

Paris, 23/12/2019

## Sujet de mémoire

2ème année master big data et data science en finance

Etudiant: Haithem Bouhlel:

### Titre :

Méthodes de data science pour optimiser la gouvernance des données dans un environnement bancassurance

### Sujet :

Intégrer de l'intelligence artificielle et du text mining dans un outil de gouvernance de données pour automatiser le rapprochement entre différentes sources de données et métadonnées.

### Problématique :

Dans un environnement de transformation digitale, de recours de plus en plus fréquent à la méthode agile, et de réglementations contraignantes (SII et GDPR) l'entreprise doit accélérer sa capacité à documenter et partager les données au travers d'un dispositif centralisé d'une documentation des données.

Face au volume significative (1.5 M de métadonnées techniques) la gouvernance des données doit avoir recours à des techniques de data science pour industrialiser l'analyse et le rapprochement des données.

### **Plan détaillé :**

Au cours de mon alternance à la société générale assurance, j'ai eu l'occasion de contribuer à la transformation digitale qui transforme en profondeur les processus de l'assurance.

Afin d'acquérir une vision des données collectées, centralisées et entreposées, l'assurance implémente un outil de gouvernance de données pour faciliter l'accès à la connaissance de la donnée et ainsi répondre aux exigences des réglementations Solvabilité 2 et RGPD.

Cependant, de nombreuses données applicatives (techniques) sont difficilement explicables du fait d'une documentation non centralisée, voire inexistante.

L'exploitation de ces dernières nécessite alors l'intervention de l'expert métier ou des équipes projets pour documenter et rapprocher manuellement les données.

L'objectif de ce mémoire est de proposer des méthodes de data science pour automatiser ces actions et élaborer des différents modèles statistiques de text mining.

Le premier cas d'usage consiste à traiter les données du périmètre non-vie dommage, priorité dans le cadre du traitement des enjeux réglementaire de l'entreprise.

En première approche j'ai eu recours à des traitements d'analyse sémantique pour rapprocher des données techniques et métiers sur la base de leurs descriptions.

Dans un 2eme temps on va évaluer des méthodes basées sur les valeurs des données (la qualité des données et leur exhaustivité).

Nous présenterons ainsi les résultats des applications et mettrons en avant le potentiel d'une telle méthodologie, en prenant soin d'expliquer point par point, les améliorations des modèles et les principes mathématiques utilisés.

Signature



Haithem Bouhlel

23/12/2019

### **Plan détaillé :**

Au cours de mon alternance à la société générale assurance, j'ai eu l'occasion de contribuer à la transformation digitale qui transforme en profondeur les processus de l'assurance.

Afin d'acquérir une vision des données collectées, centralisées et entreposées, l'assurance implémente un outil de gouvernance de données pour faciliter l'accès à la connaissance de la donnée et ainsi répondre aux exigences des réglementations Solvabilité 2 et RGPD.

Cependant, de nombreuses données applicatives (techniques) sont difficilement explicables du fait d'une documentation non centralisée, voire inexistante.

L'exploitation de ces dernières nécessite alors l'intervention de l'expert métier ou des équipes projets pour documenter et rapprocher manuellement les données.

L'objectif de ce mémoire est de proposer des méthodes de data science pour automatiser ces actions et élaborer des différents modèles statistiques de text mining.

Le premier cas d'usage consiste à traiter les données du périmètre non-vie dommage, priorité dans le cadre du traitement des enjeux réglementaire de l'entreprise.

En première approche j'ai eu recours à des traitements d'analyse sémantique pour rapprocher des données techniques et métiers sur la base de leurs descriptions.

Dans un 2eme temps on va évaluer des méthodes basées sur les valeurs des données (la qualité des données et leur exhaustivité).

Nous présenterons ainsi les résultats des applications et mettrons en avant le potentiel d'une telle méthodologie, en prenant soin d'expliquer point par point, les améliorations des modèles et les principes mathématiques utilisés.

Signature



Haithem Bouhlel

23/12/2019