

Anti-Fraude - Prédiction des alertes et auto-ajustement des seuils

Stagiaires: Matthieu DAVID, Thomas PILLOT

Encadrants: Valéry ADELEINE, Benoît GATIEN, Valentin FABIANSKI, Antoine BARRUOL

Période : 07/2020 - 03/2021



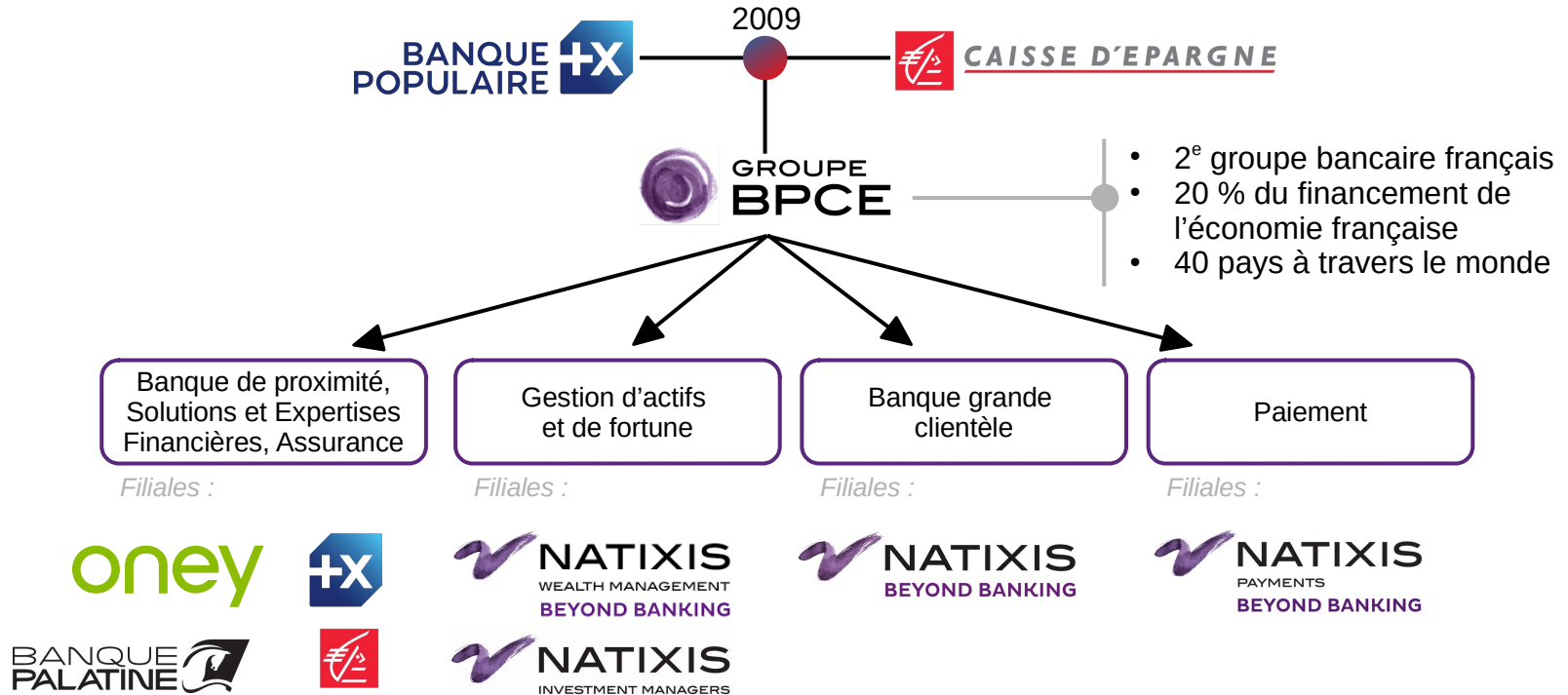
INTRODUCTION & CONTEXTE

1.1 présentation générale Natixis

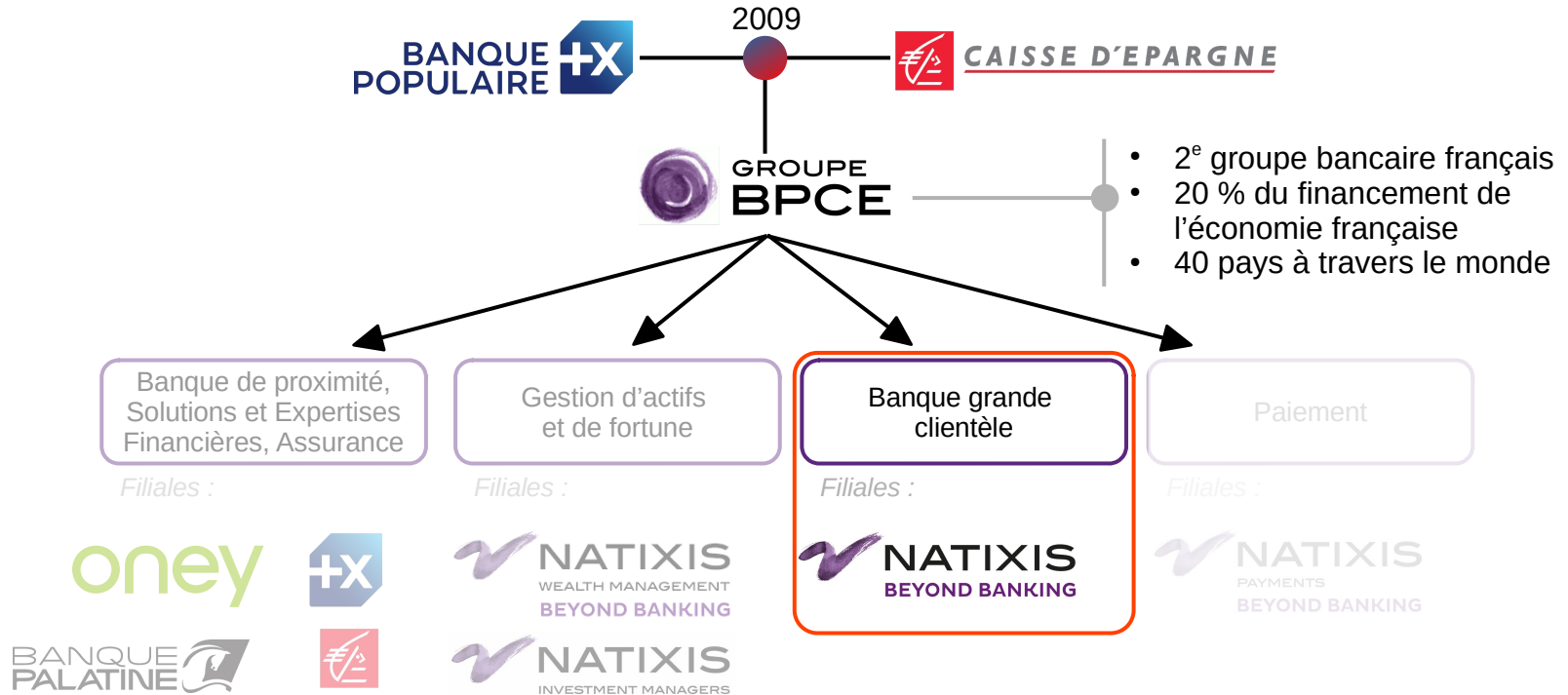
1.2 Écosystème Anti-Fraude à Natixis

1.3 Objectifs du stage

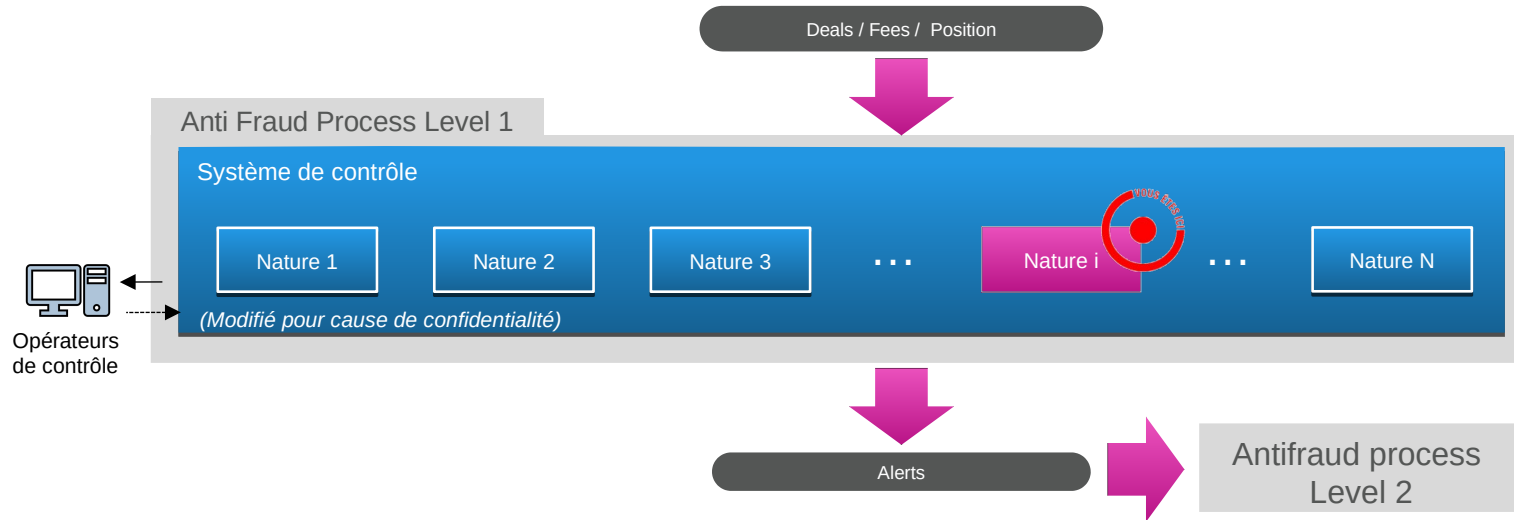
Présentation générale



Présentation générale



Ecosystème Anti-fraude à Natixis



Les missions

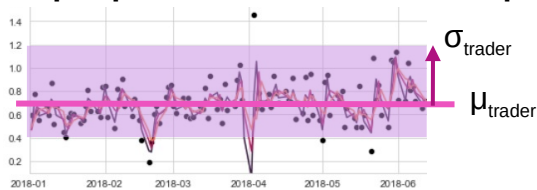
- Améliorer la pertinence des alertes.
- Améliorer la définition des seuils d'alertes.

Aujourd'hui

~50% des anomalies classées automatiquement.
(cas de POSNOTC – nombre de couples = 997.)

Alertes (05/2018 → 12/2020)	Classées auto	Traitées utilisateurs
23 625	12 240	11 385

Seuil fixe par période basé sur l'historique.

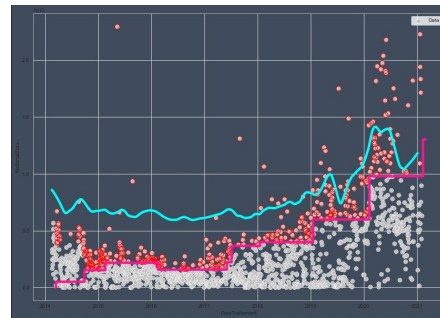


$$\text{Seuil} = \mu_{\text{trader}} + 2\sigma_{\text{trader}}$$

Demain

Réduction du “bruit” au sein des anomalies.

Seuils s'ajustant
“automatiquement”.



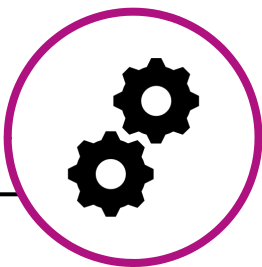
Fonctionnalité de prédiction et d'autoévaluation.
Prédire demain et dire si demain est atypique ou non.



OUTILS & MÉTHODES

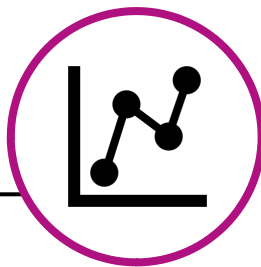
2.1 Les outils - BDD SQL & Python

2.2 Les méthodes - Isolation Forest & Time Series



 pandas

 NumPy



 matplotlib

 plotly



 SQLAlchemy

Thomas
Pillot

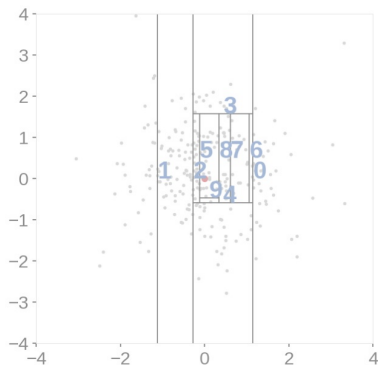
42

Isolation Forest

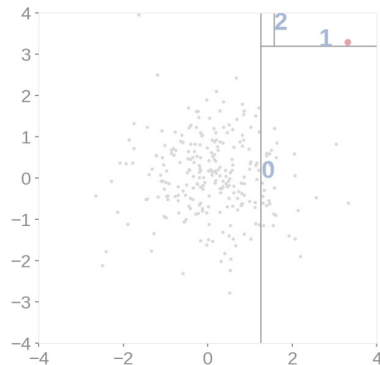
Une valeur normale est caractérisée par un nombre important de "*split*" pour l'isoler.

Une valeur atypique est caractérisée par un nombre faible de "*split*" pour l'isoler.

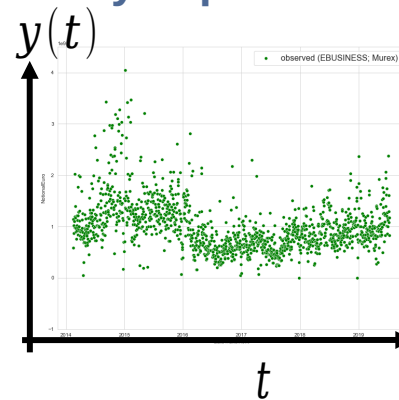
Valeur normal/nominal



Valeur atypique



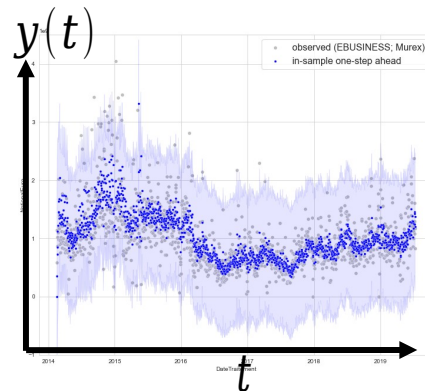
Analyse par série temporelle



- Analyse statistique
- Modélisation

Expression de y :

$$\hat{y}(t) = f(t, \hat{y}_{t-1}, \varepsilon_{t-i})$$





RÉSULTATS & DÉVELOPPEMENTS

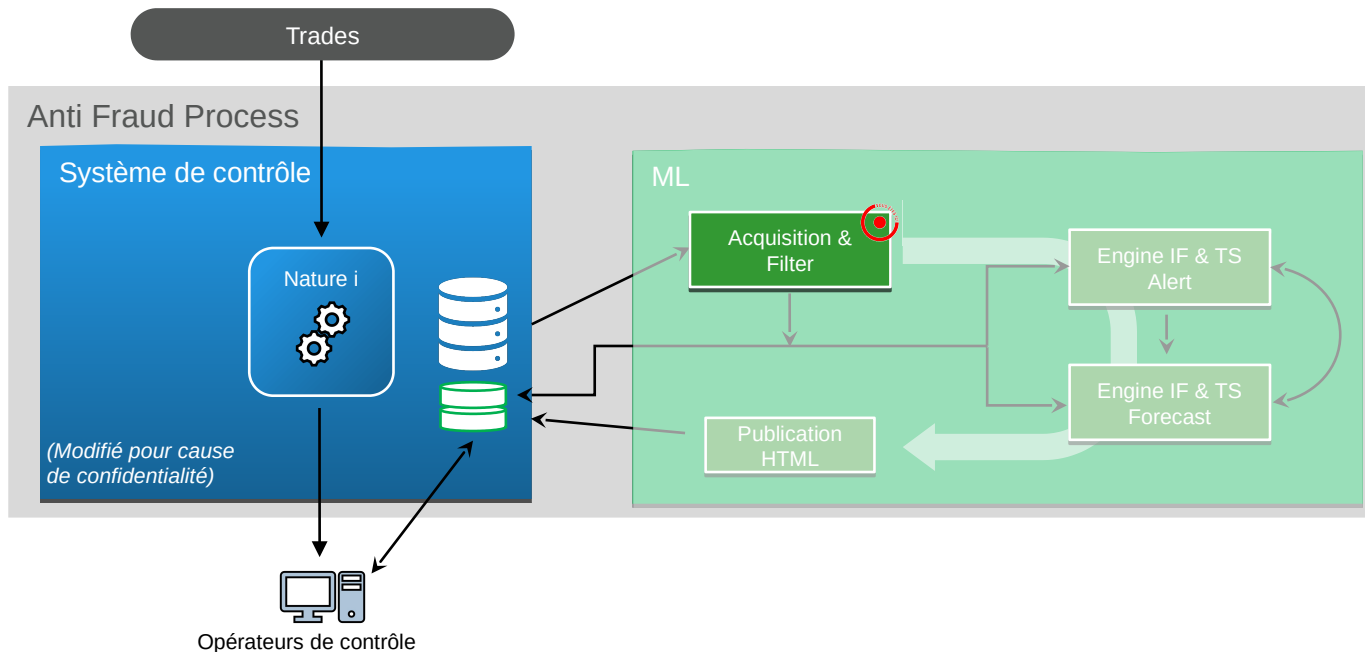
3.1 module 1 - Acquisition & Filtres

3.2 modules 2+3 - IF & TS - génération des alertes et prévision

3.3 module 4 - Visualisation

0. Vision opérationnelle globale





Rôles :

Acquisition des données brutes de la veille.

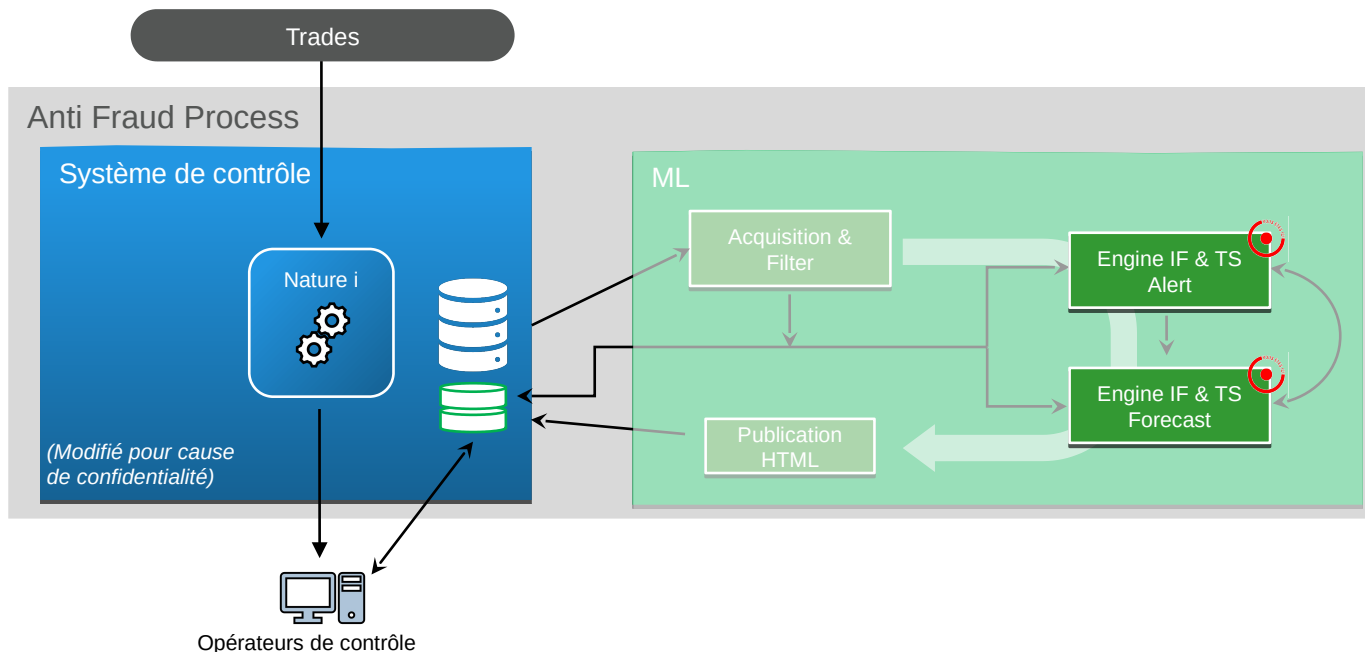
Filtrage et agrégation/transformation.

Stockage des données générées sur serveur SQL.

Outils :



3 | 2. modules 2+3 - IF & TS – Alertes et prévisions



Rôles :

Entraînement des modèles sur les données des couples nouvellement éligibles.

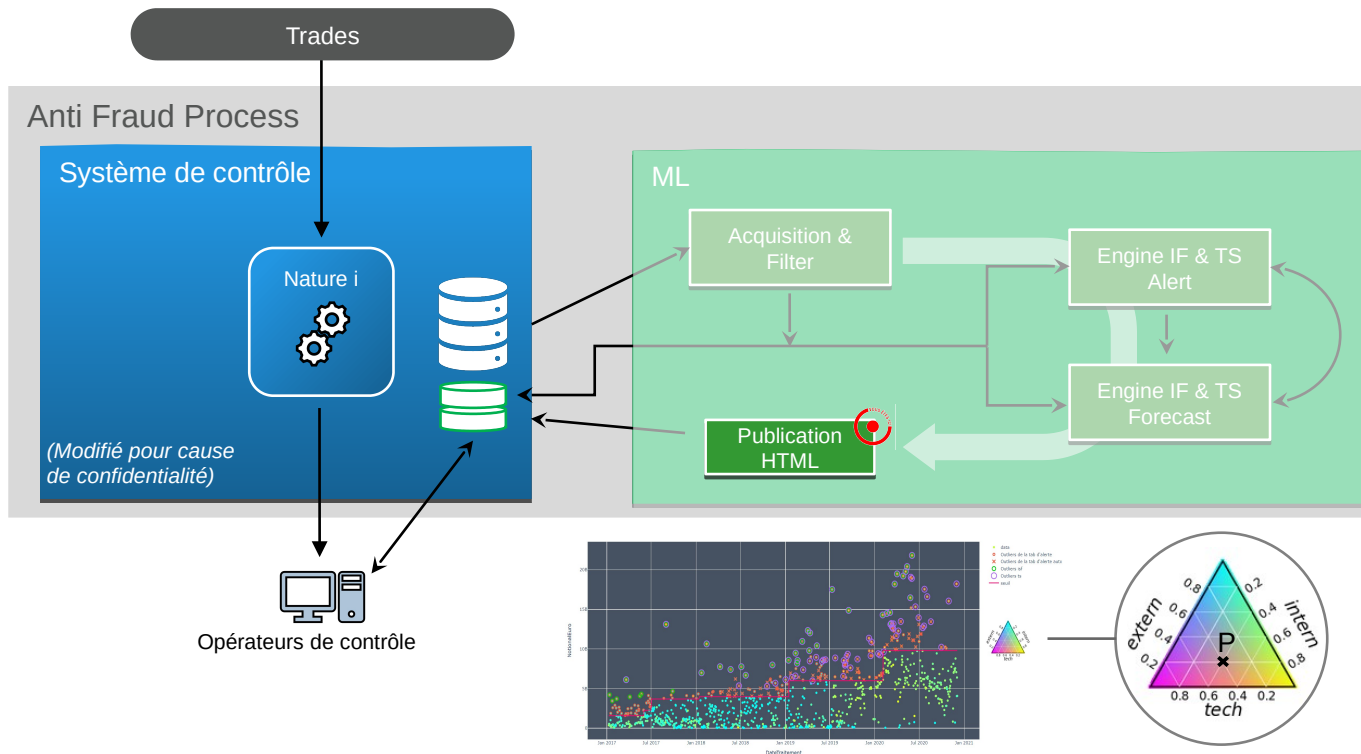
(IF+TS) **Évaluation** de la position de ce jour (position atypique ou non).

(TS) **Prévision** de la position de demain à partir du modèle.

Outils :



3 | Data visualisation



Rôles :

Visualisation des alertes classées en automatique ou traitées manuellement.

Seuil en fonction du temps.

Prévision TS de demain.

Data augmentation: fractions customers interne/externe/tech de la position journalière (couleur).

Outils :





CONCLUSION

Objectifs “*probablement*” atteints:

Amélioration sur la génération des anomalies (diminution du "bruit" des anomalies).

Proposition d'une nouvelle vision du seuil: seuil auto-ajustant.

Module de prédiction et d'autoévaluation de la position du jour du couple

L'approche explorée propose une vision complémentaire du système actuel.

Ajout de la dimension prévisionnelle sur les positions (⚠ ce n'est pas une science exacte).

La data visualisation développée apporte une vision alternative et augmentée au système actuel.

Remerciements & Conclusion personnelle



Valéry Adeleine - Head of IT+ IT BGC-GMO) -Business Support Control



Benoit Gatien - Project Leader Anti Fraud & Reconciliation



Antoine Barruol - Scrum Master



Valentin Fabianski - Datascientist – Business Support & Support

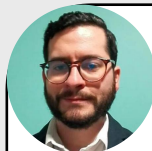


Thomas Pillot
(tpillot@student.42.fr)



Étudiant à 42 Paris.
Précédemment étudiant Licence économie
gestion.

Stagiaire developper/datascientist



Matthieu David
(mdavid@student.42.fr)



Étudiant à 42 Paris.
Secrétaire & Resp. pédagogique 42AI.
Précédemment doctorant en physique des
matériaux.

Stagiaire developper/datascientist