.readthedocs.io/en/dev/>

des messages en termes de polarité et de subjectivité. Troisièmement, nous avons utilisé un outil<sup>3</sup> pour extraire les émotions exprimées dans les messages.

Dans une troisième étape, nous nous sommes intéressés aux caractéristiques des utilisateurs. Étant donné que nous avons collecté des messages datant de plusieurs mois ou années et que le réseau Twitter est très évolutif, nous avons choisi d'extraire certains attributs au moment  $t_n$  (le ratio moyen du nombre de tweets publiés sur le nombre de jours écoulés depuis la création du compte, le ratio moyen entre le nombre followers et le nombre followings, le ratio moyen entre le nombre de tweets publiés et le nombre de followers, etc.)

### 3 Modèles prédictifs de canulars

Dans l'objectif de prédire, nous avons utilisé la cross-validation et comparé les résultats obtenus avec 4 différents types de classifieurs : Réseaux bayésiens (BN), Arbre de décision (RF), Machine à vecteurs de support (SVM) et Perceptron multicouche (MP). L'objectif est de comparer le temps nécessaire pour identifier les canulars selon le classifieur et le type d'attributs utilisés. La Figure 1 présente les résultats obtenus quand les attributs utilisés concernent (a) la diffusion, (b) les tweets, (c) les utilisateurs et (d) tous les attributs sont considérés.

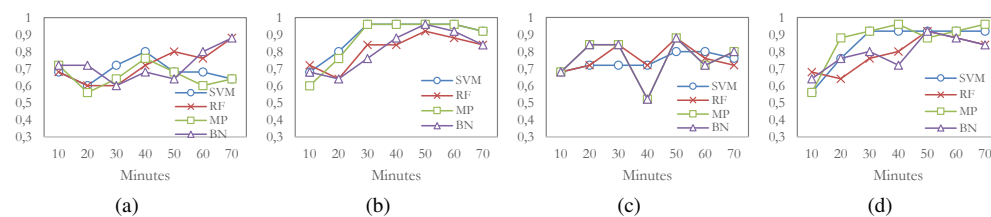


FIG. 1: Évolution du taux de classification correcte au cours du temps, quand les attributs utilisés concernent (a) la diffusion, (b) les tweets, (c) les utilisateurs et (d) tous

Le perceptron multicouche et les machines à vecteurs de support présentent les meilleurs résultats et atteignent environ 90% de classifications correctes dès la 30<sup>e</sup> minute après le démarrage du processus de diffusion.

### 4 Conclusion et perspectives

Dans ce travail, nous avons abordé le problème de détection des canulars de mort de célébrités sur les médias sociaux. En utilisant un perceptron multicouche, nous avons obtenu un taux de vrai positif à 0,9 seulement 20 minutes après le début de la diffusion de la rumeur. En perspectives, nous prévoyons d'observer les canulars dans d'autres langues et de prendre en compte d'autres caractéristiques afin d'améliorer les résultats de prédiction.

3. <https://github.com/nikiicc/twitter-emotion-recognition>