



ENSAE 1^{ère} année

Stage d'ouverture au monde professionnel

Année scolaire 2023-2024

Tilian Bourachot

Efficacité des LLMs : Analyse et Optimisation via le Clustering de Séries Temporelles

CREST

Palaiseau

Maître de stage : Azadeh Khaleghi

10 juin 2024 - 6 septembre 2024

ENSAE Paris

Service des relations entreprises et des stages

5, avenue Henry Le Chatelier - 91120 PALAISEAU - FRANCE

Tél : +33 (0)1 70 26 67 39 - Courriel : stage@ensae.fr - www.ensae.fr

Table des matières

1	Remerciements	4
2	Introduction	5
3	Le CREST comme organisme d'accueil	7
3.1	Vue d'ensemble du CREST	7
3.2	Se positionner au sein de la diversité des acteurs	8
4	Ma mission au sein du CREST	10
4.1	Évolution du travail pendant le stage	10
4.2	Description de notre mode de travail	10
4.3	Introduction au sujet du stage	11
4.4	Approfondissement du sujet et premières expériences	11
4.5	Conclusion et perspectives	13
5	Le quotidien du stagiaire	14
5.1	Mon cadre de travail	14
5.2	Une journée typique au CREST	16
6	Mes observations et impressions	17
6.1	Relations avec mon encadrante	17
6.2	Les carrières dans la recherche	17
6.2.1	Le rôle de maître de conférences	17
6.2.2	La carrière de doctorant	18
6.2.3	Autres Observations	18
7	Bilan	21

1 Remerciements

Je tiens à remercier tout particulièrement mon encadrante, Mme Azadeh Khaleghi, sans qui je n'aurais pas pu mener à bien ce stage et découvrir plus en profondeur le domaine de la recherche. Je la remercie également pour ses explications claires sur les sujets que nous avons abordés, ainsi que pour ses précieux conseils tout au long de cette expérience.

Je remercie également Théo, autre stagiaire de Mme Khaleghi, étudiant en deuxième année à l'ENSAE, qui effectuait son stage en même temps que moi. Il a été en quelque sorte mon premier "collègue" de travail. Nos sujets de stage étaient liés, et il a toujours été disponible pour répondre à mes questions et m'expliquer certains points, ce qui m'a grandement aidé.

Et enfin je remercie mes camarades en stage en même temps que moi qui m'ont fait passé un bon été sur le plateau...

2 Introduction

Au cours de ma première année à l'ENSAE, l'obligation de réaliser un stage de fin d'année suscitait en moi à la fois excitation et appréhension. En effet, je n'avais jamais vraiment eu d'expérience professionnelle. Cependant, je souhaitais que ce stage m'aide à orienter ma future carrière professionnelle et me fasse découvrir de nouveaux domaines et de nouvelles notions. En effet, les possibilités de stages et les débouchés offerts aux étudiants de l'ENSAE sont vastes, et il est crucial de ne pas perdre de temps pour définir la direction à prendre et les cours à privilégier.

Au début de ma réflexion, je n'avais pas d'idée précise quant à mon orientation professionnelle. Mon intérêt pour les mathématiques et l'informatique a toujours été présent, et c'est en assistant à plusieurs conférences organisées par les associations de l'école que j'ai découvert le domaine de la finance. Cette discipline, qui allie mes deux centres d'intérêt, a particulièrement retenu mon attention. J'ai alors entrepris d'explorer les opportunités dans ce secteur. Après de nombreuses recherches et discussions, j'ai eu la chance d'obtenir un stage au sein de la Société Générale en tant qu'analyste Equity. J'avais réussi les entretiens avec succès, et tout semblait se dérouler favorablement.

À deux semaines du début du stage, et malgré plusieurs relances par e-mail, je n'avais plus reçu aucune nouvelle de l'entreprise. Craignant de me retrouver sans stage à la date prévue, j'ai pris la décision d'explorer d'autres options. Après avoir entendu plusieurs témoignages d'élèves ayant réalisé un stage de recherche, l'idée de m'orienter vers ce domaine, notamment en mathématiques, a commencé à germer. J'ai alors contacté plusieurs chercheurs du CREST, et c'est à ce moment-là que Mme Azadeh Khaleghi m'a répondu, me permettant ainsi d'obtenir un stage en recherche.

Mon stage était initialement prévu du 10 juin au 7 août, conformément à la durée maximale autorisée pour un stage non rémunéré. Cependant, après quelques semaines de travail, Mme Khaleghi m'a proposé de prolonger le stage afin de mener une exploration plus approfondie de nos recherches. J'ai accepté cette proposition, et mon stage s'est finalement achevé le 6 septembre. Cette prolongation m'a permis de consacrer davantage de temps à des sujets essentiels et de réaliser des expériences plus poussées, renforçant ainsi la qualité de notre travail.

Ce rapport présente tout d'abord l'organisation du CREST, puis détaille la mission que j'ai accomplie au cours de ce stage. Il examine ensuite les conditions et l'environnement de travail dans lesquels j'ai évolué, avant de conclure par la présentation de mes observations.

3 Le CREST comme organisme d'accueil

3.1 Vue d'ensemble du CREST

Créé en 2015 sous le nom d'EXCESS, pour ENSAE, X et Centre d'Économie, Statistique et Sociologie, le CREST résulte de la fusion entre le GRECSTA, un groupe de recherche associant l'INSEE, l'ENSAE, l'ENSAI et le CNRS, et le Laboratoire d'Économétrie de l'École Polytechnique. En 2016, cette unité mixte de recherche adopte son nom actuel, CREST, qui signifie Centre de Recherche en Économie et Statistique.

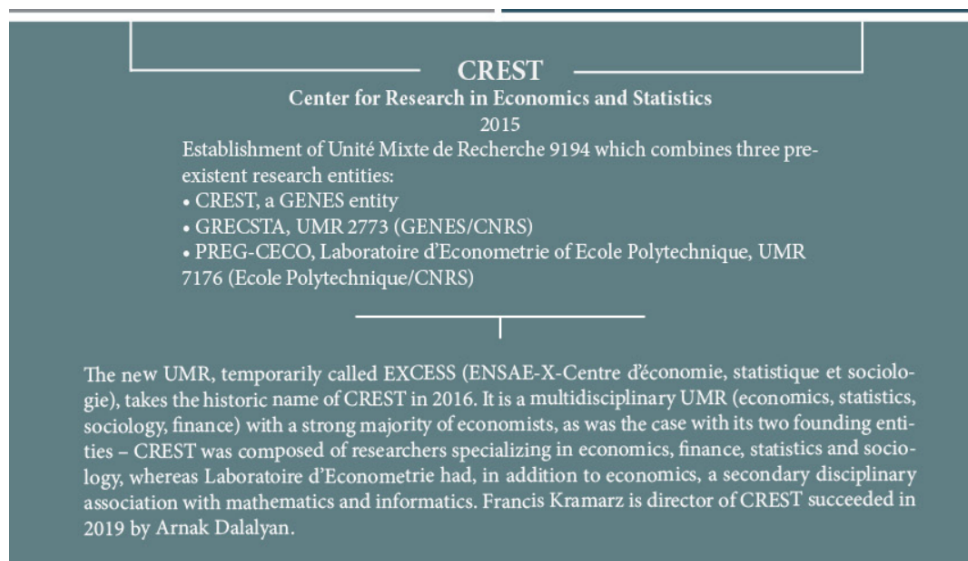


Figure 3.1: Historique du CREST

Le CREST est un centre de recherche dédié à l'application de méthodes quantitatives et de modèles mathématiques pour résoudre des problématiques en sciences économiques et sociales. Il se divise en quatre départements : le pôle Économie, le pôle Statistique, le pôle Finance et le pôle Sociologie. Le CREST est réparti sur deux campus distincts : l'un au sein de l'Institut Polytechnique de Paris, dans les étages supérieurs du bâtiment de l'ENSAE, et l'autre à Bruz-Rennes, dans le bâtiment de l'ENSAI.

Reconnue à l'échelle internationale, cette institution a vu ses chercheurs publier 420 articles entre 2018 et 2021, avec la participation de 1 045 auteurs internationaux dans le cadre de collaborations scientifiques. La répartition géographique de ces collaborations témoigne de l'influence mondiale du CREST.

Si le CREST s'impose comme un centre de recherche de premier plan sur la scène internationale, il convient maintenant d'examiner son fonctionnement interne afin de mieux comprendre les dynamiques qui animent cette institution.

3.2 Se positionner au sein de la diversité des acteurs

En explorant le CREST, on découvre une organisation dynamique, comparable à une fourmilière en pleine effervescence. Ses couloirs sont fréquentés par une diversité de personnes, allant des administrateurs aux chercheurs permanents et affiliés, en passant par les doctorants et quelques stagiaires. Avec plus de quatre-vingt-dix chercheurs permanents, le centre est un véritable carrefour de compétences. Bien que ces professionnels utilisent une multitude de langues, l'anglais et le langage mathématique facilitent leur communication, comme en témoigne le tableau des formules et démonstrations affiché dans le couloir.

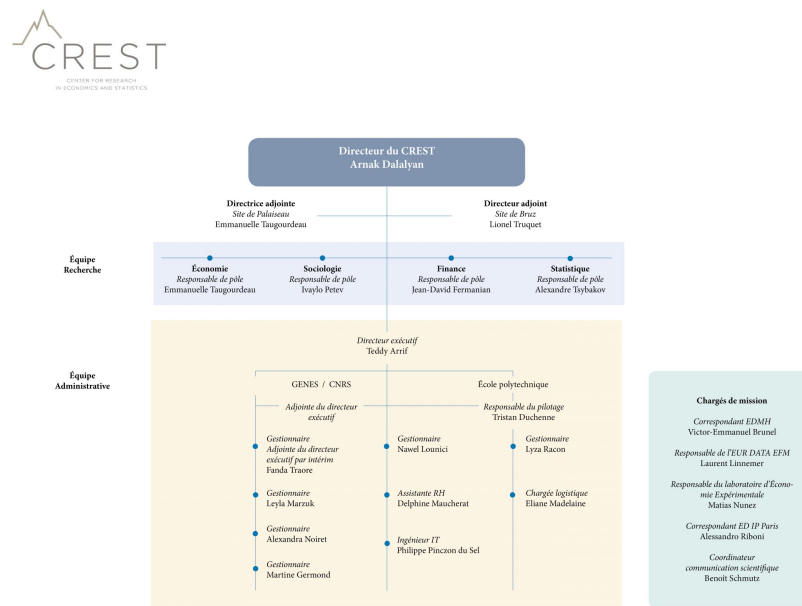
La plupart des chercheurs du CREST disposent de bureaux individuels, comme c'est le cas pour Mme Khaleghi, bien que certains partagent leur espace de travail avec d'autres collègues. En général, chaque chercheur se consacre à ses propres projets tout en collaborant fréquemment avec d'autres équipes. Ils encadrent également leurs doctorants, dont les travaux sont souvent directement liés aux recherches menées par le chercheur. Les horaires de travail sont flexibles : les chercheurs organisent leur emploi du temps autour des conférences, des réunions de suivi et des discussions avec leurs doctorants. Malgré un emploi du temps chargé, ils se retrouvent régulièrement pour déjeuner ensemble, que ce soit dans les cantines ou restaurants voisins, ou sur la terrasse du dernier étage de l'ENSAE pour ceux qui apportent leur propre repas.

Les doctorants, pour leur part, travaillent principalement dans des bureaux partagés par cinq ou six personnes, où ils se concentrent sur leurs thèses, sous la supervision des chercheurs du CREST. Ils forment un groupe soudé, évoluant ensemble pendant plusieurs années dans une atmosphère de camaraderie et de soutien mutuel.

En parallèle de leurs recherches, les chercheurs et doctorants du CREST ont également

des responsabilités d'enseignement. Certains chercheurs donnent des cours magistraux aux élèves de l'ENSAE ou de l'ENSAI, c'est le cas de Mme Khaleghi qui est professeur de Statistique pour les deuxièmes années de l'ENSAE, tandis que les doctorants assurent les séances de travaux dirigés. Ainsi, il n'est pas rare de voir des étudiants circuler dans les couloirs du CREST, et les doctorants sont toujours disposés à aider les stagiaires.

À la tête de cette organisation se trouvent le directeur M. Arnak Dalalyan et l'équipe administrative. Cette équipe se divise en plusieurs branches : GENES/CNRS et École Polytechnique. De plus, chaque pôle de recherche est supervisé par un chercheur responsable, comme l'illustre l'organigramme ci-dessous, datant de juin 2023, qui a été un petit peu modifié depuis, notamment dans l'équipe administrative.



Mise à jour : juin 2023

Figure 3.2: Organigramme du CREST

Dans cette organisation accueillante, il n'a pas été difficile de trouver ma place et de mener à bien ma mission.

4 Ma mission au sein du CREST

4.1 Évolution du travail pendant le stage

Le travail confié par Mme Khaleghi a évolué de manière significative au cours du stage. Dès le premier jour, nous avons dû adapter nos approches à plusieurs reprises et, parfois, ajuster légèrement le sujet pour explorer de nouvelles perspectives sur les grands modèles de langage (LLMs).

4.2 Description de notre mode de travail

Nous organisions au moins une réunion hebdomadaire d'une heure, en visioconférence ou en présentiel à l'ENSAE. Lors de ces réunions, je présentais mes résultats à Mme Khaleghi et lui posais toutes mes questions. Elle me proposait ensuite de nouvelles pistes de recherche et partageait son point d'expertise pour m'aider à progresser. Le reste de la semaine, nous communiquions par mail, soit pour répondre à mes questions, soit pour clarifier certains points de mes travaux.

Mme Khaleghi me fournissait souvent des pistes de recherche qu'elle n'avait pas le temps d'explorer ou des articles que je résumais afin de lui faire gagner du temps. Elle me donnait ensuite son avis ainsi que des recommandations pour orienter mes recherches. À la fin du stage, j'ai rédigé un rapport technique (voir annexe) qui récapitule l'ensemble des expériences et résultats obtenus. Ce rapport est également accessible sur le GitHub qui regroupe toutes les données du stage : <https://github.com/tilian15/summer-internship-2024>. J'ai aussi présenté une soutenance à mon encadrante pour lui exposer mon travail global et final.

4.3 Introduction au sujet du stage

Au cours de ce stage, j'ai eu l'opportunité d'explorer divers sujets liés aux grands modèles de langage (LLMs). L'objectif principal de mon travail portait sur la détection de changements de sujet dans les textes, en utilisant une bibliothèque développée par mon encadrante, permettant la détection de points de rupture dans les séries temporelles.

Étant novice dans ce domaine au début du stage, mes premières tâches ont consisté en des recherches approfondies et à la lecture d'articles sur les modèles de pointe, les réseaux de neurones et les transformeurs. La première semaine a été dédiée à la découverte des LLMs de pointe, tels que GPT, ainsi qu'à l'exploration des modèles plus récents développés par Microsoft et Meta, notamment Phi3 et LLaMA3. Comprendre leur architecture et leur fonctionnement a été un défi, surtout pour les modèles transformateurs, car je n'avais aucune connaissance préalable dans ce domaine. Cependant, à la fin de cette première semaine, j'ai pu présenter à mon encadrante une analyse détaillée de ces modèles, ce qui nous a permis de mener nos premières expériences comparatives sur les performances des modèles Phi3 et LLaMA3.

Une fois que j'avais acquis une compréhension de base des LLMs, Mme Khaleghi m'a introduit à ses propres travaux de recherche, en particulier deux de ses articles portant sur la détection de points de rupture et le clustering de séries temporelles. N'ayant aucune connaissance préalable en séries temporelles et en processus stochastiques, j'ai dû étudier ces concepts en parallèle, avec l'aide d'un ouvrage emprunté au CDI. Cette formation m'a permis de réimplémenter progressivement les expériences décrites dans les articles de Mme Khaleghi.

4.4 Approfondissement du sujet et premières expériences

Trois semaines après le début du stage, j'avais acquis une compréhension significative du fonctionnement des LLMs, ce qui nous a permis d'entamer des expérimentations plus concrètes. L'objectif initial était d'adapter un algorithme de détection de points de rupture, développé par Mme Khaleghi en collaboration avec d'autres chercheurs, appelé PyChest, pour l'appliquer aux grands modèles de langage (LLMs). L'idée était de détecter les

changements de sujet dans un texte en utilisant cet algorithme.

On peut s'interroger sur le lien entre les séries temporelles et les LLMs. Dans les modèles transformateurs et LLMs, les mots sont représentés sous forme de vecteurs. Ces vecteurs, lorsqu'ils sont mis bout à bout, peuvent être traités comme un processus stochastique.

Pour donner un peu de contexte, les LLMs utilisent des *tokens* pour représenter les mots. Un *token* est une unité de texte, comme un mot ou un sous-mot, qui est ensuite transformée en vecteur numérique. Lors de l'entraînement du modèle, les poids des vecteurs sont ajustés pour représenter au mieux le mot, c'est-à-dire que les synonymes ou les mots appartenant à une même catégorie auront des vecteurs similaires. Ce processus permet aux modèles de capturer des relations sémantiques entre les mots, facilitant ainsi les tâches de génération de texte, de traduction, et autres.

Nous avons ensuite exploré comment les modèles parviennent à représenter et à stocker les informations sur les mots à l'aide des vecteurs, un domaine connu sous le nom de *feature engineering*. Les modèles actuels, tels que GPT et les transformateurs, reposent sur des réseaux de neurones entraînés sur de vastes bases de données pour ajuster les vecteurs des mots en fonction de leur définition et du contexte dans lequel ils apparaissent.

Avec l'avènement de GPT et des modèles transformateurs, on parle de représentations dynamiques grâce au mécanisme d'attention. Ce mécanisme permet d'ajuster le vecteur d'un mot en fonction du contexte de la phrase étudiée. Cette approche est couramment utilisée dans les LLMs actuels et est à la base des capacités des IA génératives. En parallèle, il existe des représentations statiques, où un vecteur fixe est associé à chaque mot après l'entraînement. Word2Vec est un exemple de cette approche. Nous avons décidé de nous intéresser à cette méthode, car elle est plus accessible en termes de ressources et peut être facilement utilisée sur un ordinateur personnel.

Nous avons réalisé plusieurs expériences pour comprendre le fonctionnement du modèle Word2Vec, sa fiabilité, notamment à travers des techniques de clustering et de réduction de dimension. Le modèle PyChest étant adapté pour des espaces de faible dimension, il était nécessaire de réduire la dimensionnalité des vecteurs avant de pouvoir l'appliquer.

Nous avons exploré l'effet de la réduction de dimension des vecteurs à l'aide de techniques telles que l'ACP (Analyse en Composantes Principales) et avons mené des tests, notamment sur les distances entre les clusters. Pour valider nos résultats, nous avons util-

isé des méthodes de clustering basées sur des distributions gaussiennes multivariées pour modéliser les vecteurs des mots. Après avoir généré des échantillons basés sur ces distributions, nous avons analysé les distances inter et intra-cluster pour évaluer l'efficacité de nos approches.

Confrontés aux limites de la distance euclidienne, nous avons étudié, en collaboration avec Théo Monet, également stagiaire de Mme Khaleghi, l'utilisation de la distance distributionnelle, introduite par notre encadrante dans un de ses papiers de recherche. Théo m'a aidé à comprendre et à adapter cette distance à notre étude. Nous avons ensuite implémenté cette méthode avec une estimation de densité par noyau, bien que des défis soient apparus lors de l'intégration en haute dimension, notamment en raison des limitations du code Python et du temps d'exécution. C'est à l'initiative de notre encadrante que nous avons travaillé ensemble sur ce sujet, ce qui s'est avéré être une expérience très enrichissante. Théo s'est montré bienveillant et a su rapidement me mettre à l'aise, à tel point que je n'hésitais pas à lui poser des questions, même lorsque celles-ci pouvaient sembler triviales, ni à lui demander de répéter certaines explications. Sa patience et ses qualités pédagogiques ont été des atouts précieux.

Nous avons également exploré la méthode de la *Maximum Mean Discrepancy* (MMD) comme alternative pour améliorer le clustering. Ces méthodes ont été appliquées au modèle Word2Vec, où nous avons étudié l'impact de la réduction de dimensionnalité sur la précision du clustering.

4.5 Conclusion et perspectives

Bien que nos résultats aient été prometteurs, mon encadrante a décidé, vers la fin du stage, de m'orienter vers une nouvelle approche du sujet. Durant les dernières semaines, nous avons étudié les liens potentiels entre les LLMs et la théorie de l'information, en utilisant des techniques de compression telles que Gzip.

5 Le quotidien du stagiaire

5.1 Mon cadre de travail

Dès mon arrivée au CREST, Mme Khaleghi m'a accueilli dans son bureau pour me donner les premières consignes. Elle m'a informé qu'aucun bureau ne m'était réservé, me laissant libre de choisir mon lieu de travail à l'ENSAE. Les premiers jours, je me suis installé au CDI, mais l'affluence des étudiants passant leurs oraux s'est rapidement avérée gênante.

Heureusement, des camarades de promotion, également en stage au CREST mais dans le domaine de l'économie, m'ont proposé de partager leur bureau au 4e étage. Nous étions six dans cet espace spacieux, et j'y ai trouvé un environnement de travail calme et propice à la concentration. Durant mon stage, j'avais à ma disposition un grand écran d'ordinateur, ce qui s'est révélé être un avantage considérable. Au départ, je travaillais sur mon ordinateur portable, mais je me suis vite aperçu que la lecture d'articles scientifiques et de formules mathématiques sur un petit écran fatiguait rapidement mes yeux. L'utilisation de l'écran géant a nettement amélioré mon confort visuel et a augmenté ma productivité, notamment lorsque je devais jongler entre plusieurs fenêtres pour approfondir et assimiler les concepts abordés.

Les bureaux adjacents étaient occupés par plusieurs doctorants que j'avais déjà croisés lors de travaux dirigés en première année à l'ENSAE. Ainsi, nous n'étions pas isolés dans notre espace de travail. Par exemple, un doctorant venait travailler chaque début de semaine dans le bureau situé à côté du nôtre. Bien que nous nous soyons déjà croisés auparavant, notre relation ici était différente. Il se montrait toujours disponible pour répondre à nos questions ou clarifier certains points. D'autres stagiaires de ma promotion occupaient des bureaux voisins, notamment aux côtés de chercheurs en sociologie.

Nous étions donc six à partager cet espace de travail : Martin, Amrei, Paul, Hugo, Anthony et moi, chacun plongé dans ses propres recherches. Étant donné que peu de

monde était présent à cette période sur le plateau de l'ENSAE, nous avons pris l'habitude de déjeuner ensemble au Crous, de jouer au billard pendant les pauses, de faire du sport ensemble le soir, notamment du football, et même de nous balader le week-end. Cette camaraderie a considérablement enrichi mon expérience de stage, rendant la période à la fois agréable et dynamique. Sans leur présence, le stage aurait probablement été beaucoup plus long et solitaire, et je leur en suis sincèrement reconnaissant.

Au fil des semaines, alors que l'été avançait et que les vacances approchaient pour de nombreux chercheurs, le nombre de personnes que je croisais dans les couloirs diminuait progressivement. La plupart des stagiaires avec qui j'avais partagé le début du stage avaient terminé leur mission début juillet. Quand je me suis retrouvé seul, j'ai décidé de retourner au CDI, une fois les oraux terminés. C'est là que j'ai fait la connaissance de Malo, un autre stagiaire, un an plus âgé que moi, qui effectuait également un stage de recherche au CREST. Il est rapidement devenu mon nouveau collègue de travail, et nous avons partagé de nombreux repas le soir. Cependant, avec la fin de son stage début août, je me suis retrouvé de nouveau seul.

Quelques jours plus tard, lors d'une réunion hebdomadaire avec mon encadrante, elle m'a demandé de me renseigner sur un sujet d'expérience auprès de son autre stagiaire. J'ai contacté Théo qui est devenu mon nouveau "collègue" de travail. Il s'avère que Théo était dans le même groupe d'espagnol que moi cette année, mais nous n'avions jamais vraiment eu l'occasion de discuter auparavant. Nous avons donc trouvé une salle à l'ENSAE qui est devenue notre nouveau bureau. Théo m'a avoué qu'il commençait lui aussi à se sentir seul, et nous avons donc passé le reste de notre stage ensemble.

Théo, qui a un an de plus que moi, était présent durant les trois mois où j'ai travaillé. Sa présence était rassurante, d'autant plus qu'il apportait son expérience à notre collaboration. Nous avons partagé la plupart de nos repas ensemble pendant le dernier mois. Théo est une personne très intéressante, ayant fait l'ENS en physique avant de vouloir se reconvertir dans la recherche en statistique. Sa compagnie a grandement contribué à rendre la fin de ce stage plus enrichissante et agréable.

5.2 Une journée typique au CREST

J'habite à 15 minutes de marche du CREST. Cette marche quotidienne, à la fois agréable et bénéfique, symbolisait pour moi la transition entre ma vie privée et le monde du travail.

J'arrivais au CREST aux alentours de 9h, moment où peu de chercheurs étaient déjà présents. Au fil de la matinée, le couloir se remplissait progressivement. La flexibilité des horaires est une caractéristique du métier de chercheur. Les chercheurs sont souvent en déplacement pour des réunions, des conférences, ou des collaborations, ce qui était fréquemment le cas pour mon encadrante, d'où nos réunions en visio. Ma journée de travail était généralement divisée en deux périodes : trois heures le matin et 4 heures l'après-midi. Je rentrais chez moi vers 17h30, mais j'adaptais mon emploi du temps en fonction des exigences de la journée, en arrivant parfois plus tôt ou plus tard et en ajustant mon départ en conséquence. Pour équilibrer mon emploi du temps, j'avais également commencé à faire beaucoup de sport ; je m'efforçais d'aller courir au moins une fois par jour et me suis même inscrit à un marathon pour me motiver davantage.

Ce que j'apprécie particulièrement dans le domaine de la recherche, c'est la flexibilité des horaires. Travaillant sur mon propre ordinateur, il m'est arrivé à plusieurs reprises d'avoir une idée en dehors des heures de bureau, parfois en lisant quelque chose ou simplement en réfléchissant chez moi le soir. Dans ces moments-là, je n'hésitais pas à sortir mon ordinateur et à me plonger dans mon travail, même tard dans la nuit. Il m'arrivait de travailler jusqu'à 2h du matin, et souvent, ces sessions nocturnes se révélaient plus productives que plusieurs journées de travail au bureau. Le lendemain, je m'autorisais alors à dormir un peu plus et n'arrivais au bureau que pour le déjeuner. Ces moments de travail nocturne étaient parmi les plus efficaces, probablement parce que la nuit est propice à la concentration, avec son calme et son absence de distractions.

À midi, je retrouvais d'autres étudiants de l'ENSAE, également stagiaires au CREST, ainsi que mes collègues de bureau pour déjeuner ensemble au restaurant universitaire. Quand le restaurant a fermé, nous avons alors apporté nos propres repas, que nous prenions sur la terrasse du quatrième étage, tout comme de nombreux chercheurs. Parfois, pour nous détendre, nous jouions à des jeux de société ou au billard, ce qui nous permettait de déconnecter complètement de nos recherches.

6 Mes observations et impressions

6.1 Relations avec mon encadrante

Dès le début de mon stage, Mme Khaleghi m’a réservé un accueil chaleureux. Dans le cadre de notre collaboration, je n’hésitais pas à poser des questions, même si certaines pouvaient sembler moins pertinentes, et à demander des éclaircissements lorsqu’elles étaient nécessaires. Elle faisait preuve de patience et adoptait une approche pédagogique. Grâce à ma nature ouverte et à mon aisance à discuter, je me sentais libre d’exprimer toutes mes idées et préoccupations. Mon encadrante encourageait cette ouverture et s’efforçait de rendre ses explications claires en utilisant des illustrations sur le tableau ou des exemples concrets. Cette méthode s’est révélée particulièrement efficace pour moi, étant donné ma préférence pour les représentations visuelles.

6.2 Les carrières dans la recherche

Les échanges avec les chercheurs et les doctorants m’ont permis d’acquérir une meilleure compréhension du milieu de la recherche académique, notamment des rôles d’enseignant-chercheur, d’ingénieur de recherche et de doctorant.

6.2.1 Le rôle de maître de conférences

Les maîtres de conférences sont des enseignants-chercheurs, un statut particulier qui requiert un doctorat et une appartenance à un corps de chercheurs au sein d’un établissement public. Au CREST, les promotions sont limitées chaque année et nécessitent l’accord de l’ensemble de l’équipe, ce qui peut engendrer des tensions mais aussi dynamiser le département.

En tant qu’enseignants-chercheurs, les maîtres de conférences doivent équilibrer leurs responsabilités pédagogiques avec leurs activités de recherche. La préparation des cours et des travaux pratiques peut empiéter sur leur temps de recherche. Le manque de financements peut également freiner les activités de recherche, obligeant les enseignants-chercheurs à chercher des financements eux-mêmes.

Je n’ai pas observé une pression significative pour publier de la part des enseignants-chercheurs, mais cela pourrait être dû à la courte durée de mon stage. À titre d’exemple, Mme Azadeh Khaleghi, mon encadrante, a obtenu son doctorat en mathématiques appliquées en 2009 et est actuellement professeur des universités. Ses recherches au Crest portent sur les statistiques et le machine learning, plus précisément sur la conception de modèles non paramétriques pour les séries temporelles à mémoire longue.

6.2.2 La carrière de doctorant

En discutant avec des doctorants, j’ai perçu que les enseignants-chercheurs ne sont pas directement impliqués dans la recherche au quotidien, cette tâche étant principalement réservée aux doctorants. Les enseignants-chercheurs, quant à eux, définissent les orientations de recherche à long terme, les thèses ayant une durée de trois ans. En revanche, les doctorants possèdent une expertise approfondie dans leur domaine spécifique.

J’ai eu l’opportunité de déjeuner avec des doctorants et de discuter de leurs parcours, de leurs thèses, des sujets de recherche et de leurs aspirations futures. Ces échanges ont été particulièrement instructifs. Contrairement à ce que j’ai observé chez les enseignants-chercheurs, les doctorants doivent publier des articles, ce qui peut être stressant en raison des exigences de validation par des comités. Les articles peuvent être rejetés si les idées ne sont pas considérées comme suffisamment innovantes.

Les doctorants bénéficient d’une grande autonomie, ce qui peut parfois être un inconvénient. Les sujets de thèse étant souvent larges, les doctorants doivent affiner leur domaine de recherche, et un manque de perspective peut les conduire à choisir des pistes non fructueuses, entraînant une perte de temps considérable.

6.2.3 Autres Observations

J’ai constaté que, dans le domaine de la recherche, les chercheurs ont une grande liberté dans la gestion de leur temps de recherche, car il est important d’être en forme, motivé et

concentré. De plus, dans un centre de recherche, on entend une multitude de langues différentes en raison des nombreuses collaborations internationales. Les frontières semblent inexistantes.

La principale caractéristique du métier de chercheur, de mon point de vue, est la sensation de solitude, malgré la présence constante de collègues. Un chercheur n'est pas isolé ; il côtoie de nombreux autres chercheurs et doctorants, rencontre des étudiants (surtout s'il donne des cours), et échange avec divers interlocuteurs externes. Cependant, face à son sujet de recherche, il peut se sentir rapidement démuni et perdu, car il n'y a personne pour lui indiquer la direction à suivre ni pour lui fournir les réponses qu'il recherche. Cette solitude pourrait être moins ressentie si le chercheur collabore étroitement avec d'autres, ce qui est souvent le cas, mais je n'ai pas eu l'occasion de collaborer directement si ce n'est qu'avec mon encadrante. Lors de mes propres recherches, j'ai éprouvé une sensation de vertige, devant avancer de manière autonome sans savoir exactement quel serait le prochain pas. Je pense que les chercheurs développent des stratégies de réflexion, semblables à des plans d'attaque, pour aborder leur sujet. Avec davantage d'expérience, ils doivent avancer avec plus de confiance. De plus, la recherche implique constamment l'exploration et la découverte de nouvelles connaissances, et la fierté ressentie lorsque l'on aboutit est d'autant plus grande. Sans ces défis, je doute que la satisfaction de l'accomplissement soit aussi forte.

Enfin, j'ai observé que le métier de chercheur comporte de multiples facettes. Ce n'est pas une profession monotone où l'on reste enfermé dans un bureau en permanence, comme on pourrait le croire. En plus de la recherche pure, un chercheur écrit et publie des articles, peut être amené à coder, assiste à des conférences, lit de nombreux articles pour se tenir informé des avancées, donne des conférences et participe à des réunions pour obtenir des retours externes sur ses recherches, et entretient des relations fréquentes avec des collaborateurs. Ainsi, je n'ai pas eu l'impression de faire toujours la même chose ou de m'ennuyer car je jonglais entre expériences, code, et lecture d'article.

Les différents chercheurs ont l'habitude de manger ensemble et de partager leurs pauses, pendant lesquelles ils discutent souvent de sujets qui n'ont rien à voir avec leur recherche. Parfois, ils parlent de leur sujet d'étude, ce qui leur permet d'avoir de nouvelles idées, d'explorer de nouvelles approches et d'obtenir un regard extérieur différent du leur. Ils s'entraident également sur certains points. J'ai pu le constater avec Théo : un jour,

j'avais un problème de code et je n'arrivais vraiment pas à le corriger, et il a volontiers accepté de m'aider à le résoudre.

En me baladant dans les couloirs, j'ai aussi remarqué une photo d'équipe de foot sur une des portes. En regardant de plus près, je reconnaissais quelques visages qui étaient en fait des professeurs-chercheurs de l'ENSAE. En effet, les chercheurs trouvent également du temps pour décompresser le soir et faire du sport avec les autres chercheurs. Cela est très important pour eux, car cela leur permet de maintenir un contact social différent de celui lié à leur recherche et de se défouler. Certains préfèrent aller courir en groupe le midi, tandis que d'autres jouent aux cartes... En tout cas, chacun s'occupe et se détend à sa manière.

Grâce à ces observations et impressions, je suis désormais en mesure de faire un bilan objectif de mon stage.

7 Bilan

Ce stage a été une expérience très enrichissante et stimulante pour moi. J’ai eu la chance de travailler sur un sujet qui m’a réellement intéressé, ce qui m’a permis de mettre en pratique et d’approfondir les connaissances acquises durant l’année précédente. J’ai non seulement développé des compétences techniques et professionnelles précieuses, telles que la synthèse d’informations, la vulgarisation de concepts complexes, et l’utilisation de divers logiciels comme LaTeX, Python, et R, mais j’ai également eu un aperçu approfondi du métier de chercheur et des défis auxquels les doctorants sont confrontés.

Le cadre de travail flexible a été un point positif important, avec des horaires adaptables. De plus, la proximité du lieu de travail a été un confort appréciable, étant donné que j’habite à proximité et que je connaissais déjà les lieux. Cependant, je reconnais que ce confort relatif pourrait aussi être perçu comme une limite, car je n’ai pas eu l’occasion de me confronter à des situations plus exigeantes, comme des déplacements à l’étranger ou des environnements de travail très différents.

De manière objective, je reconnais que le travail que j’ai réalisé diffère de celui d’un chercheur expérimenté en raison de mon niveau de connaissances limité, de mon manque d’expérience et des contraintes de temps. Néanmoins, il est essentiel de garder à l’esprit que ce stage est avant tout une ouverture au monde professionnel. Par conséquent, mes impressions doivent être nuancées, car elles sont probablement différentes de celles d’un chercheur aguerri.

Une autre leçon importante que j’ai tirée de cette expérience est la valeur du contact humain en recherche. La solitude face à des problèmes complexes peut être un obstacle, et un regard extérieur peut souvent offrir des perspectives précieuses. La persévérance est également cruciale, surtout lorsque l’on se heurte à des défis difficiles. Ces leçons sont des aspects importants à considérer pour mes futures expériences professionnelles et académiques.

8 Conclusion

En somme, ce stage a été extrêmement enrichissant. Il m’a permis de progresser considérablement et d’acquérir des compétences précieuses en statistiques et en machine learning, en particulier dans le domaine des modèles de langage (LLMs). J’ai également constitué un répertoire précieux de modèles, d’algorithmes et de codes qui me seront certainement utiles à l’avenir. De plus, mon niveau d’anglais s’est fortement amélioré grâce à la lecture d’articles, à la rédaction de synthèses en anglais et aux interactions avec ma tutrice, qui se faisaient parfois en anglais.

Cette expérience de stage a profondément influencé ma réflexion sur l’avenir. En effet, au fil des semaines, l’idée de poursuivre une thèse a pris de plus en plus de place dans mon esprit. Les découvertes que j’ai faites dans le domaine des modèles de langage et l’opportunité de plonger dans des problématiques de recherche complexes m’ont donné un aperçu captivant du travail de chercheur. La perspective de contribuer à l’avancement des connaissances, tout en approfondissant mes compétences, est devenue une option que je considère sérieusement. Cette expérience m’a non seulement ouvert des portes, mais a également renforcé mon intérêt pour la recherche académique, faisant de la thèse une voie potentielle pour l’avenir.

Cependant, je souhaite aussi explorer une autre facette du monde du travail en effectuant un prochain stage dans une entreprise privée, axée sur la data science et le machine learning, de préférence dans le secteur financier. Cette expérience me permettra d’obtenir un autre point de vue et de travailler davantage en équipe, ce qui contraste avec ce que j’ai vécu lors de ce premier stage. Cela m’aidera à avoir une vision plus complète pour orienter mon parcours futur.