

Incident management process



30 janvier , 2020

AKOURY Hany

Descriptif

Il s'agit d'un journal des événements d'un processus de gestion des incidents extrait des données collectées à partir du système d'audit d'une instance de la plateforme ServiceNow™ utilisée par une société informatique.

Le journal des événements est enrichi de données chargées à partir d'une base de données relationnelle sous-jacente à un système d'information correspondant au processus.

Les informations ont été anonymisées pour des raisons de confidentialité.

Objectif :

**Prédire le temps restant avant
complétion d'un incident.**

Comprendre l'enjeu

Questions

- Comment prédire un temps? Une date ou une durée?
- Les attributs les plus intéressants ?

Strategie

- Feature Engineering : Adapter les colonnes à nos besoin (en rajouter - en améliorer)
- Création d'un modèle simple + premières évaluations du comportement des colonnes.

Sous Objectifs

- Modele simple verifié
- Amelioration du modele paramètres-optimisation
- Visualisations descriptives
- Création d'autres modèle
- Comparaison
- Choix du meilleur et conclusion

Comprendre le Dataset

Les variables à disposition

32 attributs sont à disposition:

- 1 case identifier
- 1 state identifier
- 32 attributs descriptifs
- 2 variables dependentes

Repartition des variables

Variables numeriques

- Reassignment_count
- Reopen_count
- sys_mod_count

3

Variables Catégorielles

Variables String/Bool:

- Incident_State
- Active
- ...
- Priority
- Closed Groupe

28

Variables Temporelles

Variables mises en forme:

- Opened_at
- Sys_created_at
- Sys_updated_at
- Resolved_at
- Closed_at

5

Proposed deliverables

Régression linéaire simple

- Premiers resultats.
- Vérification du comportement des colonnes

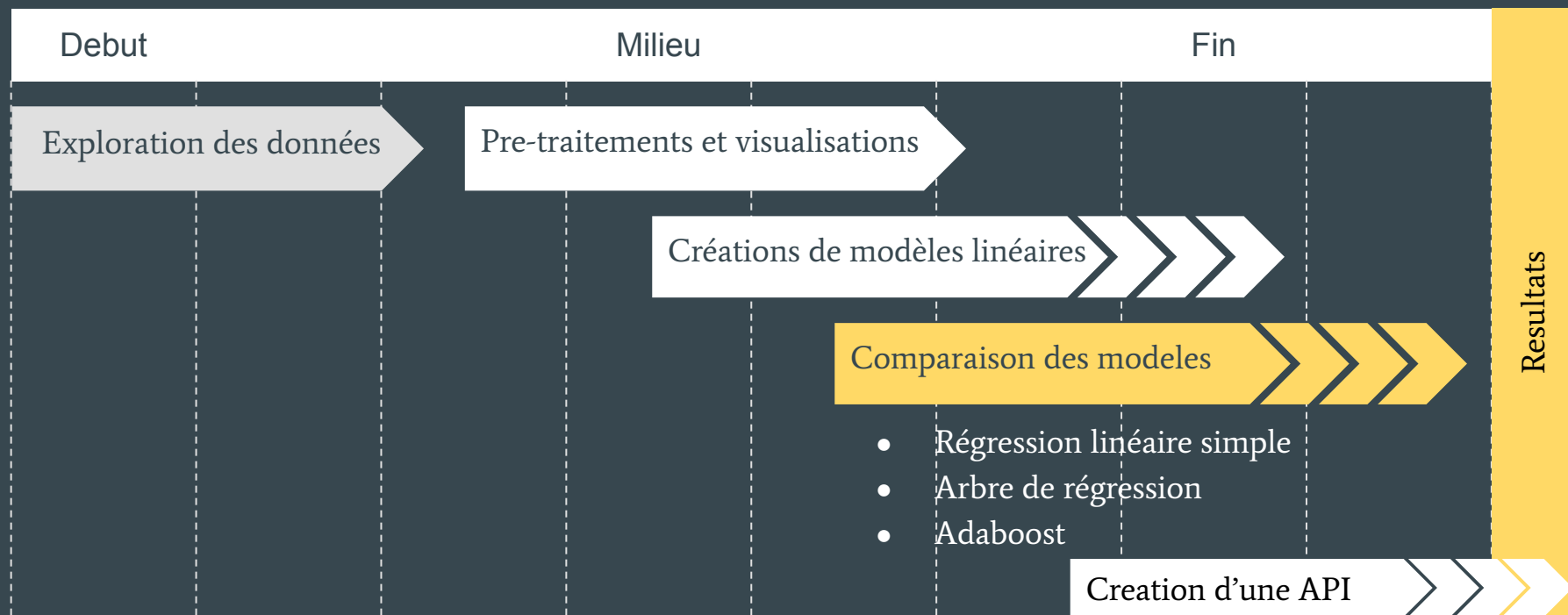
Arbre de Regression

- Parametrages plus avancés.
- Création du modèle pour pouvoir comparer.

AdaBoost

- Troisieme modele de regression
- But: comparaisons plus precises.

Méthode de travail adoptée tout au long du projet



Résultats

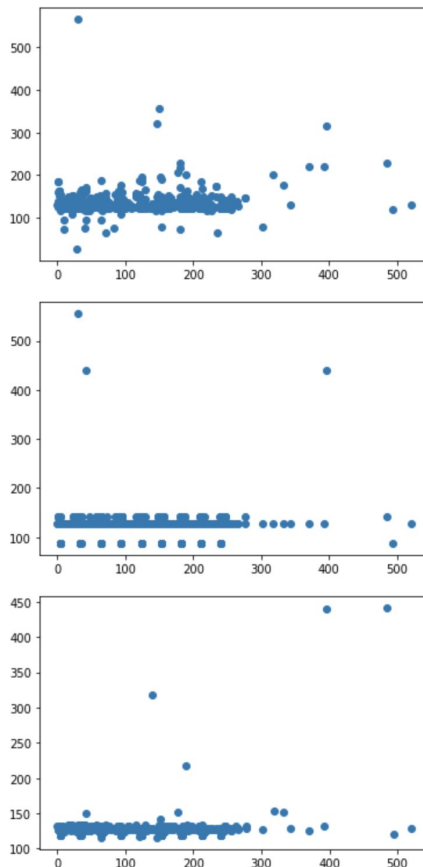
Resultats

Comme vous pouvez le voir sur le Github, lorsque les modèles sont créés testés et validés, on obtient un écart d'en moyenne 72 jours.

Cette erreur (RMSE) est relativement importante si on la compare aux durées présentes dans la colonne 'f_wait'.

Cela dit quasiment la même RMSE est trouvée pour les 3 modèles linéaires créées.

Le problème réside sûrement dans une étape préliminaire à la création du modèle.



```
LinearRegression  
['adjusted R2: -0.03791501069615921', 'rmse: 73.11362150703322', 'kfold-cross : Accuracy: Na']  
DecisionTreeRegr  
['adjusted R2: -0.03039510018148972', 'rmse: 72.84827830566391', 'kfold-cross : Accuracy: Na']  
AdaBoostRegresso  
['adjusted R2: 0.000156168736511475', 'rmse: 71.7601743899724', 'kfold-cross : Accuracy: Na']
```

Ouverture :

Suite aux résultats obtenus, il est important de se poser des questions par rapport aux données ainsi qu'à leur gestion et traitement, le problème réside sûrement la...

Merci de votre attention



AKOURY Hany

Work until you no longer
have to introduce yourself.