

GÉOSTATISTIQUE ET PROBABILITÉS APPLIQUÉES



La pratique de la Géostatistique est, d'abord, une occasion de rencontres: rencontres entre des champs d'application variés et parfois inattendus, entre des problématiques sans cesse renouvelées, et également une confrontation entre des objectifs et contraintes purement techniques d'une part et les exigences sociales, économiques, environnementales d'un monde complexe d'autre part. Autrement dit, tout en étant fiers de ce que le néologisme "géostatistique" jadis forgé à l'École des Mines ait trouvé droit de cité dans le Petit Larousse, il est satisfaisant d'observer, au fil des ans, que l'immuable définition qu'en donne le dictionnaire s'éloigne de plus en plus de la réalité, et que notre discipline trouve à s'exprimer bien au-delà de la simple estimation des gisements miniers. De fait, dans tout domaine où des jeux de données numériques présentent une organisation spatiale ou temporelle, la Géostatistique a les outils pour apporter un éclairage original, à la fois constructif et sans concession. Il semble que cet aspect transversal et non-conformiste de la Géostatistique constitue désor-

mais son caractère dominant au regard des optionnaires, et nous ne manquons pas dans le futur de justifier cette appréciation. Ainsi, chaque année, la diversité des vœux des étudiants constitue une chance exceptionnelle de tester des méthodes nouvelles et de parcourir des domaines nouveaux, et la garantie d'insuffler un surplus de dynamisme à l'équipe encadrante. Mais la mise en œuvre d'une Géostatistique de qualité exige en permanence d'assurer un équilibre, parfois délicat, entre des exigences souvent contradictoires: garantir une rigueur théorique indispensable à la fiabilité des résultats tout en conservant un point de vue pragmatique et réaliste afin que les conclusions abstraites trouvent à s'appliquer sur le terrain. Sans oublier une indispensable déontologie, dans des domaines où souvent les contraintes économiques ou environnementales soumettent le géostatisticien à des pressions qui ne relèvent plus de la science ou de la technique...

Ouverture et équilibre: c'est dans cet esprit que nous avons continué à proposer à la promotion actuelle un voyage

de deux semaines en Guyane où, dans le contexte inhabituel et parfois tourmenté d'un DOM, les visites à des laboratoires, à des industriels et à des organismes institutionnels ont permis tout à la fois d'élargir l'horizon des optionnaires et de susciter un échange ouvert et fructueux avec nos interlocuteurs. Partie intégrante de la scolarité, la mission en Guyane constitue pour les optionnaires la phase d'initiation à la réalité du terrain.

Enfin, le souci d'ouverture s'est à nouveau exprimé au niveau des sujets de l'option. Outre la variété des champs d'application, une multiplicité de méthodes statistiques est employée. Nous avons ici l'illustration que notre démarche méthodologique peut trouver à s'appliquer dans de multiples domaines et cela souligne le caractère généraliste de l'option Géostatistique et Probabilités Appliquées, tant en ce qui concerne les champs applications abordés que les méthodes mathématiques mises en jeu.

**Emilie CHAUTRU, Thomas ROMARY,
Hans WACKERNAGEL**

GEOSTATISTICS AND APPLIED PROBABILITY



The aim of Geostatistics is to study quantitative phenomena that are structured in space and/or time. Engineers are almost inevitably faced with problems related to geostatistical techniques such as the evaluation of natural reserves, the analysis of time series, cartography, etc., and, broadly speaking, any processing of "regionalized variables" according to the terminology of G. Matheron, the founder of geostatistics.

The Geostatistics and Applied Probability Minor gives priority to probabilistic models and statistical methods, and in particular focusses on their application to the processing of spatial and temporal data.

The training in this Minor is aimed essentially at providing students with a critical mastering of some of the tools which they might need to use during their internship. As an introduction to "geostatistics in the field", the Minor provides an opportunity of entering into contact with companies and personnel working with geostatistics in fields of application that correspond as much as possible to the themes of particular interest to the students. It is essential that the students acquire a sense of balance between an empirical approach and a mathematical approach

to a problem, a sense of balance which the third-year internship will put into practice in real terms.

The presentations this year again reflect the great variety of themes and applications that are encountered in this branch of applied statistics.

**Emilie CHAUTRU, Thomas ROMARY,
Hans WACKERNAGEL**

GÉOSTATISTIQUE ET PROBABILITÉS APPLIQUÉES

GEOSTATISTICS AND APPLIED PROBABILITY

Lundi 25 juin Monday 25th June ■ L213



Gaspard
DEVEYRAC

9h-10h

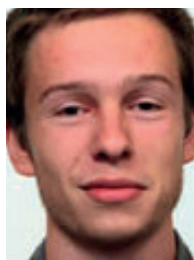
Modélisation du réseau optimal de collecte du biométhane

Biomethane collection network modeling

La transition écologique, l'indépendance énergétique de la France ou encore la recherche de nouvelles sources de revenus pour nos agriculteurs sont des sujets majeurs qui rythment les débats actuels en France. Particulièrement ambitieux, notre pays a fixé un objectif d'injection de 30% de biométhane en 2030 et de 100% à l'horizon 2050. Les contraintes technico-économiques pour créer un tel réseau sont nombreuses, notamment parce que les lieux de consommation se trouvent majoritairement en ville, alors que les lieux de production sont situés dans les zones rurales. Comment peut-on donc développer, en seulement 30 ans, un tel réseau de collecte de biométhane ? Une première étape consiste à évaluer les potentiels de production sur l'ensemble du territoire grâce aux données agricoles afin de prédire les futurs lieux de concentration. Puis, grâce à des techniques algorithmiques d'optimisation, il s'agit de tracer le réseau optimal permettant de relier les potentiels de production soit au réseau de transport, soit directement à celui de distribution. En affinant le modèle, il est alors possible de prévoir le coût de développement du réseau à l'horizon 2050, tout comme le type de canalisation à installer afin d'obtenir un réseau adapté à la fois au moment de sa mise en service, mais aussi à terme. Enfin, le calcul de différents indicateurs économiques permet d'évaluer finement la pertinence des projets ainsi que leur priorité de développement. Initialement réalisé pour la région Grand Est, l'outil développé a été choisi afin de tracer l'ensemble du schéma directeur sur la totalité du territoire français.

The ecological transition, our energy independence and also new sources of income for our farmers are some burning issues in today's debates in France. Our government has set a very ambitious target: 30% of renewable natural gas in 2030 and 100% by 2050. However, this goal raises many technical and economic constraints, especially because places of consumption are located in cities and production sites in rural areas. Thus, how to develop a comprehensive and adequate network to collect biomethane in only 30 years? The first step required to assess production potentials throughout the national territory thanks to agricultural data in order to forecast future locations of fermentation plants. Then, by using algorithmic techniques of optimisation, the developed tool draws the optimal network able to connect all production potentials to the transportation network or directly to the distribution one. By refining the model, the algorithm is then able to forecast the cost incurred developing the 2050 network, as well as predicting the kind of pipe needed in order to have a well-sized network. Moreover, the assessment of different economic indicators provide tools to evaluate in more detail the relevance of the projects and also to establish development priorities. Initially created for the Grand-Est Region, the tool has been chosen to design the full biomethane collection network throughout the French territory.

GRDF- Maitrise d'ouvrages
Nancy



Rémy
DUBOIS

10h-11h

Classification des motifs de localisation sub cellulaires par Transfer Learning

Gene's subcellular localisation patterns classification through Transfer Learning

L'expression d'un gène est le processus par lequel l'information génétique est transmise depuis l'ADN jusqu'à une protéine finale. L'étude de l'expression des gènes était, jusqu'à peu, essentiellement faite en étudiant la quantité de molécules d'ARN par cellule (quantité appelée le niveau d'expression). Néanmoins, de récentes études suggèrent que la localisation de ces molécules importe autant, sinon plus, que leur quantité. Classifier chaque gène et le faire correspondre à un motif de localisation requiert des systèmes capables d'identifier des corrélations spatiales de complexité élevée. L'analyse d'image par ordinateur a connu de grandes avancées au cours des dix dernières années,

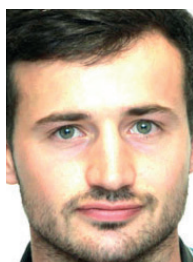
Gene expression is the process by which genetic information is transformed from DNA into functional products. The study of gene expression has traditionally been done by studying the amount of RNA molecules per cell, but recent advances indicated that localization patterns of those molecules within the cell matter as much. Classifying each gene to a given localization pattern would require systems able to identify spatial correlation of high complexity. Computer vision has seen great advances over the last decade in the field of image classification, notably through the use of supervised learning algorithms. However, as in many bioinformatics pro-

notamment grâce au développement d'algorithmes dits "d'apprentissage supervisé". Cependant, comme dans de nombreuses études de bio-informatique, l'absence de vérité connue empêche d'avoir recours à de tels algorithmes, qui demandent en retour de leur efficacité, l'accès à de très grandes quantités de données et leur réponse attendue.

Le principe de ce projet est de mettre à profit un outil de synthèse d'images artificielles, afin d'entraîner ces algorithmes puissants, suivant des méthodes dites de "Transfer Learning". Cet outil de synthèse d'images a été construit grâce à l'observation humaine de grandes quantités d'images, et la traduction de motifs récurrents en informations probabilistes par des biologistes et experts du domaine. De manière générale, le Transfer Learning fait référence à une situation où le but est de modéliser une distribution statistique (appelée cible), à partir de connaissances accumulées depuis une autre distribution (appelée source), avec un biais systématique entre la cible et la source. De fait, le principal challenge est de construire des modèles montrant une capacité supérieure à généraliser les connaissances accumulées depuis la cible, de pair avec une habileté à maintenir une variance de prédiction faible. Des méthodes de Transfer Learning célèbres font désormais partie intégrante des méthodes les plus efficaces de vision par ordinateur, comme le pré-entraînement par exemple.

blems, the absence of ground truth retains one to make any use of such techniques in the described project. The idea of this project is to leverage an existing image synthesis framework in order to train such supervised learning algorithm, in the fashion of what is mostly known as Transfer Learning. This image synthesis routine was built upon human observations of image data, and expert's hand-crafted features of the observed localization patterns. In general, Transfer Learning refers to a situation where one has to model a statistical distribution (the target domain), based on the knowledge accumulated on another distribution (the source domain), with a systematic bias in-between. Therefore, the main challenge is to build algorithms showing a superior ability to generalize knowledge acquired from the train set, along with a capacity to maintain its predicting variance low enough. Some transfer learning techniques are now famous (such as pre-training algorithms, or learning domain-independent features through gradient-reversal layers) and included in some state-of-art or prize-winner machine learning algorithms.

**Centre de Bioinformatique
MINES ParisTech**
Fontainebleau



Martin
CHARACHON

11h-12h

Effacité énergétique dans l'industrie par implémentation d'une solution complète fondée sur de l'intelligence artificielle, de la collecte à la valorisation des données

Industrial Energy Management implemented with artificial intelligence: from data collection to valorisation

L'énergie est devenu un actif que les entreprises peuvent mesurer, gérer, acheter et vendre. Ainsi, l'efficacité énergétique se révèle être un enjeu majeur dans l'optique de mieux maîtriser les consommations, et constitue un véritable atout économique pour les entreprises. La start-up Parisienne METRON a choisi une approche innovante de gestion intelligente de l'énergie. En se concentrant sur les nouvelles technologies comme le Big Data, l'Intelligence Artificielle ou encore l'IoT, METRON permet aux entreprises de maximiser la valeur de leurs ressources énergétiques. À cette fin, le choix a été fait de développer une solution complète de gestion de l'énergie pour les besoins de l'industrie (la plateforme Metronlab), basée sur la collecte et la valorisation des très nombreuses données générées en temps réel sur les sites de production. Par l'implémentation de telles technologies, METRON cherche donc à automatiser la détection de ces gisements énergétiques; ainsi qu'à estimer quantitativement leur importance pour proposer des projets d'optimisation efficaces aux industriels. L'objectif de ce stage est tout d'abord de découvrir l'univers METRON, saisir ses particularités et ses enjeux. Ainsi, plusieurs problématiques sont abordées au cours de ce projet d'option. Une première partie orientée purement R&D, avec le développement

Energy has become an asset that companies can measure, manage, procure and sell. Indeed, Energy Management turned out to be a great economic and environmental asset, with dynamic drivers for industries. METRON, a start-up based in Paris, provides an innovative approach in Big data Energy Management Technologies & Services allowing companies to maximize the value of their energy resources. So METRON chose to develop a complete energy management solution for industrial sites, basing its approach on gathering and analyzing large amounts of data that are generated in real time by automats in the manufactures. They seek to detect the major energy savings potentials in highly energy-intensive industrial processes, with the final goal of offering quantified insights to industry managers about implementable energy efficiency projects and their consequences in term of energy and economic savings.

The aim of the project is first to discover METRON universe, its characteristics and drivers. In this way, several issues are tackled during this internship. At the beginning R&D projects were explored, such as the development of drift detection or non intrusive load monitoring algorithms, or even more, a benchmark of algorithms (for regression), adapted to types and size of different data sets. Then, a case study has

d'algorithmes génériques de détection précoce de dérives, de désagrégation de courbes de charges, ou bien encore un benchmarking d'algorithmes de machine learning (orienté régression), adaptés aux différents types et tailles de données. Puis, tout en poursuivant en parallèle les sujets de recherche, une application concrète à un cas client, de certaines méthodes développées en R&D, ainsi que de nouvelles problématiques.

been performed in order to implement what has been developed during R&D and to find out about some other issues.

METRO
Paris



Guillaume
TAUZIN

13h30-14h30

Etude géostatistique de données géophysiques et géochimiques pour caractériser une pollution aux hydrocarbures dans les sols

Geostatistical study of geophysical and geochemical data to characterize hydrocarbon pollution in soils

La géostatistique a tendance à devenir de plus en plus déterminante dans le domaine des sites et sols pollués. En effet, le coût élevé des investigations incite de plus en plus à recourir à cette discipline, afin d'estimer en tout point la teneur en composants indésirables présents dans les sols ayant hébergé une activité polluante durant un certain temps. Néanmoins, les résultats obtenus ne représentent pas toujours une bonne estimation en raison de la quantité limitée d'informations dont on dispose.

Le but de ce stage est d'intégrer des données auxiliaires de géophysiques et de géochimie, afin de construire des estimateurs qui permettront d'améliorer la prédiction des teneurs en polluants sur un site ayant hébergé pendant 30 ans une usine de distillation de goudron. En effet, d'une part ces données sont beaucoup moins chères et plus facile à obtenir car non intrusives, et d'autre part celles-ci sont généralement bien plus nombreuses que les données principales. Elles représentent ainsi un atout considérable pour ce type d'étude. On cherchera ainsi à tester la fiabilité de ces estimateurs par diverses méthodes, incluant la corrélation avec les variables d'intérêt ainsi que la validation croisée, puis à estimer les teneurs en polluants sur l'ensemble du site en prenant en compte l'apport d'information par les données auxiliaires que nous aurons à notre disposition.

Geostatistics tends to become more and more determinant in contaminated sites and soils. Indeed, high costs of investigation motivate more and more to resort to this discipline, in order to estimate in all locations the content of unwanted components being in the soils which received polluting activities. Nevertheless, obtained results are not always a good estimation owing to the limited amount of information we have.

The aim of this internship is to incorporate auxiliary data from geophysics and geochemistry, to allow us to build estimators which help us to improve the prediction of the content of pollutants in a site which sheltered a tar distillation factory during 30 years. In fact, on one hand this data are far less expensive and easier to get because they are not intrusive, on the other hand they are generally much more numerous than main data. Therefore auxiliary data are a considerable asset for this kind of study. Thus we seek to test the reliability of these estimators by using various methods, such as correlation with main data and cross-validation, then we evaluate the content of pollutants in the whole site taking information provided by auxiliary data into account.

Envisol
La tour du Pin



Pierre
POPINEAU

14h30-15h30

Construction d'une base de données géospatiale et quantification des inégalités de niveau de vie au niveau communal en France

Building a geospatial database and quantification of well-being inequalities at city level in France

L'initiative du vivre mieux de l'OCDE a été lancée en 2011 par la commission Stiglitz-Sen-Fitoussi. Elle se base sur la mesure de onze indicateurs (logement, revenus, éducation, vie en société, emploi, environnement, santé, satisfaction personnelle, équilibre vie

The OECD Better Life Initiative was created in 2011 from the conclusions of the Stiglitz-Sen-Fitoussi commission. The OECD developed eleven indicators (housing, income, education, community, jobs, environment, governance, health, life satisfaction,

personnelle-professionnelle, gouvernance, sécurité) permettant de mesurer le niveau de vie à l'échelle d'un pays, afin de quantifier les inégalités de niveau de vie entre les pays membres de la zone OCDE, et de construire des politiques de développement sur mesure pour les pays membres. Dans le cadre de cette initiative, le directoire des statistiques de l'OCDE souhaite poursuivre l'étude des inégalités de niveau de vie au niveau infranational dans les pays membres.

Le but de ce stage est de construire et de maintenir une base de données géospatiale au niveau communal en France en se basant sur les données récoltées par l'INSEE (revenus, éducation, taux de chômage, espérance de vie). Dans un second temps, il faudra ensuite développer des outils de traitement de données afin de compléter les séries manquantes par des méthodes probabilistes (MCMC, échantillonnage de Gibbs). Enfin, les données seront utilisées afin de construire un indicateur de niveau de vie au niveau communal en France afin de quantifier les inégalités de niveau de vie sur le territoire. À terme, la méthodologie appliquée en France a vocation à être utilisée dans les autres pays membres de l'OCDE, afin de pouvoir quantifier localement les inégalités spatiales de niveau de vie dans tous les pays membres de l'OCDE

safety, work-life balance) to measure well-being at country-level, in order to quantify well-being differences between OECD countries and develop adequate development policies for member states. As a part of this initiative, the statistics directorate has initiated measurement campaigns at regional level in member states.

This internship aims at building and maintaining a geospatial database at city level in France, using data collected by INSEE.(income, jobs, life expectancy, housing, educational attainment). Then, we will extend coverage of the series by using probabilist algorithms (MCMC, Gibbs sampling). Finally, the data will be used to develop a better life index at city level in France in order to measure well-being inequalities in France. In the long-term, the methodology used in France will be applied to all the other OECD member states.

OECD

Boulogne Billancourt



Omar
ELIAS

15h30-16h30

Analyse et prévision d'arrêts maladies

Analysis and forecasting of sick-leave risk

Dans le conseil en data science, deux aspects sont primordiaux : la partie cadrage métier, c'est à dire la compréhension du problème d'un client, de ses enjeux et objectifs, puis la traduction en un problème data science; la partie analyse, avec le développement de modèles visant à répondre aux problèmes du client, traduits en KPI (key performance indicators) à optimiser (réduire l'attrition des clients, comprendre dans quelle situation une machine tombe en panne...). Dans le cadre de mon stage, les domaines principaux que j'explore sont l'actuariat et l'assurance, notamment l'analyse des arrêts maladies des employés et les risques financiers que cela engendre pour les entreprises. Plusieurs pistes sont étudiées, dans le but de créer des modèles par exemple de prévision du nombre d'arrêts maladies survenant en un an dans une entreprise ou de la charge financière que cela pourrait lui coûter (pour donc adapter la tarification des assurances auxquelles elle souscrit). Des méthodes classiques de machine learning, en plus d'établissement de «scoring» par individu pour évaluer leur risque à tomber en arrêt maladie, et des algorithmes de recherche de sous-groupes pour trouver des tendances parmi les sous-populations d'une entreprise, sont développées.

There are two essential aspect in data science consulting : the business framing part, ie the understanding of a customer's strategic question, its issues and objectives, then the translation into a data science problem; the analysis part, with the development of models aimed at answering the customer's issues, translated into KPIs (key performance indicators) to optimize (reduce the attrition of customers, understand in which situation a machine breaks down ...).

As part of my internship, the main areas I explore are actuarial science and insurance, including the analysis of employee sick leave and the financial risks entailed for companies. Several possibilities are to be studied, with the aim of creating models such as forecasting the number of sick-leave occurring in a year in a company or the financial charge it would generate (in order to adapt the pricing of insurances the company subscribed). Classical machine learning methods, in addition to individual scoring to assess their risk of falling into sickness, and subgroup discovery algorithms to find trends among the subpopulations of a company, are developed

QUINTEN

Paris

Ximeng
QIU

16h30-17h30

Performance de la collection et de l'assortiment accessoires

Collection and assortment performance of accessories business

Au sein de la direction Merchandising et plus particulièrement de l'équipe en charge de l'analyse business et du monitoring de la Collection et de l'assortiment des accessoires, la mission est de rationaliser les grands équilibres aux accessoires et d'émettre une recommandation sur la taille de collection pour chacune des catégories et de construire un modèle d'assortiment adapté aux enjeux du business accessoires et permettre ainsi de couvrir le business tout en ayant une offre optimisée et efficace.

Dans ce contexte, le but de ce stage est d'analyser les historiques de ventes au niveau catégorie/produit, zone et groupement de magasin (cluster), afin de proposer des assortiments différenciants, efficaces et en ligne avec la stratégie business des magasins. L'enjeu est de donner une recommandation trouvant un équilibre entre le cœur business, la place de la nouveauté, l'expérimentation de territoires nouveaux et le visuel merchandising en magasin. Enfin, l'objectif est de construire des outils d'aide à la prise de décision pour les prochaines collections.

Within the Merchandising Direction and particularly in the team of Business Analysis and Collection and Assortment Monitoring, our mission is to rationalize the balance of accessories, to make recommendation of collection size for each category, and to create an assortment model adapted to the challenge of accessories which allow to cover the business while having an optimized and efficient offer in each retail store.

In this context, the aim of this internship is to analyze historical sales at the level of category / product, zone and store (cluster) in order to offer differentiating and efficient assortments to stores which are in line with the business strategy. The challenge is to give a recommendation integrating a balance between the core business, the place of novelty, the experimentation of new territories and the merchandising visual in store. Finally, the goal is to build decision-support tools for next collections.

Louis Vuitton
Paris

Mardi 26 juin Tuesday 26th June ■ V111

Mohammed
CHLIEH

9h-10h

Réseaux de neurones récurrents appliqués à la détection de «hotword» pour des assistants vocaux

Robust recurrent neural models for hotword detection in voice assistants

Ces dernières années, la détection de hotword à l'aide de modèles contraints en temps de calcul et en mémoire a attiré l'attention de la recherche dans le domaine de la reconnaissance vocale. Les assistants vocaux comme Alexa par Amazon, Siri par Apple et Google Home contiennent un système de détection de hotword qui fonctionne en continu sur le produit même ("Alexa", "OK Google", "Hey Siri") et qui n'envoie les données de parole au cloud qu'après détection du hotword pour des raisons évidentes de respect de vie privée. Snips est un assistant vocal qui n'envoie jamais de données dans le cloud. Pour de telles applications, le système de détection de hotword doit être précis et contraint à fonctionner avec un faible CPU et une petite mémoire. Ces contraintes de calculs font des réseaux de neurones récurrents de mauvais candidats, cependant ils parviennent à capturer des tendances très intéressantes sur des fenêtres temporelles assez grandes. Comment alors repenser ces modèles dans ces contraintes de calculs et de mémoire?

Small-footprint hotword spotting systems have been increasingly attracting attention. Voice assistant systems such as Alexa on Amazon Echo, Siri from Apple or Google Home deploy an always-running keyword spotting system on device ("Alexa", "OK Google", "Hey Siri"), and only stream audio to the cloud for automated speech recognition after detecting a hotword instance for obvious privacy matters. Snips is a voice assistant running completely on device, and faces the same challenges. For such applications, accurate on-device hotword spotting running with low CPU and memory is critical. This low latency constraint makes RNNs or bidirectional LSTM not a proper fit in principle, however they manage to capture long-term dependencies in speech sequences. How can we re-think such models to make them work on device with low computational cost and memory footprint?

SNIPS
Paris



Romain
VIAL

10h-11h

Analyse du langage naturel appliqué au traitement automatisé de contrats juridiques

Natural language processing for mindless contract management

La gestion des contrats est une importante source de risque et d'optimisation des coûts pour les entreprises. Les solutions existantes sont lourdes et complexes (saisie de l'information fastidieuse, interfaces chargées) et ne permettent pas d'aider les juristes/avocats à se positionner comme partenaires business (possibilité de recherches d'infos limitées, absence d'automatisation des tâches à faible valeur ajoutée). Hyperlex développe une solution simple et intelligente de gestion de contrats pour les entreprises en ligne, afin de leur offrir les services juridiques dont elles ont besoin.

L'objectif d'ensemble de ce stage est d'explorer les différentes méthodes récentes d'analyse de texte afin de résoudre des tâches variées : classification automatique de documents, de clauses ou extraction d'entités nommées. De plus, la quantité de données accessibles pour entraîner des modèles étant variable selon les tâches et les clients, il sera proposé un ensemble varié de techniques permettant de résoudre au mieux les différents problèmes en fonction des données présentes (de non-supervisé à largement supervisé).

Contract management is an important source of risk and of cost optimization for companies. Existing solutions are complex and require tedious and repetitive data collection through a non-intuitive user interface. Hence, lawyers have limited search capabilities and cannot automate low value-added tasks. Hyperlex is thus creating a simple and intelligent on-line contract management solution for companies. This internship aims at exploring recent natural language processing and machine learning methods to solve various tasks such as document and clause classification or named entity recognition. Moreover, as the amount of data available to train models is varying according to the task and the clients, several techniques will be proposed to provide an efficient approach among all possible data regime (from an unsupervised to a fully supervised scheme).

Hiperlex
Paris



Gabriel
MELKI

11h-12h

Utilisation de l'Intelligence Artificielle pour combattre la fraude

Use of artificial intelligence to combat fraud

Le phénomène de fraude est un enjeu capital pour les assureurs français mais également internationaux, puisqu'on estime aujourd'hui qu'en moyenne 15% des déclarations de sinistres sont frauduleuses à travers le monde. Que ce soit de la fraude organisée par des particuliers, par des groupes structurés, par des entreprises ou de la fraude à l'assurance-vie (blanchissement d'argent), la gestion est longue, complexe et souvent coûteuse. C'est la raison pour laquelle les groupes d'assurance cherchent à se doter de solutions automatisées.

Shift Technology est née en 2013 pour répondre à ce besoin, en proposant une solution de détection des cas suspects basée sur des algorithmes de machine learning et dont l'objectif est de faciliter le travail des gestionnaires anti-fraude. La tâche des Data Scientists de Shift Technology consiste à extraire toute l'information utile à la détection de fraude à partir de données souvent incomplètes, redondantes, mal-orthographiées et peu structurées. La détection de fraude individuelle et collective soulève de nombreuses difficultés pratiques et théoriques, en particulier la reconstruction d'entités dans une base de données, l'estimation de montants dépensés, la détection de communautés, celle de cycles dans un graphe non orienté ou encore le scoring non-supervisé sur un ensemble de sinistres.

The phenomenon of fraud is a major issue for both French and international insurers, since today it is estimated that on average 15% of claims are fraudulent worldwide. Whether it's organized fraud by individuals, organized groups, businesses or life insurance fraud (money laundering), management of the fraud is long, complex and often expensive. That's why insurance groups are looking for automated solutions.

Shift Technology was born in 2013 to respond to this need by proposing a suspicious case detection solution based on machine learning algorithms and whose objective is to facilitate the work of anti-fraud managers. The task of Shift Technology's Data Scientists is to extract all the information needed for fraud detection from often incomplete, redundant, misspelled and unstructured data. The detection of individual and collective fraud raises many practical and theoretical difficulties, in particular the reconstruction of entities in a database, the estimation of amounts spent, the detection of communities, that of cycles in an undirected graph or even unsupervised scoring on a set of claims.

Shift Technology
Paris



Nadir
BOUCHENE

13h30-14h30

Data science pour le « revenue management » : segmentation de l'offre et prédiction de la demande

Data science for revenue management: offer segmentation and demand forecasting

Le revenue management (ou gestion de revenus en français) consiste en l'application d'outils analytiques pour prédire les comportements clients et en l'optimisation des stocks et de la demande dans le but de maximiser le chiffre d'affaire d'une entreprise. Pricemoov est une plateforme de gestion de revenus basée sur l'intelligence artificielle. L'entreprise collecte les données clients issues de l'historique des transactions ainsi que certaines données externes pouvant influencer sur l'achat. Ces données sont classées de manière à segmenter l'offre et les typologies clients. Ensuite, il s'agit de prédire la demande dans ces différents segments puis d'optimiser le prix en fonction de celle-ci et du stock de produits.

Mon stage s'est divisé autour de deux tâches. La première a été un travail de recherche sur l'amélioration du module de prédiction de la demande en utilisant des méthodes d'ensemble learning et de deep learning (LSTM). Une part importante a été dédiée au feature engineering à partir de l'analyse des séries temporelles. L'autre mission qu'on m'a confiée consistait à travailler sur des missions courtes réalisées pour des clients afin de livrer des analyses quantitatives sur leurs méthodes de pricing.

Revenue management is the application of analytics that predict consumer behavior and optimize product availability and price to maximize revenue growth. Pricemoov is a SaaS multi-sector dynamic pricing solution. It provides an AI-powered revenue management platform that allows key stakeholders to collaborate around their company's revenue strategy. From the beginning, it collects a great amount of data from transaction records and also external data. Then, it builds customer and offer segmentations and provides highly accurate prediction algorithms to enable companies to adjust their prices in real time.

My internship was divided into two parts in order to discover the large range of skills and services. Firstly, I have been in charge with developing a new module of forecast to provide more accurate demand prediction. I have focused also on feature engineering based on the analysis of time series. My research has targeted ensemble learning methods and recurrent neural networks to forecast heterogeneous time series. Moreover, I have been involved in short-term projects as proof of concepts that aims to give to customers an overview of the solution delivered by Pricemoov.

PRICEMOOV
Joinville-le-Pont



Joseph
REQUILLART

14h30-15h30

Optimisation du taux d'autorisation du paiement sur internet de e-commerçants

Online payment authorization rate optimization for e-merchants

Le paiement sur internet pour les e-commerçants est un sujet très complexe avec beaucoup d'argent en jeu. Le management de son flux de paiement est une tâche très difficile et beaucoup d'e-commerçants perdent beaucoup d'argent par défaut de compréhension. Améliorer la performance de paiement d'un e-commerçant implique à la fois de diminuer ses frais de transaction et d'augmenter son taux d'autorisation. Une transaction sur internet peut échouer pour plusieurs raisons incluant des raisons financières, techniques ou encore pour suspicion de fraude. Il s'avère que les banques du porteur de la transaction refusent trop souvent des transactions pour suspicion de fraude, ce qui entraîne un manque à gagner immédiat pour les e-commerçants. Une transaction est caractérisée par un

Online payment for e-merchants is a very complex workflow with a lot of money at stake. Managing this workflow is not an easy task and e-merchants often lose a lot of money. Improving payment performance deals with both fees and authorization rate. Online transactions can be refused for many reasons including financial or technical issues and fraud suspicion. Issuing banks refuse too many transactions for fraud suspicion, which leads to direct shortfall for e-merchants. A transaction is described by a lot of parameters (around 20) and changing one of them can lead to a very different result. By collecting payment information of thousands of e-merchants (with top French e-merchants among them), we can have a unique understanding of the payment workflow. Statistical analysis helps us to understand which

certain nombre de paramètres (une vingtaine) et la variation de l'un d'entre eux peut avoir un impact important sur l'issue de cette transaction. En collectant les données de paiement de plusieurs milliers de e-commerçants (dont les plus gros français), nous sommes en mesure d'analyser et de comprendre le fonctionnement du flux de paiement de manière unique. Par l'analyse statistique de nos données nous comprenons mieux les paramètres significatifs d'une transaction afin de modéliser une transaction par un jeu de variables pertinentes. La suite de notre approche est de construire des modèles de scoring à l'aide d'algorithmes de machine learning afin de mesurer la chance de succès d'une transaction et d'identifier la combinaison de paramètres qui optimise les chances de succès de cette transaction. Naturellement notre dataset est très déséquilibré du fait de la proportion de transactions réussies, ce qui nous oblige à recourir à différentes techniques pour palier à ce problème. Dans la mesure où nous ne pouvons pas rejouer une transaction qui a échoué dans le passé pour valider notre choix de combinaison de paramètres, il est très difficile d'évaluer nos modèles. De plus le passé récent du comportement du flux de paiement a un impact important à prendre en compte sur les transactions présentes, ce pourquoi nos modèles ont besoin d'être très réactifs au passé récent. Une catégorie d'algorithmes, les multi armed bandit, sont très adaptés à ces problématiques dans la mesure où ils s'adaptent très bien au comportement récent d'un des modèles qu'il a disposition et peut rapidement adapter son choix vers un autre modèle plus adapté à la situation.

parameters are meaningful and allows us to model a transaction by a set of relevant features. The next step of our approach is to use machine learning algorithms to build a scoring system for the chance of success of a transaction which will point out the best parameter combination to optimize the chance of success. Naturally our dataset is very imbalanced and we have to use several tricks to fight this issue. Since we cannot retry a past transaction to validate our parameter combination choice, it is hard to evaluate the performance of our models. Moreover, the recent past of transaction behavior has a huge impact on current behavior, thus our models need to be highly reactive to recent past. Multi-armed bandit algorithms are very adapted to those problematics since they will constantly choose the model that provides the best parameter combination for a given transaction profile, regarding previous performance of each model.

ProcessOut (Station F)
Paris



Mamadou
NGOM

15h30-16h30

Calibration d'un modèle macro-financier du rendement à long terme des actions

Calibration of a macro-financial model of long-term yield of bonds

EDF gère un fonds d'actifs ayant vocation à sécuriser le financement des charges de long terme du nucléaire, notamment le démantèlement des centrales et la gestion des déchets radioactifs. Un décret datant de 2007 précise que « le taux d'actualisation utilisé pour le calcul des provisions ne peut excéder le taux de rendement des actifs dédiés, tel qu'estimé avec un haut degré de confiance ». Ce contexte réglementaire nécessite de justifier les hypothèses prises à moyen / long terme (10 ans voire 20 ans) sur le rendement des actifs financiers en portefeuille dans un contexte de ralentissement de la croissance économique. La R&D d'EDF travaille pour la division ALM (Asset Liability Management) de la Direction Financière d'EDF sur la problématique du rendement à long terme des actifs financiers et sa dépendance aux fondamentaux macroéconomiques. L'objectif du stage est de s'appuyer sur des modèles macro-financiers de type Long Run Risk (LRR) pour expliquer les primes de risques sur le marché action ainsi que les taux d'intérêts à long terme via une modélisation

EDF manages an asset fund intended to secure the financing of long-term nuclear charges, in particular the dismantling of power stations and the management of radioactive waste. A 2007 decree states that «the discount rate used for the calculation of provisions cannot exceed the rate of return on dedicated assets, as estimated with a high degree of confidence». This regulatory context necessitates justifying the assumptions made in the medium / long term (10 years or even 20 years) on the return on financial assets in the portfolio against a backdrop of slowing economic growth. EDF R & D works for the ALM (Asset Liability Management) division of EDF's Financial Department on the issue of long-term return on financial assets and its dependence on macroeconomic fundamentals. The objective of the internship is to rely on Long Run Risk (LRR) macro-financial models to explain risk premiums on the equity market as well as long-term interest rates via econometric modelling including macroeconomic variables (typically the growth rate). Based

économétrique incluant des variables macro-économiques (typiquement, le taux de croissance). A partir d'une implémentation existante d'un modèle novateur de type LRR, un objectif majeur du stage est de mettre en œuvre une méthode de calibration sur données historiques permettant in fine d'interpréter les rendements passés des actions et les taux historiques, afin de produire des scénarios prospectifs plausibles sur le long terme conditionnés à des hypothèses macroéconomiques (évolution de la tendance du taux de croissance, évolution des taux). Le stage comporte d'abord une phase bibliographique permettant de se familiariser avec la littérature existante sur les modèles LRR et de comprendre le positionnement du modèle retenu par EDF puis une phase de d'estimation du modèle sur données historiques, possiblement par méthode des moments.

on an existing implementation of an innovative LRR-type model, a major objective of the internship is to implement a calibration method based on historical data, which ultimately allows the interpretation of historical equity returns and historical rates and to produce long-term, plausible prospective scenarios subject to macroeconomic assumptions (evolution of the growth rate trend, interest rate trends). The internship includes first a bibliographic phase to become familiar with the existing literature on LRR models and to understand the positioning of the model chosen by EDF and then a phase of estimating the model on historical data, possibly by the method of moments.

EDF R & D
Palaiseau



Yuanjui
WANG

16h30-17h30

Application du « Approximate String Matching » dans l'intégration d'informations bancaires

Application of Approximate String Matching in bank information integration

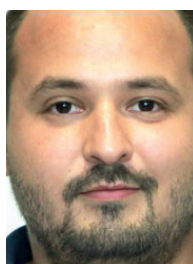
L'appariement des noms est important dans les banques où l'analyse et les rapports d'affaires sont invoqués sur des informations intégrées provenant de sources internes et externes. Les divergences dans de nombreuses bases de données opérationnelles locales et centrales, telles que les variations de noms de clients et d'adresses, rendent difficile l'intégration et la construction d'une grande base de données centrale. En outre, la grande échelle de la quantité de données exclut presque la possibilité d'effectuer l'intégration manuellement. Par conséquent, une approche automatique est en demande pour correspondre à l'information lancée à partir de diverses sources avec un rendement élevé et une précision acceptable.

L'objet du projet est d'appliquer des techniques d'Approximate String Matching à l'intégration des informations bancaires. La première étape consiste à rechercher dans la littérature et à identifier les algorithmes disponibles, tels que Levenshtein, sou-dex, appariement probabiliste, liés à la question de l'intégration de l'information bancaire. Ensuite, étudier le modèle de données à assortir, comparer les algorithmes de correspondance actuels, mesurer l'efficacité et la précision, et trouver le meilleur algorithme dans le contexte. Enfin, développer et affiner un outil pratique basé sur Alteryx ou python afin d'appliquer l'algorithme à l'intégration de la base de données.

Name matching is important in banks where business analysis and reporting rely on integrated information from both internal and external sources. Discrepancies in numerous local and central operational databases, such as variations of client names and addresses, make it difficult to integrate and construct a grand central database. In addition, the large scale of data amount almost excludes the possibility to conduct the integration manually. Hence, an automatic approach is in demand to match information initiated from various sources with high efficiency and acceptable precision.

The object of the work is to apply approximate string matching techniques to bank information integration. The first step is to research in literature and identify available algorithms, such as Levenshtein, Soudex, probabilistic matching, related to the question of bank information integration. Second, study the pattern of data to be matched, compare the current matching algorithms, measure the efficiency and accuracy, and find the best algorithm in the context. Finally, develop and fine tune a practical tool based on Alteryx or Python so as to apply the algorithm to the database integration.

BNP Paribas
Paris

Sami
JOUABER

9h-10h

Classification de séries temporelles appliquée aux trajectoires estimées d'aéronefs

Time series classification applied to estimated aircrafts trajectories

Les centres de contrôle aérien sont chargés de fournir une vision claire des aéronefs en vol. Cependant à ce jour la labélisation des appareils détectés par radar ne repose que sur la collaboration de ceux-ci et l'émission par le Automatic Dependence Surveillance Broadcast (ADS-B) des informations les concernant (modèle, propriétaire, etc.). Avec la multiplication des appareils privés, et notamment les drones, il devient nécessaire d'être capable d'identifier les aéronefs dont seule la trajectoire peut être suivie au radar.

Ce stage proposé par le Centre de Robotique (CAOR) et le Centre de Morphologie Mathématique (CMM) des Mines de Paris, en association avec Thales, a pour but de développer une solution de classification de trajectoires estimées d'aéronefs par apprentissage artificiel basée uniquement sur les données cinématiques des appareils observés.

The first goal of a traffic control center is to yield a clear vision of the state of the air traffic. However the information about detected aircrafts (model, owner, etc) is only provided by themselves thanks to Automatic Dependence Surveillance Broadcast (ADS-B). With the increased use of private aircrafts and especially drones, one should be able to automatically detect a number of features solely based on the radar measurements.

This internship proposed by the Centre de Robotique (CAOR) and the Centre de Morphologie Mathématique (CMM) of Mines ParisTech, in cooperation with Thales, aims to develop the tools to classify aircraft trajectories based on supervised machine learning using only cinematic information.

Centre de Morphologie Mathématique, MINES ParisTech
Fontainebleau

Pierre
GLASER

10h-11h

Réseaux de neurones et données malpropres

Neural Networks and dirty data

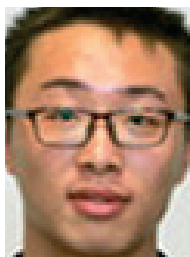
Le but de ce stage est de comprendre mieux la puissance des réseaux de neurones dans le calcul de similarité entre entités sémantiques (mots, phrases...). Les encodages classiques des vecteurs de mots de type one-hot génèrent des vecteurs orthogonaux qui ne tiennent pas compte des similarités entre mots. Pour pallier ce problème, des encodages denses tels "word2vec" ont été élaborés. Cependant, ces derniers ne résolvent pas le problème des bases de données "sales", où certains champs ont pu être rentrés à la main. (exemple: comment interpréter "Highway", "Highwayy", de la même manière?). Une première solution est de "nettoyer" manuellement la base de données, mais cela est souvent long et assez aléatoire. Une seconde solution est de déterminer des mesures de similarité entre les entités sémantiques.

Mon travail pendant ce stage d'option est d'explorer le pouvoir explicatif des réseaux de neurones à inférer eux mêmes les bonnes mesures de similarités, et le gain qu'ils apportent.

The goal of this internship is to get a better understanding of neural networks trying to learn non-linear similarities on dirty categorical variable (typically semantic entities). Classical categorical encodings like one-hot create orthogonal vectors that do not take textual similarity into account. Further improvements led to dense vector representations such as "word2vec", and more recently, "fasttext". However, those algorithms lose a lot of their power if being trained on limited dataset with dirty categories, generally manually filled by humans. A first solution would be to manually clean the data, but this can be time-intensive. A second solution would be to transform the categorical entities into count vectors and create a similarity measure in this space.

My work during this internship will be to explore the power of neural networks in learning those non linear similarities that can be potentially more problem-specific and flexible than classic similarities such as 3gram similarity

Parietal (Inria-CEA)
Gif-sur-Yvette



Jungfen
CHEN

Septembre

Estimer la probabilité du retard d'un vol par l'apprentissage automatique

Estimation of flight-delay probability by machine learning

Comme de plus en plus de passagers choisissent de voyager en avion, le retard du vol a un impact de plus en plus important sur les passagers. Si une compagnie aérienne offre aux passagers une assurance pour le retard de vol, ils ont tendance à être fidèle parce que l'assurance réduit leur perte. Sur la base des données historiques, l'assurance paramétrique présente de nombreux avantages par rapport aux produits actuariels traditionnels. Lorsqu'un vol est en retard de plus de 2 heures, les passagers reçoivent leur compensation immédiatement et automatiquement. Il apprend des données historiques pour calculer la probabilité d'un retard de vol, aide ainsi les assureurs à déterminer le montant de cette assurance.

Le principal défi de ce stage est d'abord de sélectionner des compagnies aériennes avec des données de bonne qualité, puis d'appliquer des algorithmes d'apprentissage automatique pour prédire la probabilité de retard des vols dans les compagnies aériennes.

As more and more passengers choose to travel by airplanes, the delay of flights has an increasingly important impact for passengers. If an airline company offers the passengers an insurance for flight delay, they tend to be loyal because the assurance reduces their loss. Based on historical data, parametric insurance has much advantages over traditional actuary products. Once a flight is late for more than 2 hours, the passengers get their compensation immediately and automatically. It learns from historical data to compute the probability of a flight delay, thus helps the insurers determine the price of the insurance. The main challenge of this internship is first to select airlines with good quality data, then to apply machine learning algorithms to predict the probability of delay for flights in certain airlines.

AXA Global Parametrics
Paris



Marc
CHETCUTI

Septembre

Utilisation de méthodes statistiques dans le marché de l'énergie solaire

Statistical Modelling Applied to the Solar Energy Market

L'effondrement des coûts de l'énergie solaire sur les cinq dernières années, ainsi qu'un nouvel engouement pour les énergies renouvelables suite à la COP21 est en train de restructurer fondamentalement le marché de l'énergie à l'échelle nationale ainsi qu'à l'international. Une très large pénétration des renouvelables, ainsi qu'un recul du charbon sont à prévoir dans les mix énergétiques des pays à travers le monde. Face à ces changements structurels de grande ampleur il est nécessaire pour les acteurs majeurs du solaire d'avoir à leur disposition des outils performants pour prévoir les quantités d'énergies qui pourront être générées par un projet solaire, ainsi que des outils précis permettant la valorisation financière de projets. Les grands acteurs de la division Renouvelables de Total, à savoir Total Solar, SAFT, Total EREN et Sunpower sont largement conscients de ces enjeux et recherchent des méthodes économiques et statistiques pour pouvoir avoir des estimations fiables et précises.

The recent collapse of solar energy costs in the past five years and a heightened awareness of the need for the energy transition after the Paris Agreement are fundamentally altering the structure of the energy markets both on a local and international scale. A large penetration of solar energy and a relative decline of coal are to be expected in the energy mix across the globe. With these major shifts in mind, it is now necessary for the large players in solar energy to have access to performant tools to predict the quantities of useful solar energy that could be generated by a project. Efficient models for project financial valuation will also be needed. Total Solar, a subsidiary of Total dedicated to solar energy project (Utility Power Plants and Industrial and Commercial sectors mostly), as well as SAFT, Total EREN and Sunpower - all part of Total Renewable Energy division - are well aware of this requirement and are actively seeking new financial and statistical models to have more reliable and precise data.

Total
Paris



Quentin
PETIT

Septembre

Jeux à champ moyen discret

Discreet mean field games

Les jeux à champ moyen sont une branche des mathématiques appliquées à la frontière entre analyse et probabilités. C'est une théorie qui a été introduite il y a une dizaine d'années par Pierre-Louis Lions et Jean-Michel Lasry. Son intérêt est qu'elle permet de modéliser une situation où il y a un grand nombre de joueurs indistinguables et leur influence sur le système est négligeable. On retrouve de telles situations particulièrement en économie mais aussi en biologie et en physique. Une des forces de cette approche est qu'elle résume l'évolution des décisions et l'évolution de la répartition des agents en un système d'équations aux dérivées partielles couplées. Il est alors naturel de vouloir résoudre numériquement ce système. Certains travaux ont déjà été faits en ce sens, on peut citer par exemple les articles d'Yves Achdou où il propose une discrétisation du système en partant des équations continues.

Mon travail durant ce stage est d'étudier un autre type de discrétisation, voir si cette discrétisation est consistante avec le système continu, comprendre le lien qu'il peut y avoir avec celle déjà établie et en dire ses avantages.

Mean field games are an applied mathematics branch linked to probability and analysis. It is a theory introduced about ten years ago by Pierre-Louis Lions and Jean-Michel Lasry. Its interest is that it permits to model a situation with a large number of players who are indistinguishable, and their action on the global system is negligible. This kind of situation can be found, in particular in economics but also in biology and physics. One strength of this approach is that it summarizes the evolution of the decisions and the evolution of the repartition of players in one system of coupled partial differential equations. It is natural to want to solve numerically its system. Several works have already been done in this way, we can quote some articles written by Yves Achdou who proposes a discretisation starting from the continuous system.

My work during my internship is to study another type of discretisation, to see if it is consistent with the continuous system, to understand the link with the other discretisation and to name its advantages.

CEREMADE, Université Paris-Dauphine
Paris



Wei
XIONG

Septembre

Détecteur et générateur de commentaires automatiques sur la cause des exceptions de Value-at-risk quotidiennes

Automated detection and commentary generator on the causes of daily value-at-risk

La Value-at-risk (VaR) est une mesure de risque importante appliquée par des instituts financiers et régulateurs pour d'une part quantifier le risque du marché des différents portefeuilles et d'autre part calculer la charge de capital exigée par les régulateurs qui couvre la perte possible. Elle est définie comme la somme de perte possible d'un certain niveau de probabilité ($\alpha = 5\%$ ou 1%), sur un horizon de temps (une journée par exemple). Intuitivement elle peut s'interpréter comme l' α -quantile de la distribution de perte. Au jour t , une $VaR(\alpha)$ est calculée pour estimer la perte maximale du jour $(t+1)$. Une exception a lieu si la perte du jour $(t+1)$ dépasse la $VaR(\alpha)$. Cette exception est probablement provoquée par le mouvement du marché, le problème du modèle, etc. En cas d'exception, la gestion de risque cherche sa cause en inspectant les portefeuilles et le modèle de VaR.

Ce projet de stage vise à développer un outil qui peut localiser automatiquement la cause d'exception basée sur des données historiques et la distribution de PnL. Un commentaire sera également généré entretemps. Cela exige l'application de différents modèles statistiques et techniques de Machine Learning.

Value-at-Risk (VaR) is an important risk measure now widely used by financial institutions and regulators to quantify market risk of various portfolios and to compute the capital charge required by regulators to cover possible losses. It is defined as the amount of possible loss given a probability level (for example $\alpha = 5\%$ or 1%), over a time period such as a day. It can be intuitively viewed as an α -quantile of the loss distribution. On day t , a $VaR(\alpha)$, which is a negative number, is calculated to estimate the possible loss of day $(t+1)$: the probability that the loss on day $(t+1)$ is no more than $VaR(\alpha)$ is $(1-\alpha)$. An exception happens when the actual profit and loss (PnL) on day $(t+1)$ is inferior to $VaR(\alpha)$. The possible causes of exception could be market move, model issues, etc. Every time when an exception happens, the risk management function in a firm needs to find out its causes by scrutinizing the portfolios as well as the VaR model, and then give a commentary on the losses.

This internship project aims to develop an automated tool based on historical distribution of PnL that can automatically localize the cause of daily VaR exception. A commentary is also generated with the detected cause. This requires the application of various statistical models and machine-learning techniques.

J.P. Morgan
Londres (Royaume-Uni)