UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

MoSEF
MASTER ECO STAT - QMEF

NEXIALOG
CONSULTING

Kick-Off Challenge Professionnel en Data science

Armand L'Huillier
Tel : 06.73.03.49.61
Mail: alhuillier@nexialog.com

THINK SMART ACT DIFFERENT

Février 2025

Kick-off Challenge Nexialog 2025

01 **NEXIALOG**

02 **CHALLENGE**

Tel : 06.73.03.49.61
Mail: alhuillier@nexialog.com

THINK SMART ACT DIFFERENT

Retour sur le partenariat

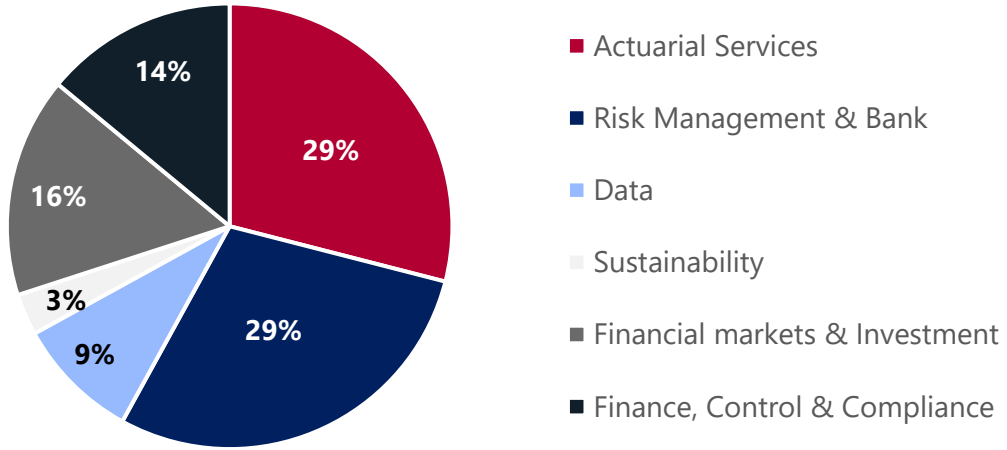
Nexialog X Mosef

Nexialog Consulting

Chiffres clés de 2024

Répartition de notre Chiffre d'Affaires en fonction de nos Business Units

En 2024



Ils nous font confiance

BNP PARIBAS · CRÉDIT AGRICOLE · CRÉDIT AGRICOLE CORPORATE & INVESTMENT BANK · CNP assurances · AXA · AG2R LA MONDIALE · SwissLife · KLÉSIA · PREDICA · Carrefour banque et assurance · Allianz · PRO BTP GROUPE · NATIXIS · SOCIETE GENERALE · malakoff humanis

2006	16 %	7	180	40	21,5 M€
Année de création	**Croissance Prévisionnelle** En 2025	**Nombre de Business Unit** En 2025	**Effectif** En 2024	**Nombre de Clients actifs** En 2024	**Chiffre d'Affaires** En 2024

Data Consulting
Business Unit

NOS SAVOIR FAIRE

Data Engineering

- Diagnostics et pilotage de projets de transformation des architectures des données.
- Expertise sur l'ensemble des métiers de la donnée : collecte, préparation, modélisation, analyse.
- Expertise métiers : financement, titrisation, marché, marketing, finance durable.
- Formation des équipes opérationnelles, accompagnement et acculturation.
- Veille technologique et réglementaire

Data Gouvernance

- Pilotage des projets de data management.
- Maitrise des architectures de données et IT.
- Maitrise des systèmes de données financières, extra-financières ; données structures et non structurées.
- Formation des équipes opérationnelles, accompagnement et acculturation.
- Veille technologique et réglementaire.

LES OUTILS

python, R, SQL, SAS, Power BI, SAP BO, data iku, alteryx, Collibra

Et bien d'autres…

NOS PROFILS :

- **Data Analyst** (Statisticien, Chargé d'études, Data Miner, etc.)
- **Data Visualist** (Data Analyst Viz, Data Analyst BI, DataViz Engineer, etc.)
- **Data Scientist** (ML Engineer, etc.)
- **Data Engineer** (DataOps, Cloud Engineer, Big Data Engineer, etc.)
- **Tech Lead Data** (Expert technique, AMOE, etc.)

Challenge : Nexialog X Master MOSEF

2021	2022	2023	2024	2025
2021 : Invitation à des évènements divers (afterworks…)	**Septembre :** Présentation de rentrée de Nexialog Consulting	**Septembre :** Présentation de rentrée de Nexialog Consulting	**Septembre :** Présentation de rentrée de Nexialog Consulting	**Septembre :** Présentation de rentrée de Nexialog Consulting
		Challenge Nexialog PD FORWARD LOOKING IFRS 9	Challenge Nexialog X DEXIA PD BALOISE ET SEGMENTATION	Challenge Nexialog ?
		Mars, avril, juin : Divers invitations	**Mars, avril, juin :** Divers invitations	
1 CDI recruté (en risque)	1 CDI recruté (en risque)	4 CDI recruté (en risque et data)	2 CDI recruté (en risque et data)	

	2023	2024	2025
Type	✓ **Risque de crédit**	✓ **Risque de crédit**	✓ **Data science & IA**
Sujet	✓ **Modéliser la Probabilité de défaut IFRS 9 Foward Looking**	✓ **Modéliser la Probabilité de défaut Baloise et segmenter en risque les clients**	✓ **?**
Livrables	• Modélisation • Créer un outil • Présenter au jury	• Modélisation • Créer un outil • Présenter au jury	• Modélisation • Créer un outil • Présenter au jury
Evaluation	**• Présentation (savoir être + application)**	**• Rigueur de modélisation**	**• Innovation de modèle**

Challenge : Nexialog X Master MOSEF

Ambition pour 2025

1)

- Depuis 2 ans, Nexialog fait des missions en Data, c'est le moment de faire évoluer le challenge pour permettre au MOSEFs de travailler sur des sujets IA en partenariat avec le cabinet

✓ Objectif : Faire un Projet professionnel, vendre le challenge pendant votre entretien. Votre projet résout des vrais problématiques d'entreprise

2)

- Montrer à **Nexialog** que le master MOSEF est un master puissant en data science : on a surtout des Mosef en risque au cabinet !

✓ Objectif : Recruter encore + de Mosef en Data au cabinet

3)

- Montrer au **partenaire** que le master MOSEF est un master puissant en data science

✓ Objectif : Faire travailler les MOSEF sur une nouvelle problématique IA

4)

- Vous travaillez sur un sujet Data science que vous préférez au risque de crédit xD
- Un sujet qui sera utile à la majorité pour les entretiens de CDI des étudiants

✓ Objectif : Vous faire travailler sur un sujet que vous pourriez aimer

Il y a un an… En Master 1 :

Quand vous connaissiez à peine python et le machine learning

De : Melanie Lucas <Melanie_Lucas@edinburghairport.com>
Envoyé : jeudi 13 juin 2024 16:37
À : Armand L'HUILLIER <alhuillier@nexialog.com>
Objet : Re: Partnership Sorbonne

Hi Armand,

Thanks again for the presentations yesterday, we were all impressed with the standard!

7 mois plus tard on se souvient encore de vous

Vous aviez fait du bon travail

De : Melanie Lucas <Melanie_Lucas@edinburghairport.com>
Envoyé : lundi, janvier 13, 2025 10:17 AM
À : Armand LHUILLIER <alhuillier@nexialog.com>
Objet : 2025 Student Projects

Hi Armand,

Hope you had a good holiday period!

Just touching base to see if you are interested in collaborating again on student projects this year?

Many thanks,
Melanie Lucas
Capacity Manager (Landside)
m: 07867 160 717

Edinburgh Airport
Where Scotland meets the world

Juin 2024

Janvier 2025

Faires encore mieux qu'en Eco Stat !

Maintenant que vous avez 1 an d'expérience en plus, plusieurs mois d'expérience professionnelle, appris davantage sur les techniques ML/IA, vous pouvez aller bien plus loin

NEXIALOG CONSULTING

THINK SMART ACT DIFFERENT

Dévoiler le Challenge 2025

Le partenaire 2025

Un leader du numérique français

GROUPE ALTICE

Acteur majeur des télécommunications

Top 2 téléphone mobile

Top 3 réseau fixe

Présent en Europe (Benelux)

Ce projet vous ouvre des portes dans les grands groupes des Télécoms
Quelques anciens mosef sont actuellement chez Free, Bouygues…

Le contexte
Détection d'anomalie réseau

Problématique

Régulièrement, les box internet subissent des interruptions réseaux : ERROR 404…

Quand plusieurs box reliées ensemble subissent ces problèmes de manière simultanée, alors les box ne sont pas défectueuses : c'est la faute au réseau lui-même. Il faut faire intervenir des techniciens pour faire la réparation sur le réseau au niveau qui est atteint par la panne (PEAG, OLT, PEU…)

Il est préférable d'anticiper où et quand des problèmes vont survenir.

Idée de résolution

Travail **non-supervisé.** Il n'y a pas de label/variable cible. On ne sait pas ni quand, ni où, il y aurait pu avoir une panne/une intervention d'un technicien.
Mais il y a un gros chantier de valoriation de la donnée : De nombreux tests sont mis en place dans toutes les box : permet de voir implicitement les interruptions réseaux.

Faire de la détection d'anomalie pour résoudre le problème

Chaîne technique : Les équipements sont interconnectés.

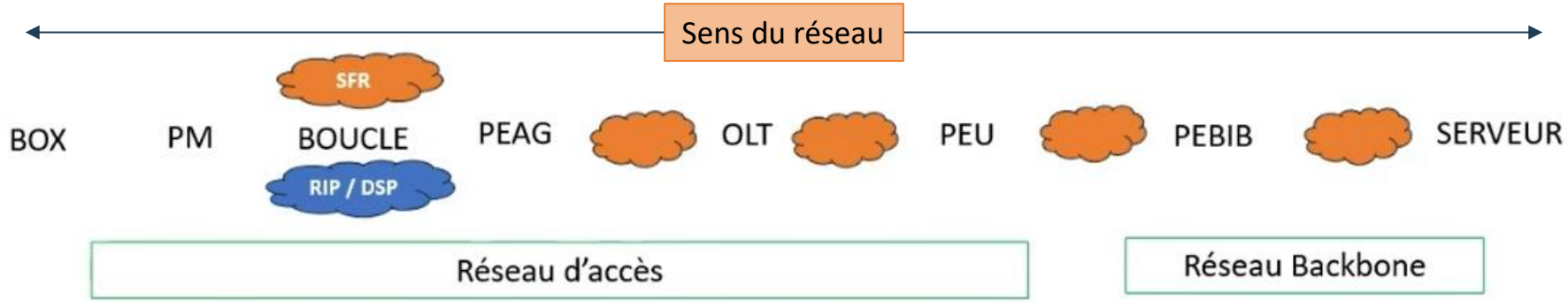
Chaîne technique
Détection d'anomalie réseau

Il y a plusieurs niveaux d'agrégation des équipements pour renseigner sur les flux envoyés/reçus:
Réseau d'accès (partagé entre SFR, Orange… Chacun gère des PEAG, des OLT, qu'ils mutualisent tous ensemble). Par exemple, certains PEAG sont gérés par SFR et d'autres non.

- **BOX** : ce qu'il y a chez chaque client. Elle envoie des informations/requêtes. Le but est de la relié au serveur le plus vite possible pour chaque échange d'information.
- **PM** (points de mutualisation) : Rassemblement de plusieurs BOX en un agrégat. Souvent à l'échelle d'une copropriété, d'un immeuble. Ca ressemble à un gros compteur, à une armoire.
- **BOUCLE** : N'est pas un agrégat comme les autres. C'est un nuage, un cercle. Ca sert de régulation du réseau. Le but est de transmettre les informations des PM aux PEAG les plus disponibles/moins surchargés pour améliorer la rapidité du réseau. Une boucle a plusieurs PEAG reliés à elle. Et plusieurs PM sont rattachés à une boucle.
- **PAEG** : Une box/un PM peuvent ont une boucle mais peuvent avoir plusieurs PAEG. Par contre à partir du moment où on se situe du le PAEG, on est à peu près sûr d'être rattaché au même OLT, PEU et PEBIB ensuite.
- **OLT** : Niveau d'agrégation intéressant.
- **PEU** : Niveau d'agrégation intéressant.

Réseau Backbone = uniquement du réseau SFR

- **PEBIB** : SFR a 8 PEBIB en France
- **SERVEUR** : Internet…



Définir le sujet
Détection d'anomalie réseau

VOTRE TRAVAIL SERA :

Vous aurez donc des données réseau à traiter. Il s'agit de nouvelles données : SFR a mis en place des « tests » depuis décembre sur les box. Donc vous avez les résultats des tests de décembre à janvier. Vos modèles seront exploratoires, ils n'ont encore rien modélisé : modèles non-supervisés.

Objectif : utiliser ces tests pour détecter des futurs problèmes sur une partie du réseau → Détection « d'anneaux ».

On se focus sur un seul nœud (PEAG ou OLT) par modèle.

Faire un modèle plutôt **complexe** par nœud (un modèle qui prend en input plusieurs critères/indicateurs/variables). A la maille **heure**. Dire X heures avant le besoin d'une intervention pour désamorcer l'anneau potentiel. Vous n'effectuerez pas les mêmes modèles et utilisation de métriques pour chaque niveau de réseau.

Autrement dit, vous construirez un modèle de détection d'anomalie au niveau PEAG, un autre différent au niveau OLT…

Tests

Test **DNS** = SFR demande régulièrement à la box d'envoyer des données vers le serveur. Typiquement, un test qu'ils mettent en place est de mesurer le temps que met la box à recevoir une adresse IP lorsqu'elle se connecte au DNS (Domain Name Server)/à un navigateur internet. Déjà SFR vérifie que la box reçoit en effet l'adresse IP, SFR mesure le temps qu'il a fallu pour la récupérer, et recommence le process pour déceler des variations de temps.

Test de **Score** = le test-score est un test qui agrège 5 tests et résume l'info en 1 (temps, débit, perte de paquet…).

Agrégations suggérées des tests : Boucle X PAEG X OLT, PAEG X OLT,… Tout en prenant en compte qu'un chemin pour un même PM change de PEAG pour différent test…

Dans les données brutes, les tests sont agrégés par heure et par chaîne technique (BxPxOxPxP).

Faut paaas agréger les résultats des deux tests. Faire temps_test_dns + temps_test_score : ca n'a pas de sens. (j'ai mis 3 « a » à pas donc ne le faites vraiment pas ok ?? Je vous verrai de toute façon)

Test DNS
KPI : Temps de réponse (ms)

2,4 M test/jour

Test Scoring
KPI : MoS Latence (ms) – retranscription (%)

1,2 M test/jour

Description des bases
Détection d'anomalie réseau

Informations à savoir sur les variables « ID »

Old_model = ancien vs nouveaux modèle. Peut-être que l'ancienneté causera un biais sur la détection d'anomalie. Il faudra peut-être faire un modèle de détection pour chaque.

Boucle : Htth sont les anciennes boucles, récupère les infos au niveau du client directement. BU=Boucle unifiée. Le numéro associé=l'ID de la boucle.

DSP = correspond à l'opérateur qui gère la boucle (si dsp_1 oui alors c'est SFR qui gère sinon c'est un autre opérateur).

PEBIB = est mal renseigné. Peut-être que vous serez amené à remplir les val. Manq. Prenez le mode entre PEAG et OLT si besoin. (le + souvent pas de variation de PEBIB avec ce croisement)

PEU = PE « unifié ». Rien à spécifier de particulier

Informations à savoir sur les variables numériques

Nb_test = c'est le nombre de test…DNS (DOMAIN NAME SERVEUR)

Avr_dsn_time=résultat de latence moyenne du test.

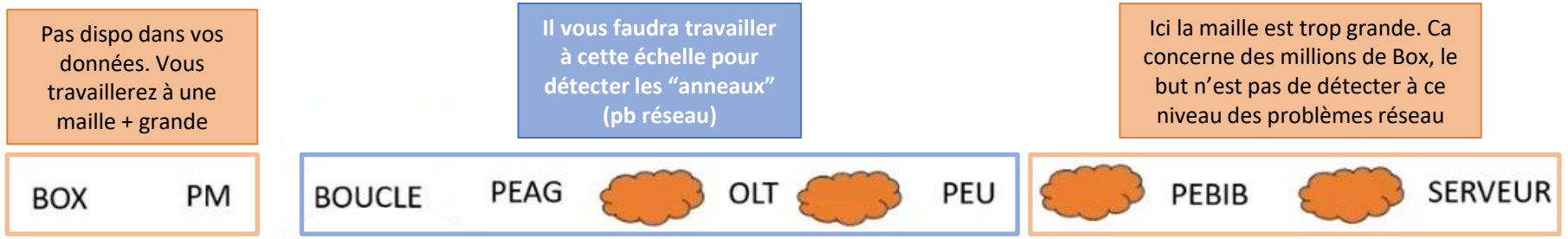
Vous avez aussi **l'écart-type**

…même chose avec les tests Scoring.

Métrique de lantence =en milisecondes.

Nb de clients= pas forcément intéressant d'analyser quand c'est manquant car aucun client… ou très peu. Se concentrer sur les tests qui concerne bcp de clients (moins de 10 c'est rien)

Olt_name (les 1ers chiffres sont son département) : Idée de métrique à calculer en plus : Distance entre le département de la box/pm/boucle et le département de qui traite l'appel DNS (n'est pas en soit une anomalie)



Présentation du sujet
Détection d'anomalie réseau

Complexité du sujet

La complexité du sujet est réelle :

❑ Proposer des modèles de détection d'anomalie **rigoureux** (en termes métier et data science).

Vous avez champ libre : des modèles qui traitent la série temporelle du PEAG et trouvent une anomalie sur son test de rapidité, des modèles ML, des modèles DeepL.

N'oubliez pas que vous devrez présenter vos modèles, à un jury qui souhaitera comprendre vos choix, qui voudra comprendre si le modèle fonctionne, s'il résout vraiment le problème. (vous devrez **expliciter** vos choix, vos pistes, vos découvertes)

❑ **Surtout un problème de taille** : les box/PM se sont pas toujours rattachés aux mêmes PEAG, OLT, PEU.. Donc faites bien attention quand vous agréger les lignes ensemble et quand vous comparer les résultats des tests entre eux. Un test passe une fois par une root, mais le test d'après de la même box/MP peut très bien passer par une root différente/par un chemin différent. Dans les données, c'est tracé.

Quelques conseils

Codez en **classe**/fonction : comme ca vous pourrez essayer plusieurs fois vos modèles à des niveaux d'agrégations différents rapidement.

Si la base est trop volumineuse :

- Vous pouvez essayer d'uploader votre dataframe sous format **Polar**. En Pandas ca sera clairement pas efficace. Sinon Pyspark est une bonne option aussi, comme vu en python avancé.

- Codez en ligne sur Colab (possible d'avoir accès à des GPU gratuitement peut-être). Ca sera de toute façon + efficace qu'en local une fois que votre base sera uploadée.

- Sinon vous pouvez supprimer des lignes : en agrégeant les données par Heure en demie-journée/journée. Mais soyez très vigilant à ne pas perdre des informations.

Les modalitées du challenge

Dates, jury, prix…

Asso mosef : svp faites signer les accord NDA pour que j'envoie ensuite les bases

11 avril

Vous présenterez votre travail, sous forme d'application (ou autre si + approprié). Le 11 avril, l'après-midi dès 13H. S'en suivra un cocktail au cabinet

Jury

- SFR (2 membres)
- Nexialog (2 membres, associés…)

Ils choisiront quels sont les groupes qui gagneront les prix. (2 lots : 1er groupe et 2ème groupe)
Rania et moi serons présents

Modalités d'évaluation

COMPRENDRE LE BESOIN/ PROPOSER UNE SOLUTION ADAPTEE

INNOVATION MODELE RIGOUREUX, METHODOLOGIE JUSTIFIEE
Ce qui est attendu de SFR : que vous proposiez des solutions nouvelles. Des méthodologies, des trucs smarts

PRESENTATION/ APPLICATION/ DESIGN

Constituez des groupes de 4, et commencez à travailler le + vite possible (pour gagner)

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

MoSEF
MASTER ECO STAT - QMEF

NEXIALOG
CONSULTING

Kick-Off Challenge Professionnel en Data science

Bon courage et bon challenge !

Armand L'Huillier
Tel : 06.73.03.49.61
Mail: alhuillier@nexialog.com

THINK SMART ACT DIFFERENT

Février 2025