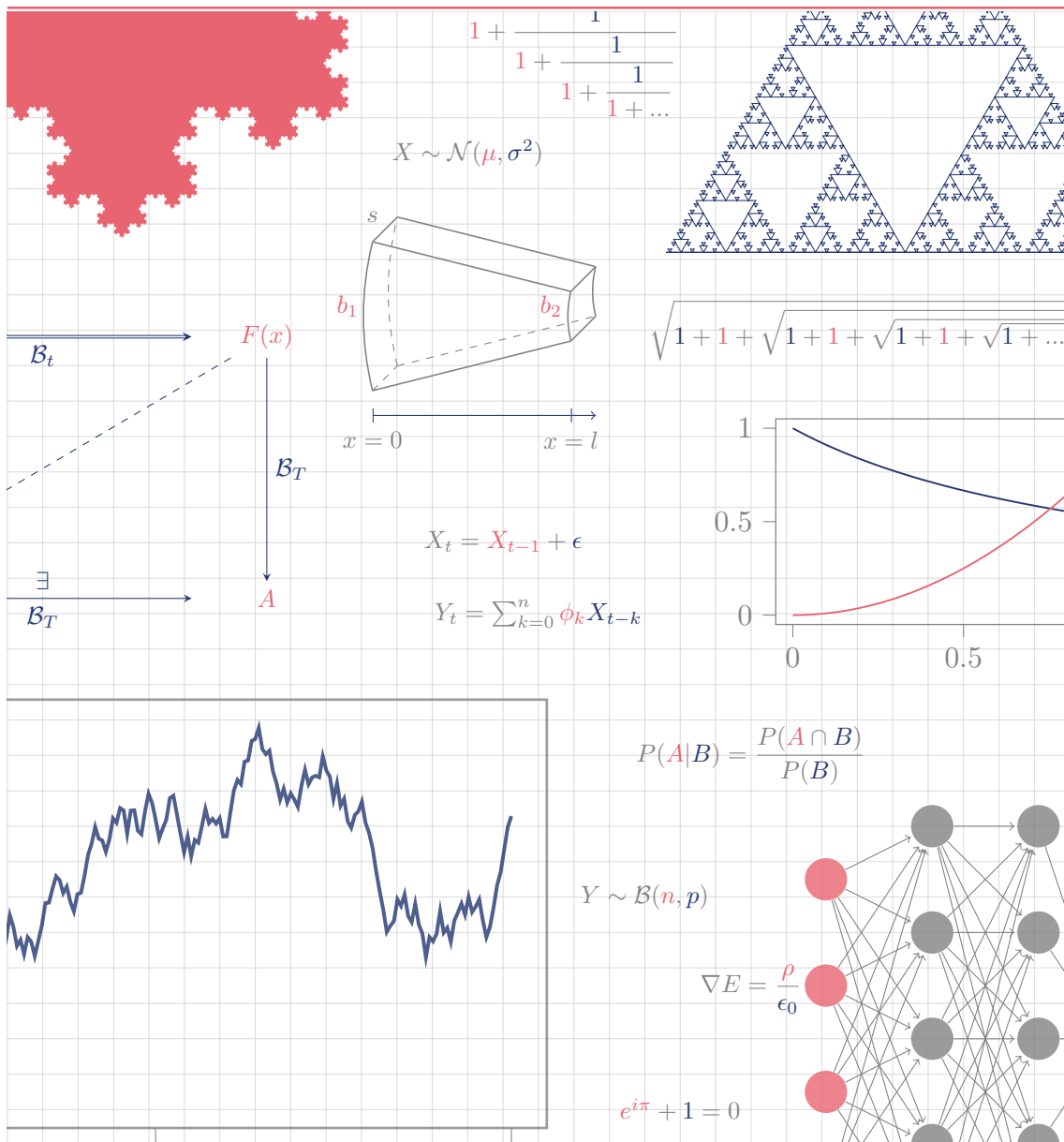


## RAPPORT DE STAGE DE TROISIÈME ANNÉE

Robin CAMARASA

August 12, 2023



# Contents

---

<b>Introduction</b>	<b>3</b>
<b>1.1 Remerciements</b>	<b>3</b>
<b>1.2 Présentation générale</b>	<b>3</b>
1.2.1 Mon parcours . . . . .	3
1.2.2 Mes missions . . . . .	3
<b>Présentation de l'entreprise</b>	<b>4</b>
<b>2.1 DataGenius dans son contexte</b>	<b>4</b>
2.1.1 Description générale de DataGenius . . . . .	4
2.1.2 L'offre de DataGenius . . . . .	4
2.1.3 DataGenius une entreprise accompagnée. . . . .	4
<b>2.2 Organisation</b>	<b>5</b>
2.2.1 Organisation globale . . . . .	5
2.2.2 Organisation technique . . . . .	6
<b>2.3 Perspectives et enjeux</b>	<b>6</b>
2.3.1 La vision de DataGenius . . . . .	6
2.3.2 Les enjeux à venir . . . . .	6
<b>Présentation des missions réalisées</b>	<b>7</b>
<b>3.1 Technologies</b>	<b>7</b>
<b>3.2 Produits internes</b>	<b>7</b>
3.2.1 Face Count . . . . .	7
3.2.2 AI compare. . . . .	8
3.2.3 Data Voice . . . . .	8
<b>3.3 Mission clients</b>	<b>8</b>
3.3.1 Fongecif . . . . .	8
3.3.2 Michel Condemine . . . . .	9
3.3.3 BioMéca . . . . .	9
3.3.4 Lunettes pour tous. . . . .	9

<b>3.4 Rédaction d'articles de Blog</b>	<b>10</b>
3.4.1 Article sur l'apprentissage profond bayésien . . . . .	10
3.4.2 Article sur la participation au challenge ISIC2019. . . . .	10
<b>Conclusion</b>	<b>11</b>
<b>4.1 Bilan</b>	<b>11</b>
<b>4.2 A venir</b>	<b>11</b>

# Introduction

---

## 1.1 Remerciements

Je tiens à débiter ce rapport en remerciant les personnes sans lesquelles ce stage aurait été impossible. Tout d'abord un grand merci à madame Mathieu Jugnaru pour sa supervision attentive et avisée. Je tiens également à montrer toute ma gratitude à monsieur Melaine pour la maîtrise technique qu'il a réussi à me transmettre. Une pensée toute particulière me vient pour madame Cappai qui a su m'enseigner la rigueur et le sérieux nécessaire à l'activité de conseil. Je remercie monsieur Zemmouri de m'avoir accueilli dans son entreprise et m'avoir initié aux problématiques de conseil dans le monde de la Data Sciences. Un grand merci, également à l'équipe RER pour son accueil chaleureux.

## 1.2 Présentation générale

Ce rapport de stage entre dans le cadre du travail de fin d'étude du cursus d'Ingénieur Civil des Mines (ICM) de Saint-Etienne de Robin CAMARASA. Il a pour objet de présenter dans une première partie l'entreprise DataGenius.

### 1.2.1 Mon parcours

Afin de donner certains éléments de contexte sur les différentes missions qui m'ont été confiées durant ce Travail de Fin d'Etude (TFE), je tiens à rappeler brièvement mon parcours à l'école des Mines de Saint-Etienne. Ainsi au niveau des choix de cours pour ma deuxième année, j'ai opté pour la majeure Informatique accompagnée de la toolboxe Intelligence Artificielle. Afin d'approfondir les notions vues dans la toolboxe Intelligence Artificielle, j'ai opté en troisième année pour la majeure Sciences des Données. Enfin, pour avoir une vision métier des 2 majeures que j'ai choisies à l'Ecole, il a été logique de les lier en choisissant le défi Big Data.

### 1.2.2 Mes missions

Donnons à présent un rapide aperçu des différentes tâches que j'ai pu réaliser pour DataGenius durant ce stage de fin d'études. Mon travail se subdivisait en 2 classes de missions. D'une part, les missions clients où j'intervenais en qualité de data scientist auprès d'entreprises. D'autre part, les projets internes où je réalisais des travaux de développement informatique pour créer ou poursuivre l'avancée de produits internes. Ces différentes missions seront détaillées dans le chapitre "Présentation des missions réalisées".

# Présentation de l'entreprise

---

## 2.1 DataGenius dans son contexte

### 2.1.1 Description générale de DataGenius

DataGenius[2] est une start-up implantée à Villeurbanne depuis avril 2017. Elle se définit comme un DataLab externalisé, offrant à ses clients une cellule d'innovation et de recherche externe en Data Science. L'objectif ambitieux de la jeune entreprise est de devenir un pure player en matière d'intelligence artificielle dans la région lyonnaise. En parallèle de cette activité de consulting, des produits internes sont développés dans le but d'accroître l'offre et la rentabilité de l'entreprise sur le long terme.

### 2.1.2 L'offre de DataGenius

DataGenius propose 3 offres B2B à ses clients. Ces 3 offres sont adaptées au niveau de maturité de l'entreprise quant à ses problématiques de valorisation des données des entreprises. Voici la présentation de ces dernières :

1. Diagnostic IA et DATA : Cette offre est réservée aux clients s'intéressant à la valorisation de données mais ne sachant pas comment la mettre en place dans leur entreprise. Cette première approche permet de démystifier ces nouvelles technologies auprès des clients et ainsi de proposer des projets réalistes pour leur entreprise.
2. Accompagnement : Cette offre permet généralement de réaliser un projet défini durant la phase de diagnostic et aboutit à la réalisation à minima d'un POC, au mieux à la mise en production de la solution développée.
3. Expertise : Cette offre permet de mettre à la disposition du client un data-scientist sous la modalité d'un forfait jour.

### 2.1.3 DataGenius une entreprise accompagnée

Comme de nombreuses start-up, DataGenius est accompagnée par des organismes qui visent à l'aider à mettre en place sa stratégie entrepreneuriale. L'un de ces organismes est le Réseau Entreprendre Rhône qui héberge ses bénéficiaires dans son espace de co-working. En étant lauréat du réseau entreprendre, DataGenius bénéficie d'un accompagnement personnalisé réalisé par un professionnel de l'entrepreneuriat. De plus, un membre du réseau, soit une personne ayant occupé

ou occupant une position de direction dans une entreprise, effectue du mentorat auprès de DataGenius. Enfin, des ateliers thématiques permettent aux lauréats d'échanger sur les difficultés qu'ils rencontrent au quotidien.

## 2.2 Organisation

### 2.2.1 Organisation globale

L'effectif de l'entreprise à mon arrivée était de 7 personnes comme le montre l'organigramme de la figure 2.1 à la page 5. Parmi ces 7 membres du personnel, 3 membres sont employés en contrat à durée indéterminée et forment le comité de direction : Taha Zemmouri, Sammy Melaine et Marina Cappai sont respectivement Chief Executive Officer (CEO), Chief Technical Officer (CTO) et Chief Operating Officer (COO). Ainsi, monsieur Zemmouri est responsable du travail rendu au client et de la vision à court et long terme de l'entreprise. De son côté, monsieur Melaine, est garant du recrutement technique et de la qualité technique des livrables apportés aux clients. Enfin, madame Cappai, supervise la qualité et l'amélioration continue. A ces fonctions de direction s'ajoutent les fonctions de data scientist (pour monsieur Melaine) et data consultant (pour monsieur Zemmouri et madame Cappai).

Je me permets ici un aparté pour clarifier la différence entre data scientist et data consultant. En effet bien que les terminologies soient proches il existe une réelle différence entre ces deux métiers. Le data scientist a une approche technique de la mission client là où le data-consultant possède une approche métier. De plus, le data-consultant se place en tant que chef de projet.

L'effectif était complété de 4 stagiaires. Monsieur Negri, en stage alterné entre novembre et juillet, était en charge de la commercialisation de DataVoice, un des produits internes de l'entreprise. Madame Paradis, également en stage alterné sur la même période, assurait les fonctions de secrétaire de direction. Madame Turiaud, quant à elle, était engagée sur une durée de 2 mois à partir d'avril, avait pour mission de stage de moderniser la charte graphique de l'entreprise. Enfin, je complétais l'effectif en tant que data scientist comme j'ai pu l'expliquer précédemment.

Malgré la petite taille de la structure, la volonté de hiérarchie pyramidale est clairement affichée par le comité de direction de l'entreprise. Cependant dans la vision de l'entreprise, cette verticalité de l'organisation n'est pas là pour brider la créativité, elle encourage au contraire à faire remonter les problématiques rencontrées le plus rapidement possible.

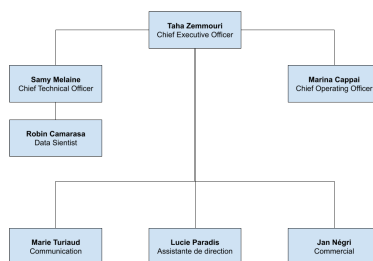


Figure 2.1: Organigramme de DataGenius, avril 2019

### 2.2.2 Organisation technique

Afin de réaliser un bilan et d'aborder des pistes d'améliorations aux process actuels, l'équipe technique se réunissait une fois par semaine lors d'un point technique. Ce point technique durait 1h30, le lundi de 9h à 10h30 et s'articulait en 3 parties :

1. Reprise des problématiques rencontrées sur la semaine: Durant approximativement 30 minutes, cela permet l'échange sur les difficultés rencontrées et sur les réussites techniques de la semaine.
2. Planification des challenges à venir: Point sur les différentes problématiques soulevées par les produits internes et les projets clients du moment.
3. Pistes d'amélioration continue: Ce dernier moment du point technique permet de réfléchir à long terme aux choix techniques de l'entreprise. Sur la période de mon stage, cela a été l'occasion d'intégrer des démarches qualité aux code notamment par la génération de documentation automatique (librairie sphinx), l'amélioration de la robustesse des applications en ligne de commande (librairie click) et l'utilisation d'intégration continue (Circle CI)

## 2.3 Perspectives et enjeux

### 2.3.1 La vision de DataGenius

Actuellement, DataGenius a spécialisé son offre sur la partie technique et non sur la partie métier. Effectivement, sur le plan technique la start-up a acquis une expertise sur le domaine du machine learning qui est une sous-branche de l'intelligence artificielle. En ce qui concerne la branche métier, on s'aperçoit que l'entreprise possède des clients venant de domaines très différents allant des domaines de l'immobilier au domaine de la reconversion professionnelle. C'est donc sur ce point que DataGenius souhaite travailler en adaptant son offre au secteur d'activité de la santé.

De plus, l'entreprise souhaite augmenter son panier moyen en se positionnant sur des projets plus ambitieux. Un des éléments clés mis en place par DataGenius pour arriver à cette augmentation du panier moyen est de modifier la cible de sa prospection. En effet, durant mon stage j'ai pu observer le passage d'une prospection de tout type d'entreprise à une prospection ciblée sur les grands groupes.

### 2.3.2 Les enjeux à venir

Cette spécialisation de DataGenius tant sur le métier que sur la typologie de projets clients apporte son lot d'interrogations et de problématiques. En effet, sur le plan métier il est nécessaire de former les équipes aux métiers de la santé. De même, la volonté de se positionner sur des projets de plus grande envergure auprès de plus grandes entreprises oblige la division technique à mettre en place des stratégies de qualité de codes et de mettre en place une politique de maintenance.

De plus, DataGenius se trouve à un moment charnière sur le plan économique. En effet, comme de nombreuses start-up la société a effectué des investissements tant sur le plan économique que sur le plan humain (notamment par la formation de ses employés). C'est ainsi dans les prochains mois que l'on pourra déterminer si ces investissements se sont avérés rentables.

Enfin, DataGenius rencontre un problème de gestion RH qui représente actuellement l'enjeu principal. En effet, durant mon stage DataGenius a connu de nombreux départs qu'il va falloir remplacer. Ces recrutements s'opèrent à des postes clés tant dans le domaine technique que dans le domaine commercial. Ils seront déterminants pour l'avenir de DataGenius.

# Présentation des missions réalisées

---

## 3.1 Technologies

Avec l'arrivée de monsieur Melaine, au poste de CTO, un certain nombre de choix techniques ont été réalisés :

- Système opérationnel: Au niveau des systèmes d'exploitation, il a été choisi de travailler avec la distribution Ubuntu 18.04 basé sur un noyau UNIX.
- Technologie data science: Les projets de data science sont réalisés en python accompagné des bibliothèques pandas[3], scikit-learn[4], tensorflow[1] et plotly qui permettent respectivement de traiter les bases de données, appliquer des algorithmes de machine learning classiques, créer des réseaux de neurones et réaliser des missions de data vizualisation légère.
- Technologie data vizualisation: Pour les missions nécessitant uniquement de data vizualisation, l'équipe technique opte pour les technologies qlick sense ou PowerBI en fonction des besoins clients.
- Technologie web: Dans ce domaine, il est nécessaire de différencier 2 types de technologies : le back end qui correspond aux codes exécutés sur serveur et le front end qui correspond aux codes exécutés par le client. Pour les technologies de serveurs, DataGenius a opté pour le framework django qui permet un développement et une intégration rapide. En ce qui concerne le front end, HTML, css et Javascript sont utilisés sans framework.
- Technologie big data: L'entreprise souhaitant développer ce pan de son activité, une formation est dispensée aux salariés souhaitant développer leurs compétences en scala afin d'utiliser les technologies Spark[6].

## 3.2 Produits internes

### 3.2.1 Face Count

Face Count fut le premier projet sur lequel j'ai pu travailler la première semaine de mon stage. l'objectif était dans un premier temps la reprise d'un code permettant la détection de visages humains et d'émotions en temps direct. Dans un second temps, la mission consistait à configurer l'ordinateur pour que le code fonctionne sur carte graphique et non sur processeur. Enfin, la création d'un algorithme visant à compter le nombre de visages distincts détectés sur la journée complétait le projet.



Ce projet était principalement un outil de communication de DataGenius au salon du SIDO, salon lyonnais réunissant les principaux acteurs de l'IoT et de l'Intelligence Artificielle. Ainsi, la transformation d'un algorithme fonctionnant sur processeur en un algorithme fonctionnant sur carte graphique a permis d'obtenir un programme robuste assurant le fonctionnement en temps réel sur une journée entière. La seconde partie de comptage quand à elle, avait pour objectif la réalisation d'un post LinkedIn.

### 3.2.2 AI compare

De la même façon que le projet précédemment décrit, la version bêta de ce projet interne devait être réalisée avant un salon, le salon 5I de Grenoble. L'objectif de cet outil est la comparaison des principales solutions de Machine Learning disponibles sur internet. Pour ce faire, dans le cadre d'une version bêta, nous nous sommes concentrés sur les solutions d'Amazon, IBM, Google et Microsoft sur les tâches suivantes : la reconnaissance d'images, la retranscription d'audio sous forme textuelle, la diarisation d'audio, l'analyse de sentiments et l'extraction de mots clés dans un texte.

Initialement prévu comme une reprise d'une solution créée par un ancien collaborateur de DataGenius, en accord avec mon directeur technique nous avons décidé de réorienter le projet en repartant de zéros. Les grandes difficultés de ce projet furent d'interroger les différentes API et SDK de Google, Amazon, Microsoft et IBM et de trouver un formatage des réponses obtenues permettant une comparaison équitable des résultats. Ce projet prit 3 semaines de travail équivalent temps plein. La solution web fut mise en ligne avant le début du salon 5I est encore disponible ici : <http://ai-compare.com>

### 3.2.3 Data Voice

Le produit ici développé est un site web à destination des centres d'appels qui a pour but de réaliser la retranscription et l'analyse de ces appels. J'ai pu travailler sur ce projet sur une durée de deux semaines tant sur la partie front-end que sur la partie back-end de la solution. Initialement prévu comme la reprise des codes d'un ancien collaborateur de Data Genius, après une analyse plus poussée du travail effectué par mon directeur technique et moi-même, nous avons décidé de garder une grande partie du front-end et de reprendre de zéro le back-end. Beaucoup de codes ont pu être repris d'autres projets internes de l'entreprise car ce produit faisait appel aux mêmes API et SDK que AI Compare. Au niveau de la méthodologie, cela a permis à l'équipe technique de mettre en pratique des solutions agiles telles que Trello.

## 3.3 Mission clients

### 3.3.1 Fongecif

Fongecif est un organisme visant à accompagner les salariés dans des processus de formation ou de réorientation professionnelle. Cet organisme nous a sollicité sur une offre de Diagnostic IA auquel j'ai pu apporter mon expertise technique. De ce diagnostic nous avons pu retirer différents besoins de l'entreprise en terme de valorisation des données et proposer un POC sur la visualisation de parcours des demandeurs de formation au sein du Fongecif.

Le livrable réalisé d'un point de vue technique fut un jupyter notebook permettant d'effectuer des statistiques élémentaires sur les demandeurs de formations ainsi qu'un diagramme de visualisation des différentes étapes de leur parcours. Ce notebook permettait également de filtrer les demandeurs selon des caractéristiques simples (sexe, tranche d'âge, catégorie socio-professionnelle etc.).

### 3.3.2 Michel Condemine

Michel Condemine fut un client exotique pour DataGenius en cela que, contrairement aux autres clients qui sont des entreprises, monsieur Condemine a fait appel à nous en tant que particulier. Il nous a sollicité sur une offre de type Expertise pour le développement d'un algorithme de Spiking Neural Networks.

Cette mission a occupé la majorité de mon stage est s'est effectuée en 2 phases. La première consistait à reprendre des scripts octave réalisés par un chercheur en neurosciences italien et de les "traduire" en python. Une fois cette transposition de langage réalisée, il a été nécessaire de modifier le modèle du chercheur en passant d'un neurone à un réseau neuronal. La deuxième phase était soumise à la condition d'un GO/No GO de la part de monsieur Condemine. Etant satisfait des résultats, monsieur Condemine signa la phase 2 qui comprenait la modification du modèle et l'application à deux problèmes, le problème du XOR et la reconnaissance de chiffres de la Base MNIST.

Madame Cappai et moi-même avons travaillé en binôme sur ce projet. Cette mission fut très enrichissante car elle comprenait chez le client une forte part d'affecte qui a pu rendre le dialogue complexe et tendu par moment. Ainsi cela fut formateur sur le plan technique mais surtout sur le plan humain et commercial.

### 3.3.3 BioMéca

Bioméca est une start-up de biomédical également lauréate du réseau entreprendre. Cette petite structure nous a sollicités sur une mission de type accompagnement. La réalisation souhaitée fut un réseau de neurones permettant de déterminer le type de lésions cutanées d'un patient. L'objectif à long terme est d'une part la prédiction du type de lésion grâce à l'algorithme que nous avons mis en place et d'autre part la détection du stade d'avancement de cette dernière par une analyse mécanique des tissus.

Ce projet fut très paradoxal car d'un côté nous avons largement rempli les objectifs attendus par le client et il fut très enrichissant sur le plan technique ; mais d'un autre côté de nombreux problèmes d'organisation et un dépassement conséquent du temps furent à déplorer sur ce projet. Ces problématiques furent l'occasion d'avoir une réflexion sur la place de la gestion de projet et s'interroger sur la communication et la collaboration entre data scientists et data consultants.

Sur le plan purement technique, nous avons mis en place avec Samy Melaine des méthodes de transfer learning[5] afin de réduire le temps d'apprentissage des différents réseaux convolutifs que nous avons utilisés. Les différents modèles testés sur cette tâche ont été les suivants : ResNet50, InceptionNet, MobileNet, DenseNet. Le réseau démontrant les meilleurs résultats a été le ResNet50. Les deux grandes difficultés de ce projet sur le plan technique ont été le fait que le jeu de données d'entraînement était déséquilibré et le temps d'entraînement des réseaux qui était très long par rapport à la durée de la mission.

### 3.3.4 Lunettes pour tous

Je travaille actuellement sur une mission pour l'entreprise Lunettes Pour Tous. Cette mission correspond à l'offre Accompagnement mise en place par DataGenius. Ainsi, dans une première phase nous avons réalisé une analyse exploratoire des données qui a permis de définir avec le client un projet possible de valorisation de leurs données. Ce projet consiste à réaliser des prédictions de ventes par demi-journée sur une semaine.

L'approche que j'ai développée sur ce projet fut tout d'abord d'utiliser des séries temporelles pour effectuer mes prédictions. DataGenius travaillant sur python, j'ai réalisé une revue des différentes bibliothèques proposant de l'analyse de times series. Celle qui répondait le mieux aux besoins fut prophet développée par Facebook. Une fois la prédiction réalisée via la times-serie, une forêt aléatoire vient corriger cette prédiction en fonction de variables exogènes telles que la météo ou le nombre de visites sur le site internet.

## 3.4 Rédaction d'articles de Blog

L'objectif premier de la rédaction d'articles de Blog pour DataGenius est de créer un contact différent avec ses anciens clients et prospects. Il s'agit donc d'un outil de démarchage permettant également de démontrer les capacités techniques de l'entreprise en vulgarisant les travaux de R&D effectués par ses data scientists.

### 3.4.1 Article sur l'apprentissage profond bayésien

Pour m'habituer à l'exercice, monsieur Zemouri et monsieur Melaine ont décidé de me donner un sujet que je maîtrisais bien, le sujet de mon projet recherche. Cet article portait donc sur l'apprentissage profond bayésien. Ce fut un premier exercice très intéressant et complémentaire de mon projet recherche en cela que le public cible et intentions étaient clairement différents. Ainsi, mon projet recherche se destinait à des scientifiques là où ce post de blog avait un objectif commercial. Il était donc nécessaire de mettre en avant les intérêts que pouvait trouver les potentiels clients de ces solutions tout en réduisant au minimum les détails trop techniques. Vous pourrez ainsi trouver l'article au lien suivant : <https://www.datagenius.fr/post/bayesian-deep-learning-soyez-sur-de-vos-incertitudes>

### 3.4.2 Article sur la participation au challenge ISIC2019

Ce second article fait écho au projet de classification des lésions cutanées proposé par BioMéca. En effet, suite à notre mission en collaboration avec la start-up lauréate du réseau entreprendre, le CTO de DataGenius a décidé de poursuivre l'aventure en participant au challenge international proposé par ISIC, une organisation internationale dédiée au diagnostic médical. J'ai donc rédigé un article expliquant les enjeux du challenge et notre méthode, afin de permettre à DataGenius de communiquer sur sa première participation à un challenge de ce type.

# Conclusion

---

## 4.1 Bilan

Cette conclusion sera pour moi l'occasion de dresser un bilan sur mon étonnement personnel vis à vis de la facette du monde de l'entreprise à laquelle j'ai pu être confrontée durant ce stage. A titre personnel, j'ai pu apprécier la diversité des missions proposées durant ce stage, ce qui m'a permis de me former sur de nombreuses techniques et technologies. Cependant, ce n'est pas sur le plan technique que cette expérience fut la plus riche en enseignement. En effet, grâce à cette expérience je me suis aperçu que la pierre angulaire d'une petite structure est la gestion des ressources humaines. Gestion qui est rendue très difficile par les pressions financières que peuvent connaître les start-up. Cela constitue sans l'ombre d'un doute la leçon la plus importante de mon stage.

## 4.2 A venir

Dans un avenir très proche, DataGenius va se voir confrontée à de nombreux challenges. Ainsi, la start-up connaît actuellement de grands bouleversements de son effectif qui vont apporter de grandes réflexions sur le plan RH. Effectivement, avec les départs de nombreux membres de l'effectif, la jeune start-up se voit dans l'obligation de mettre en place une politique de recrutement avisée dont l'orientation future de l'entreprise dépend.

En ce qui me concerne, les différents projets que j'ai pu réaliser ont réveillé en moi mon goût pour la recherche scientifique. Passion que je vais assouvir en réalisant un doctorat en Deep Learning appliqué à l'imagerie médicale à l'université Erasmus Medical Center de Rotterdam.

## Bibliography

---

- [1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.
- [2] DataGenius. Datagenius website. <https://www.datagenius.fr>.
- [3] Wes McKinney et al. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX, 2010.
- [4] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- [5] Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pages 242–264. IGI Global, 2010.
- [6] Matei Zaharia, Mosharaf Chowdhury, Michael J Franklin, Scott Shenker, and Ion Stoica. Spark: Cluster computing with working sets. *HotCloud*, 10(10-10):95, 2010.