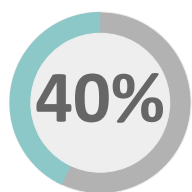




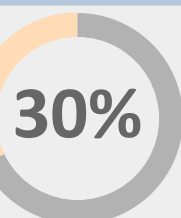
Maîtrisez votre e-reputation grâce à l'IA

3 approches pour anticiper le « bad buzz » sur twitter

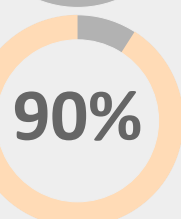


des entreprises estime que l'e-réputation représente le risque numéro 1 pour leur image

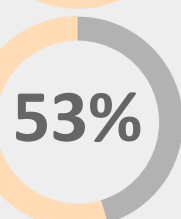
Pourquoi évitez le « bad buzz » ?



des bad buzzs sont aussitôt relayés par les médias internationaux dans l'heure qui suit la première annonce.



des crises durent moins de 72 heures mais peuvent laisser des traces indélébiles.



des entreprises touchées par un bad buzz n'ont pas retrouvé leur cours de bourse initial 1 an plus tard

'Il faut 20 ans pour bâtir une réputation
et cinq minutes pour l'anéantir'
- Warren Buffet -

L'agence MIC vous propose aujourd'hui un tour d'horizon sur 3 approches permettant d'identifier les tweets portant un sentiment négatif vous permettant ainsi d'engager une conversation, d'empêcher un éventuel bad buzz et de transformer une menace en opportunité !

Objectifs et méthodes

Notre but étant d'analyser les sentiments d'un tweet en utilisant les outils d'intelligence artificielle fournis par Microsoft Azure, nous avons retenu les trois approches suivantes

1. API sur étagère

Cette approche utilise un réseau de neurone profond existant mis à disposition par le service cognitif de Microsoft. Le réseau est pré-entraîné, une simple requête vers l'API Microsoft suffit. La mise en œuvre de cette approche est très rapide, cependant l'architecture du réseau et le jeu d'entraînements sont imposés.

2. Interface Graphique

Les différentes étapes de traitement des données, d'entraînement et d'évaluation des modèles sont représentées dans une GUI. Il suffit de connecter entre eux les éléments et de lancer les calculs. Des connaissances en Data Science et Machine Learning sont nécessaires. Aucune programmation n'est requise. Aucun algorithme d'apprentissage profond n'est disponible. Vous maîtrisez les jeux de données.

3. ML Studio

Dans ce cas, de fortes compétences en programmation sont nécessaires. Le concepteur a alors le contrôle sur toute la chaîne de traitement décisionnelle. Dans notre cas, nous avons utilisé des réseaux de neurones récurrents à mémoire. Le modèle final est servi sur le cloud. Cette approche est la plus complexe mais aussi celle qui vous donne le plus de possibilités et les meilleures performances.

L'expérience

Pour les approches 2 et 3, nous avons entraîné les modèles avec 400'000 tweets extraits de la base de données Twitter open source Kaggle. Les détails des pré-traitements et des plongements de mots sont visibles sur le GitHub donné en lien. Pour établir une base de comparaison, nous avons isolé 1600 tweets de la base de données puis nous avons évalué les performances des 3 modèles sur ce jeu de données.

Résultats

Avec un jeu de données équilibré, nous avons utilisé l'exactitude (accuracy) comme métrique de comparaison, elle représente le pourcentage de prédictions correctes faites par le modèle.

71 %

74 %

80 %

Comparatif

	API sur étagère	Interface Graphique	ML Studio
Accessibilité pour néophytes	★★★★	★★	★
Économie à l'exploitation	★	★★★	★★
Temps de mise en œuvre	★★★★	★★★	★
Performance	★	★★	★★★★

Conclusion

Votre choix se portera vers une approche ou une autre en fonction de vos besoins et de vos ressources. Pour un projet pilote, préférez une approche type API sur étagère, rapidement vous aurez un produit à présenter et à mettre en production, mais les coûts liés aux services seront élevés.

Si vous souhaitez avoir le contrôle sur votre modèle de prédiction, la possibilité de le faire évoluer et de le maintenir, préférez alors une approche ML Studio que vos équipes de Data Scientist sauront mettre en œuvre. Enfin, si vous n'avez pas de compétences fortes en science des données dans votre organisation mais que vous souhaitez tester plusieurs modèles avant d'investir, tournez-vous vers la solution interface graphique.

Plus de détails et le code:

<https://github.com/yhereng/OpenClassrooms/tree/master/OC-P7-Sentiment%20Prediction>

Sources :

(1) Selon l'éditeur Digimind (www.digimind.fr)

(2) Étude du cabinet Deloitte avec Forbes Insight réalisée au printemps 2013 auprès de 300 patrons et managers : http://www.deloitte.com/assets/Dcom-France/Local%20Assets/Documents/publications/1311%20Exploring%20Strategic%20Risk/Exploring_strategic_risk_nov2013.pdf

(3) <http://www.cadic-services.com/gerer-son-e-reputation/#.Uu9Gz3d5PxZ>

(4) Selon l'agence américaine de communication Digital Firefly.