



Modèle de prévision de la durée des incidents sur les offres cloud d'OB

Pâquarse Mahouvi

Septembre 2022 – Août 2023, Alternant Data Analyst



Business

PLAN

1

PROBLEMATIQUE

2

OBJECTIF

3

METHODOLOGIE

4

RESULTATS

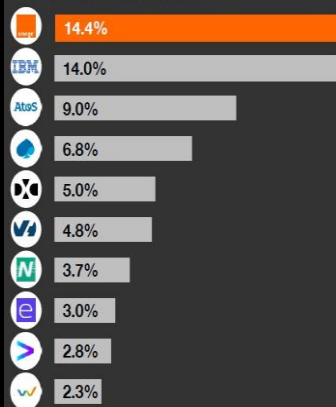
5

CONTINUITE

Private Cloud

#1 in France⁽¹⁾, #7 in Europe⁽²⁾

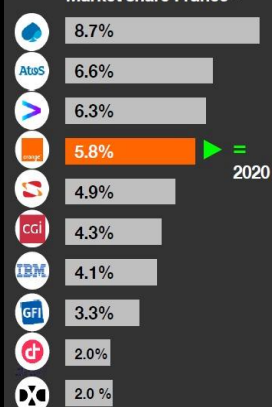
Market share France⁽¹⁾



Multi-Cloud Services

#4 in France⁽³⁾, #15 in Europe⁽³⁾

Market share France⁽³⁾



Intégrateur réseau et numérique de référence. C'est l'entité du Groupe Orange qui accompagne les entreprises et organisations dans leur transformation digitale.



Orange Business Services
Cloud infrastructures

Private Cloud

Cloud Avenue

Flexible Engine

New 2022

Sur les services cloud ...

Surviennent des incidents

Interruption non planifiée d'un service, la réduction de la qualité d'un service ou un événement qui n'a pas encore eu d'impact sur le service au client

Que faire ?

Rétablir le service convenu pour le client dans les plus brefs délais. C'est le Garantie de Temps de Rétablissement

Toujours respectés ?

Non (ou pas de GTR)

Dénomination

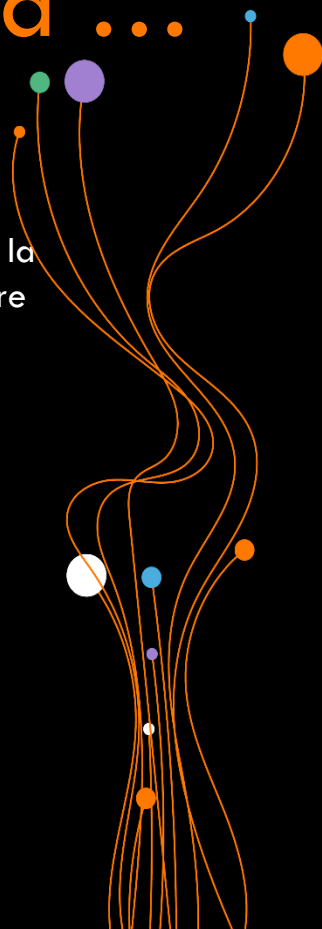
Non respect contractuel et donc des conséquences

Conséquences

Des pénalités pour orange Business calculées en fonction de l'offres et de la durée d'indisponibilité des services

Réflexion

Anticipé la durée des incidents



Objectifs



Outils de prévision de la durée des incidents

Modèle de Machine Learning capable de prédire la durée des incidents dès la détection



Optimiser la résolution des incidents

Outils de gestion des incidents permettant d'optimiser l'intervention des équipes

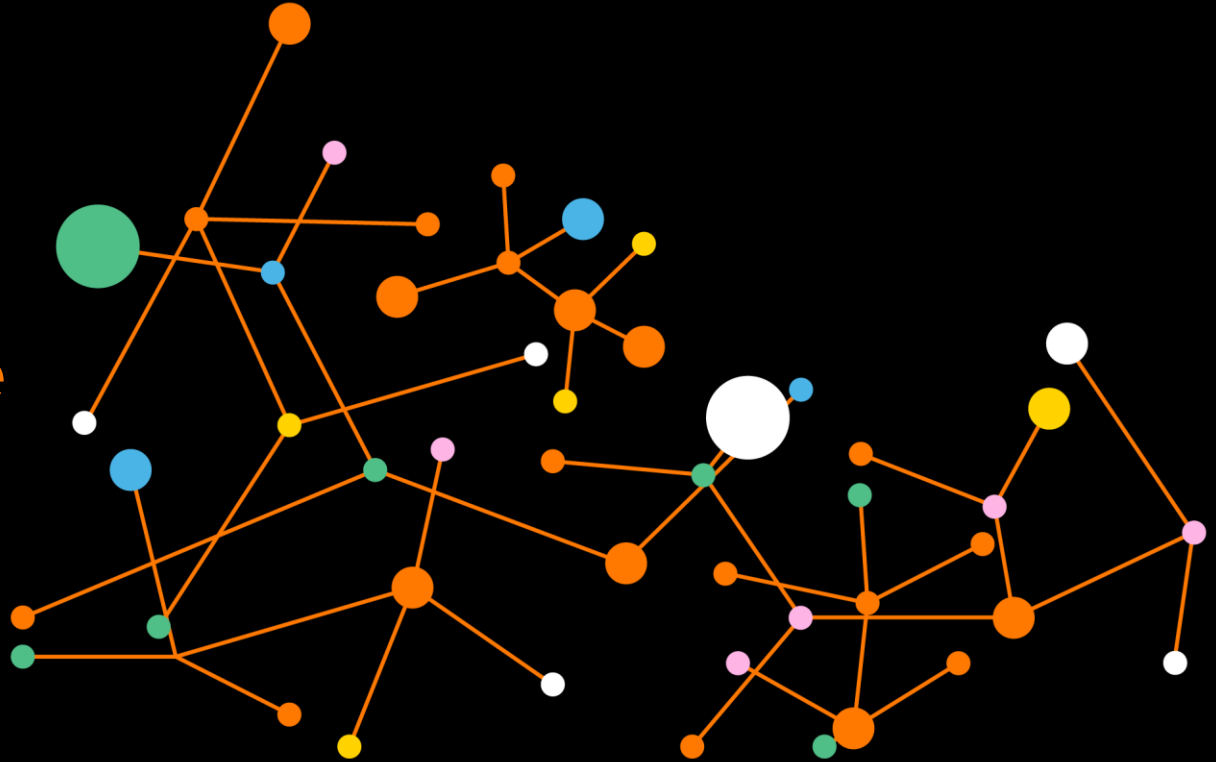


Réduire les pénalités

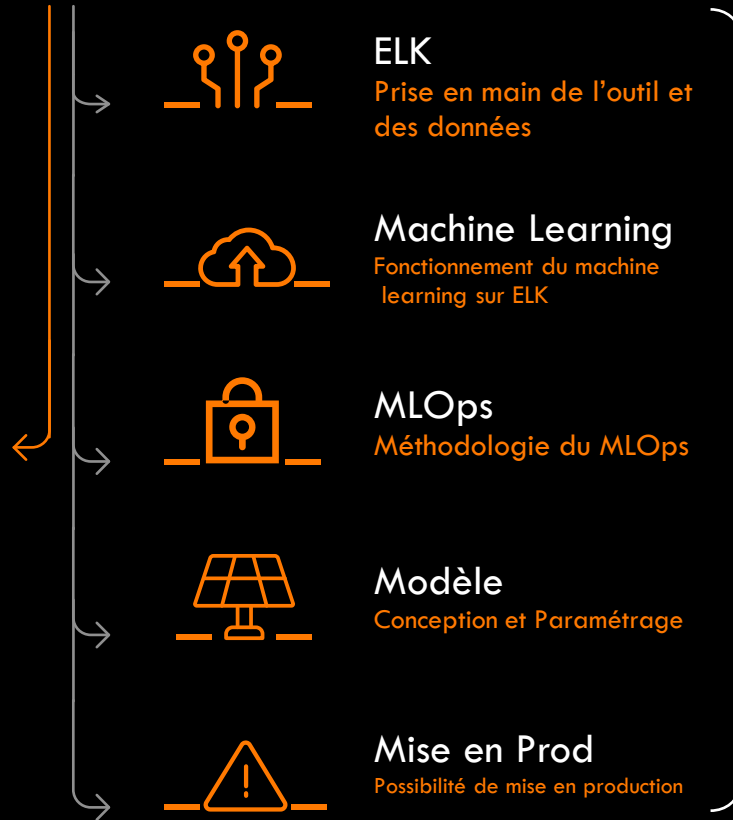
Il permettra sur le long terme d'être un outil de fidélisation de client et réduction des pertes

S'inscrit également dans le cadre d'une mise en production pour une utilisation sur le long terme
Réalisé de bout en bout sur la suite ELK (Elasticsearch-Logstash-Kibana), l'outil d'analyse et de traitement statistique de mon entité d'accueil.

Méthodologie

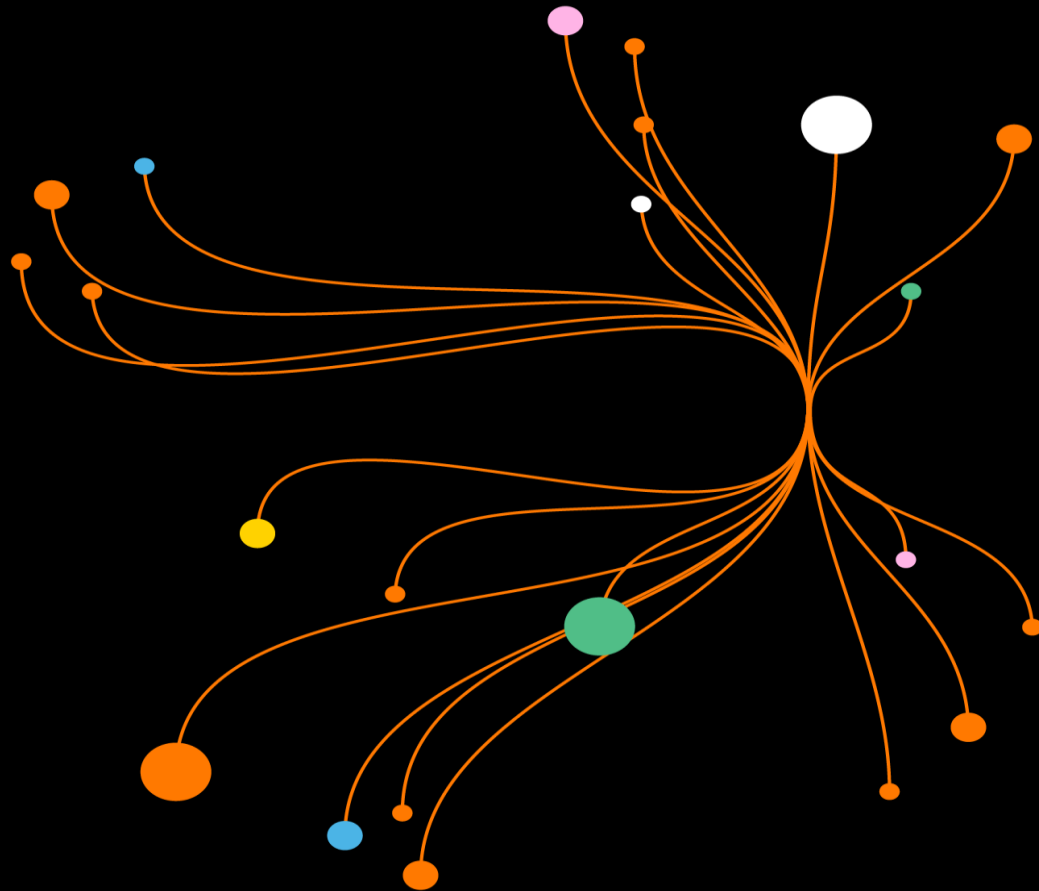


Pour atteindre les
objectifs de cette
mission



Résultats

Résultats



Résultats



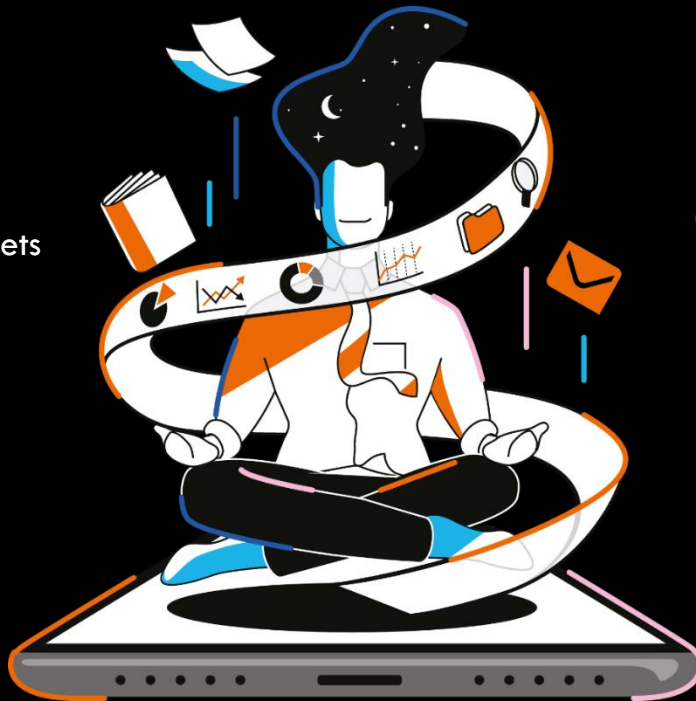
Index

Oceane_closed_tickets



99 897

Tickets Clôturés



Incident_duration

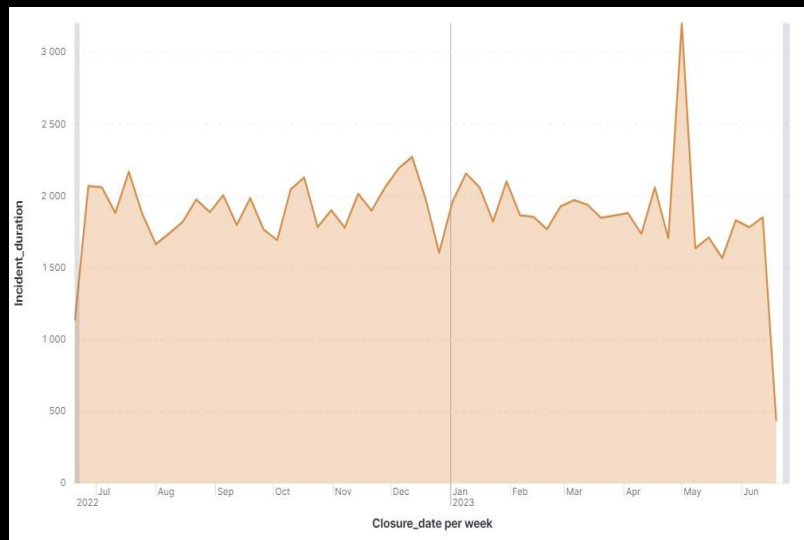
Sur les offres cloud



Kibana

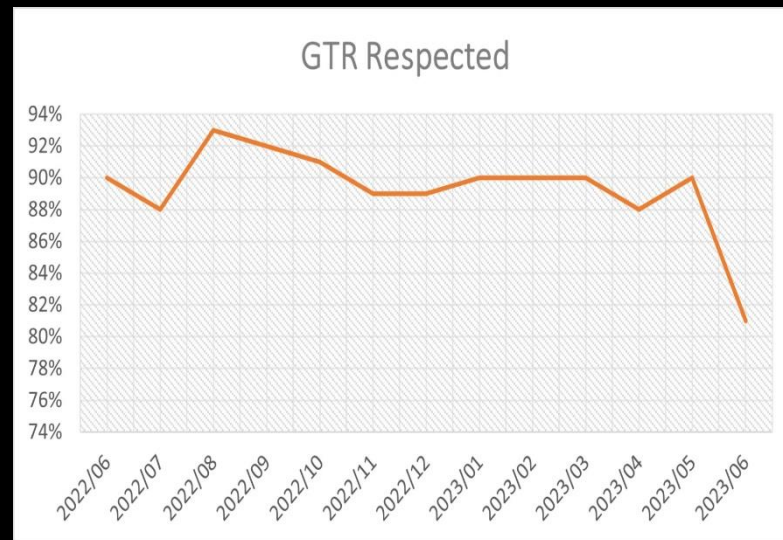
Outil de visualisation
de données

Résultats



Evolution de la durée moyenne des incidents

Source : Dashboard, ELK



Evolution du pourcentage de GTR Respecté

Source : Discover, ELK

Résultats

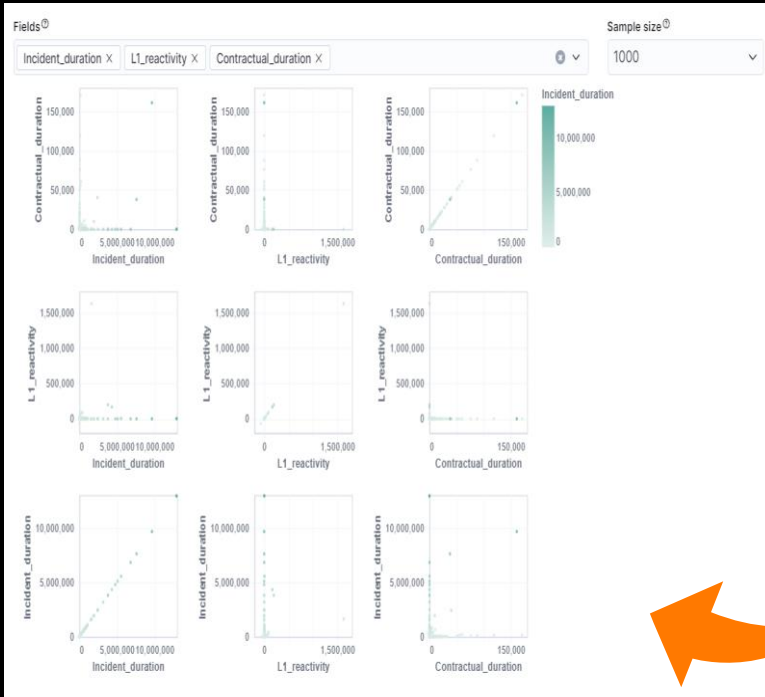
Analyse faite dans les limites des possibilités d'ELK

ELK n'est pas un outils permettant des tests statistiques avancées

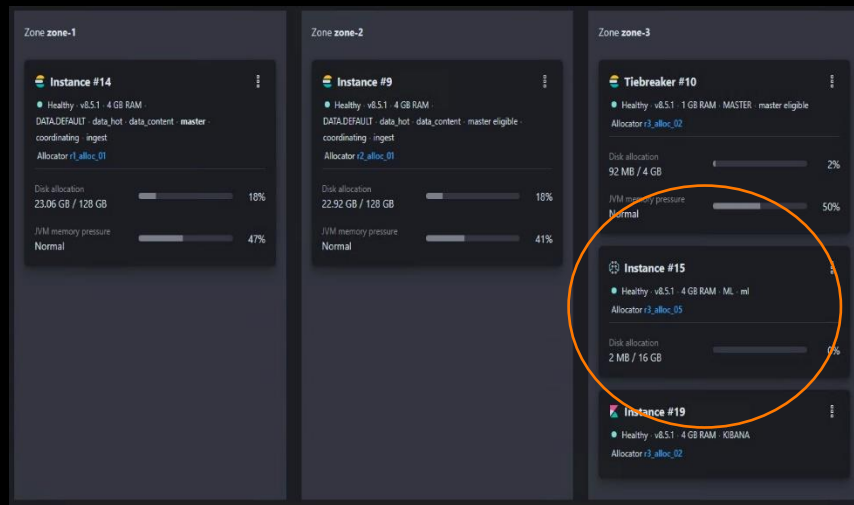
Discussions avec les SDM pour faire le choix des champs importants

Corrélation entre les fields numériques et long

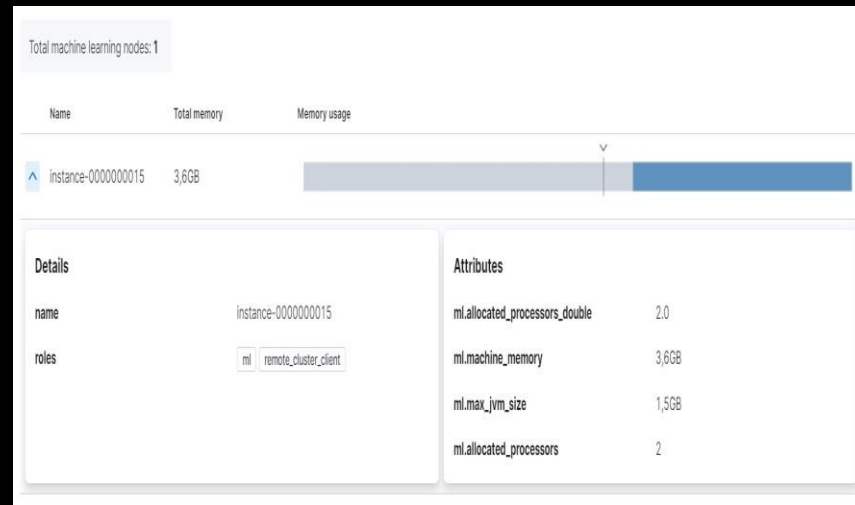
Echantillon de 1000 observations



Résultats



Instance de machine learning



Capacité de l'instance

Paramétrage du modèle

Sur l'instance ML

Capacité limitée de la
puissance de calcul

Très forte volumétrie des
données

10%

Des données utilisés

Résultat concluant compte
tenu des données utilisées

Modèle XGBoost d'ELK
(unique)

Paramétrage du modèle

171 000

Erreur Pseudo Huber Training

168 000

Erreur Pseudo Huber
Generalisation

36

Max trees

2

Max Optimisation

10,16

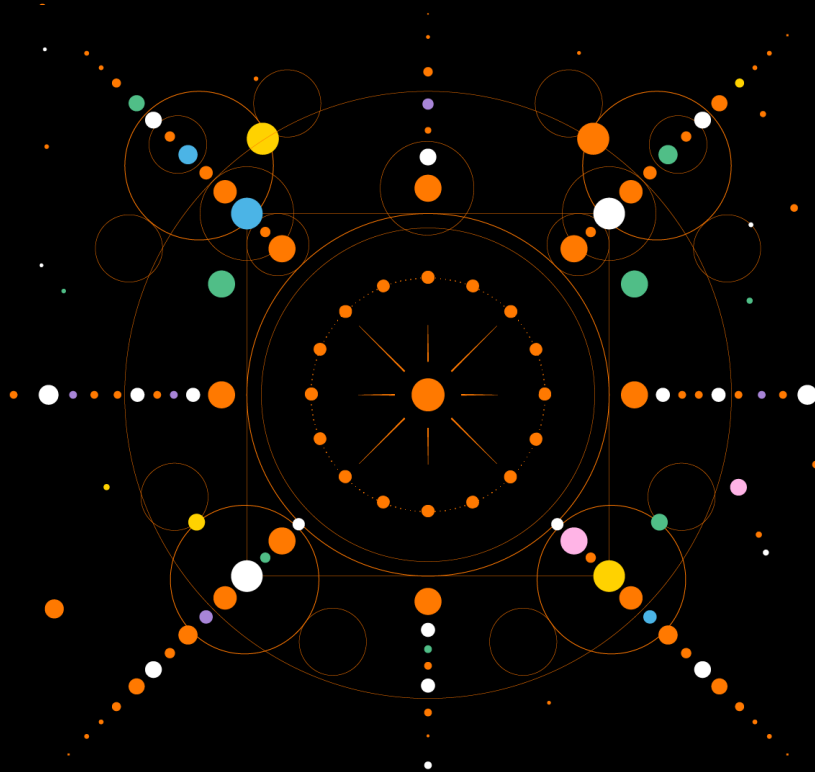
Soft tree depth limit

0,15

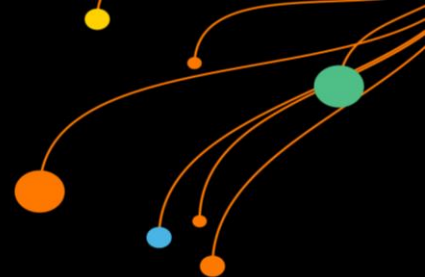
Soft tree depth tolerance

Modèle

XGBoost dans ELK



Difficultés & Limites



Pourcentage de données

Au delà de 15% des données, le cluster tombe. Ré-entraîner sur plus de données



Puissance de calcul

Cluster avec plus de ressources pour améliorer la rapidité d'exécution des modèles



Analyse de la cause profonde

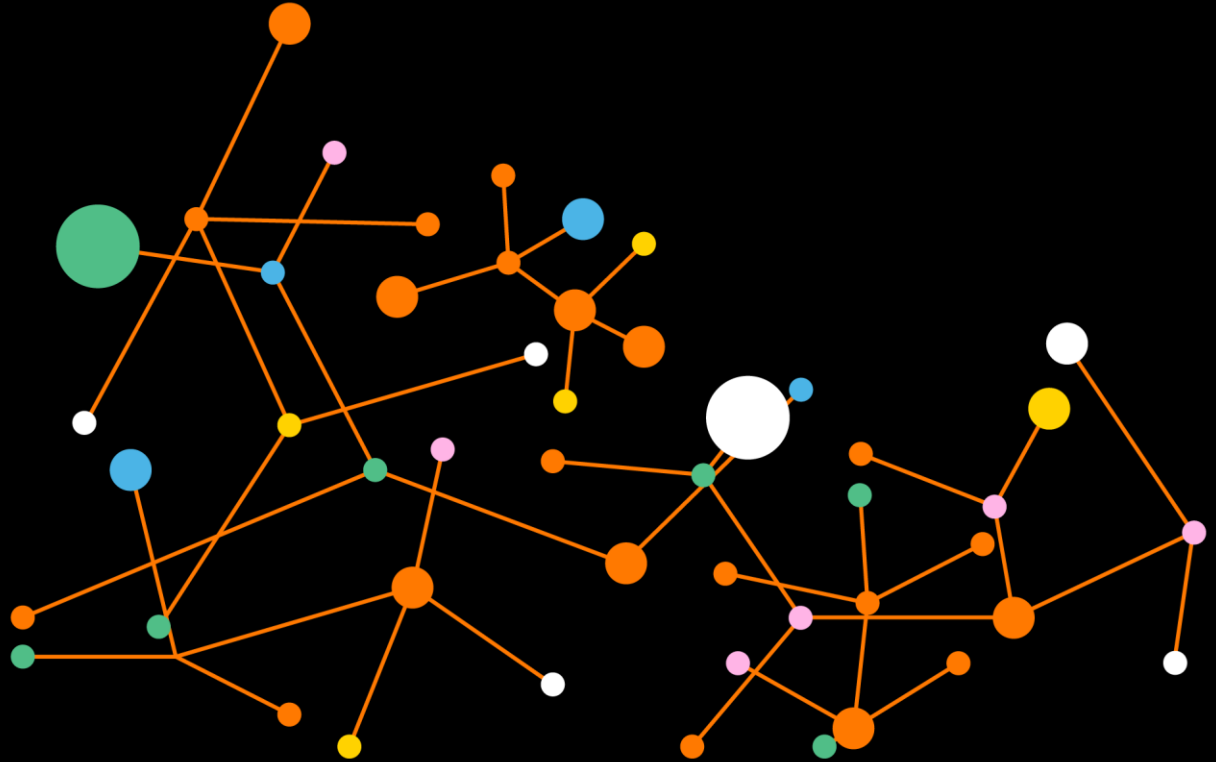
Corrélations entre les incidents et analyse de la cause profonde



Mise en production

ELK est également un outil de mise en production des modèles

Pour la suite



... Pour aller **plus loin**

Pipeline de préparation des données

Mettre en place le pipeline permettant d'adapter les mappings des données

Logstash

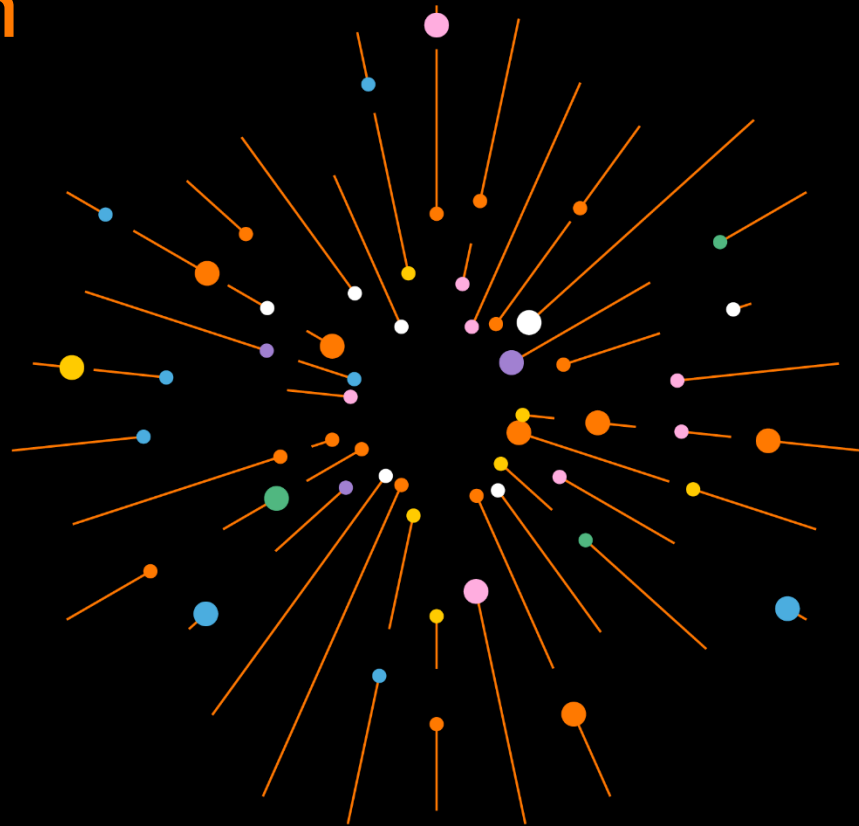
Pré-traitement des données et intégration du modèle

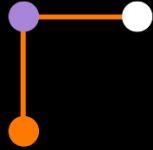
Déploiement

Tests approfondis avant la mise en production

Surveillance

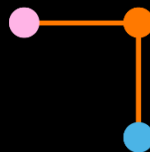
Contrôler régulièrement les performances





Complexité de la mise en application des techniques de MLOps sur des outils non spécialisé dans le déploiement continu

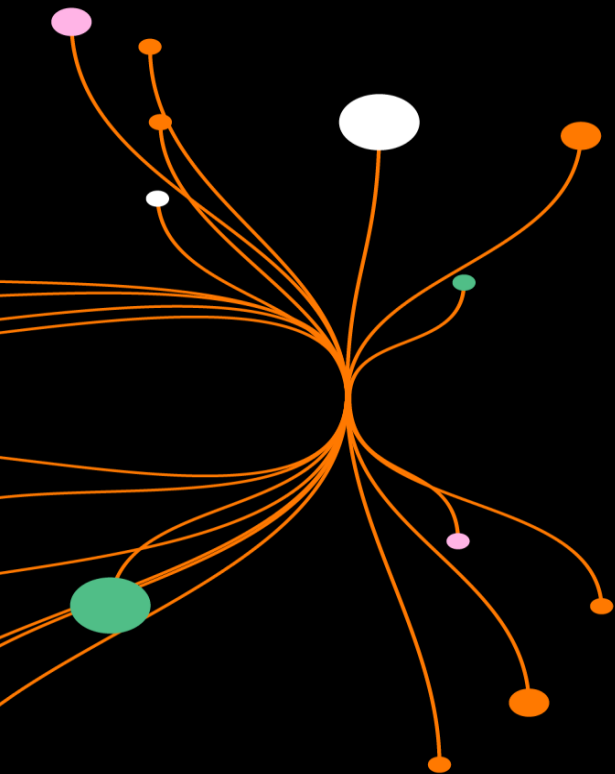




Nécessité de ressources et de puissance de calcul
pour l'utilisation de l'IA et du ML en entreprise

(Green IT)





Merci 🙏