

# APPRENTISSAGE AUTOMATIQUE POUR LA SURVEILLANCE DE MARQUES

J. Tytgat<sup>1,2</sup>, G. Wisniewski<sup>1</sup>, A. Bétrancourt<sup>2</sup>

<sup>1</sup> Université Paris Cité, LLF, CNRS 75013 Paris, France

<sup>2</sup> IPSIDE, 31100 Toulouse, France

# INTRODUCTION

- *Legaltech* : utilisation de techniques informatiques dans un cadre juridique
  - *Smart contrat*, classification automatique de brevets, recherche d'antécédents, etc.
  - Grands volumes de données structurées, libre d'accès
- Surveillance de dépôts de marques
- Une tâche de classification  $\approx$  similarité

# NOTION DE MARQUE

- Différentes natures (logo, hologramme, son, couleur...)
- Marque verbale : combinaison de caractères typographiques standards.
  - « *Yoplait* », « *Guy Degrenne* »,
  - un slogan « *Parce que vous le valez bien, des chiffres* »
  - des chiffres et/ou des lettres « *307* » (Peugeot)
- Enregistrée comme représentant des biens et/ou services spécifiques
  - Classification de Nice
  - Risque de confusion
  - Décision du juge (similitude des signes, identité des produits & notoriété de la marque)

# CRITÈRES DE SIMILARITÉ

- **Phonétique** Caloon c/ Kalon
- **Sémantique** Magic Hour c/ Magic Times
- **Graphie** Swile c/ Sm !le
- **Classification** Classe 32 : Bières, eaux minérales et gazeuses c/ Classe 33 : Boissons alcoolisées
  
- Mais aussi : la notoriété, la zone sur laquelle s'étend la protection...

# LA TÂCHE

- On ne prend pas en compte les critères extrinsèques
- Similarité  $\neq$  "Similarité juridique"
- Se situe en amont du travail des juristes
- $\Rightarrow$  Les faux négatifs sont critiques
- 
- Le corpus
    - $\sim$  70 000 exemples positifs issus d'un contexte de surveillance
    - $\sim$  700 000 exemples négatifs créés aléatoirement
- $\rightarrow$  Simuler une situation de surveillance
- Tâche de classification binaire

# CARACTÉRISTIQUES

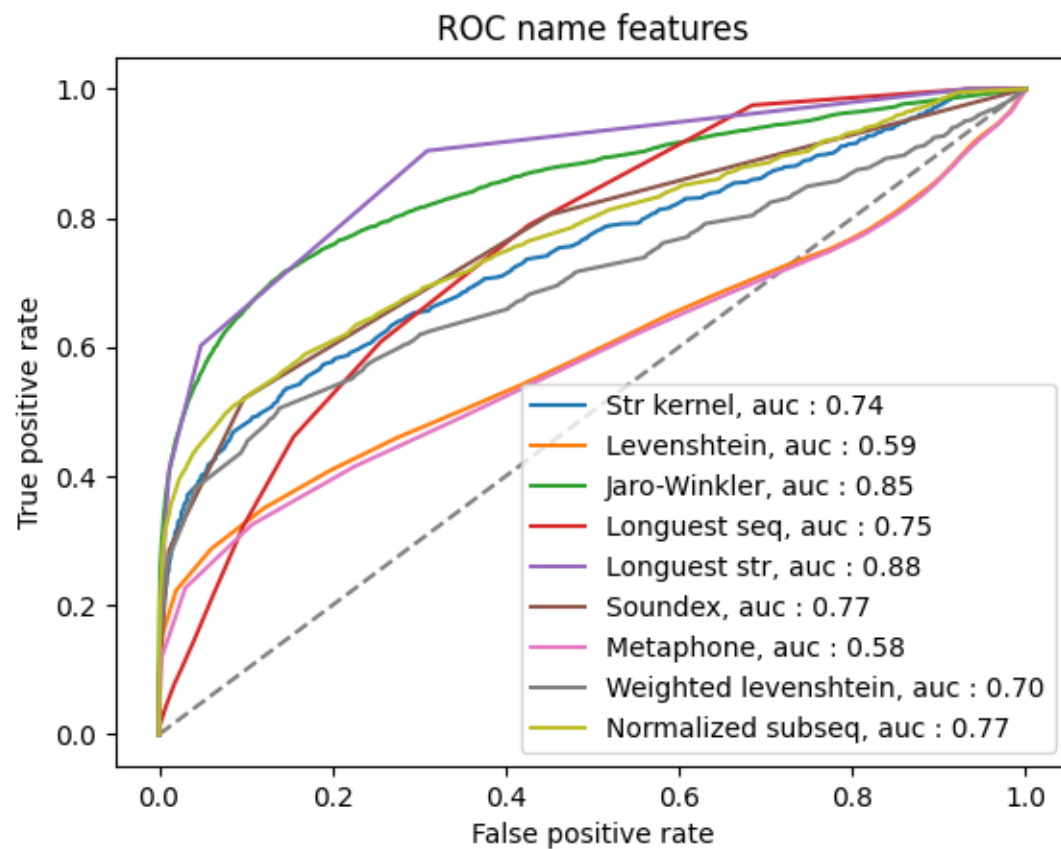
- Distance d'édition
  - Levenshtein, Levenshtein pondéré, Jaro-Winkler
- Similarité de string
  - String kernel, longest subsequence (normalized), longest string
- Similarité phonétique
  - Soundex, metaphone

+ un pré-processing pour la graphie

# ÉVALUATION

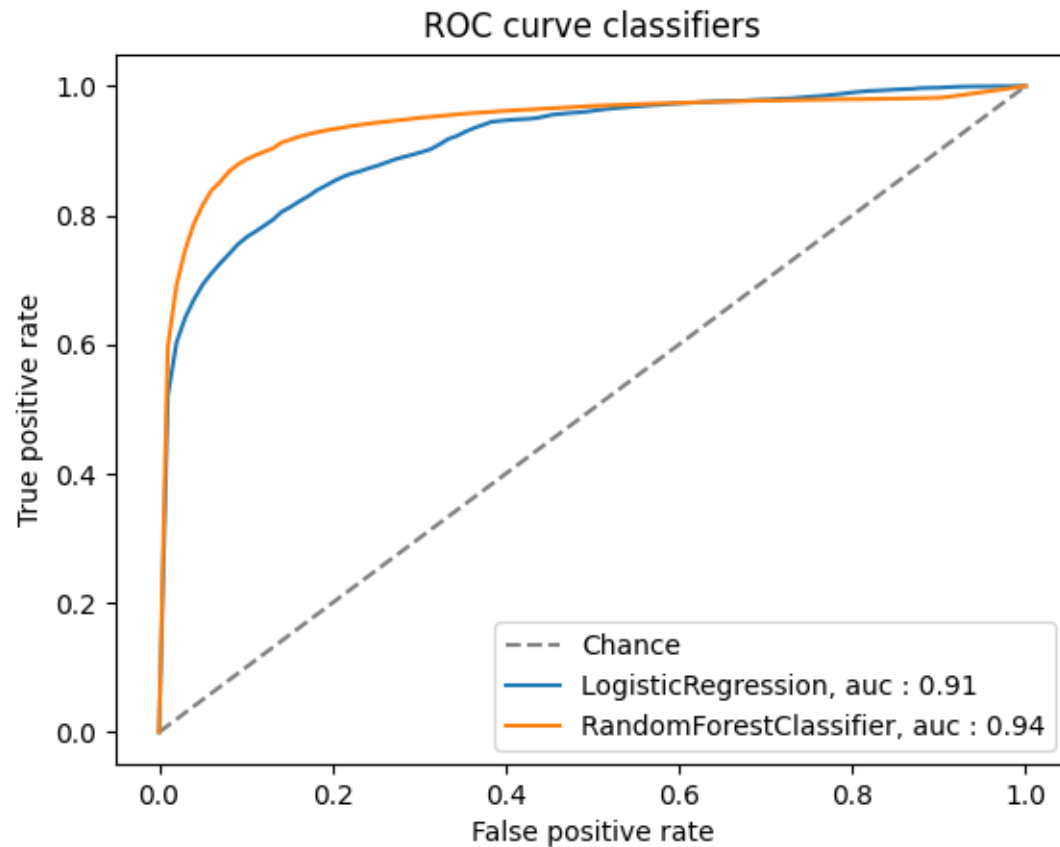
- Score F1
- Courbes ROC
- Optimisation standard : cross-validation + gridsearch

# COMPARATIF FEATURES





# MODÈLES D'APPRENTISSAGE



La régression logistique a obtenu un score F1 de 0.60 et la forêt aléatoire de 0.72 .

# DISCUSSION

- Mise en place protocole simulant une surveillance
- Analyse des différents algorithmes de similarité
- Combinaison de ces métriques au sein d'un système d'apprentissage
- Utiliser des méthodes d'apprentissage a bien un intérêt
- Travail sur les *features* reste ouvert
  - *Distinctivité, multilinguisme, meilleurs appréciations des produits et services...*
- Revoir l'évaluation
  - Faux négatifs restent critiques
  - Pérennité du système ?
  - Métriques garantissent-elles une bonne qualité des prédictions ?

# MERCI DE VOTRE ATTENTION

- Questions ?
- [j.tytgat@ipside.com](mailto:j.tytgat@ipside.com)
- [a.betrancourt@ipside.com](mailto:a.betrancourt@ipside.com)
- [guillaume.wisniewski@u-paris.fr](mailto:guillaume.wisniewski@u-paris.fr)