

PHILÉAS CONDEMINE

En bref

Lead Data Scientist NLP chez Covéa depuis Octobre 2021.

4 ans au ministère de la Santé pour modéliser les **parcours de soins** du SNDS et mettre en place des outils pour le **centre de crise sanitaire CoViD-19**.

3,5 ans au centre d'excellence technique d'AXA AGPC : amélioration des techniques de **tarification** et de **gestion des sinistres** grâce au *machine learning* et aux données externes.

📄 Télécharger en PDF sur [github](#)



EXPÉRIENCES PROFESSIONNELLES

2021-
Aujourd'hui

Lead Data Scientist NLP


 Conseil Interne Data Science

📍 Paris, France

- Multiple projets **NLP en Production** avec des problématiques de **Labellisation**, **Annotation**, **Monitoring**, **MultiLabel Topic Classification**, **Sentiment Analysis**, **Anonymisation** grâce au fine-tuning et re-entraînement de modèles type RoBERTa/CamemBERT : **Transformers HuggingFace** sur **Azure + Databricks**.

2018-
2021

Senior Data Scientist et Entrepreneur d'Intérêt Général (EIG)

 DREES et centre crise sanitaire CoViD-19

📍 Paris, France

- **Modélisation de séquences de soins** pour la prédiction de pathologies et la détection de ruptures dans les séquences de soins.
- Classification par **active-learning** et **diffusion** d'indicateurs de santé.
- Outil pour les ARS : **saisie interactive** du zonage des professions de santé
- Contribution au **centre de crise sanitaire CoViD-19** : collecte de données auprès **des hôpitaux - respirateurs** et **des laboratoires - tests, réactifs, centre de dépistage, EPI**.
- Production d'*open-data* sur des données hospitalières avec implémentation de contraintes d'**anonymisation** k-anonymat et l-diversité
- Mise en place de **formations à R et Shiny** pour les agents de la DREES

2014-
2017

Actuarial Data Scientist

 AXA Global P&C

📍 Paris, France

- **Innovation dans la tarification** en assurance auto et mrh en améliorant les zoniers, véhiculiers. Tarification avec des techniques de *boosting* et d'interprétation du *machine-learning*.
- Analyse du **coût des sinistres** : prédiction du coût théorique pour *scorer* les garages en tenant compte du *case-mix*. Poc avec AXA-Spain, déployé en Espagne, Italie, France avec les *Data Engineers* locaux.
- Utilisation de données textuelles pour la **gestion des sinistres litigieux**.
- **Modèle de risque des routes** basé sur les données GPS des utilisateurs.
- Mise en place de hackathons et formations à la data science pour les actuaires.

2014

Mémoire d'actuariat

 AXA Belgium

📍 Télétravail

Exploitation des données de tarification du portefeuille auto d'AXA Belgium pour mesurer l'impact *ceteris paribus* de l'âge sur le risque d'accident de la route avec des techniques économétriques. **Sinistralité des conducteurs âgés** Le mémoire était finaliste du prix SCOR.

CONTACT

✉ phileas.condemine

🐙 github.com

🔑 gitlab.com

📈 hackerrank.com

🐦 twitter.com

in linkedin.com

📞 +33 643 549 576

COMPÉTENCES

Expert en R, avancé en Python et confirmé en SQL & NoSQL.

Machine-learning **GLM**, **XGBoost/GBM**, **SVM** pour la résolution de problèmes en assurance : tarification, gestion de sinistres, fraude, attrition.

Traitement de données massives en **Python-PySpark** mais aussi **R-data.table** quand ça suffit.

Deep-learning en **Python - Pytorch & Keras** pour la modélisation de séquences de soins ou le traitement du langage naturel.

Dashboard et outils interactifs avec **R-Shiny**, leaflet, plotly, DT, rhandsontable...

2012-2013

- **Stages de césure**
SCOR - Exane 📍 Paris, France
6 mois sur la tarification de Cat-Bonds par techniques MCMC 6 mois en structuration de produits dérivés et construction d'un indice synthétique basé sur un panier d'actions optimal



FORMATION

2017-2020

- **Deep learning**
Découverte, approfondissement et mise en pratique des techniques de *deep learning*.
📍 fast.ai, datascientest, deeplearning.ai

Formation au *deep learning* : vision par ordinateur et traitement du langage naturel. Entraînement de modèle *from-scratch* mais surtout *transfer-learning* / *fine-tuning*. Application aux séquences de soins : LSTM et 🧠 transformers.

2019

- **Spark & Scala**
Formation au langage Scala et à la programmation avec Spark & PySpark
📍 Coursera par Martin Odersky & Heather Miller

En parallèle de ce MOOC Coursera, je mettais en pratique sur le *High Performance Computer* de la DREES avec les données du SNDS.

2016

- **Formation aux langages web**
Introduction à HTML, CSS, Javascript & JQuery. 📍 CodeSchool.com

Formation avec projet sur la pile logiciel *Stack* MEAN. Développement d'une API de tarification fast-quote avec Express. Compétences Web cruciales pour le développement d'applications R-Shiny.

2014

- **Introduction à la data science**
Apprentissage supervisé et non supervisé 📍 Coursera par Bill Howe

Support Vector Machine, Gradient Boosting, Random Forests, k-means etc.

2010-2014

- **ENSAE Paris - IP Paris**
Cursus ingénieur 📍 Paris, France

Spécialisation actuariat et cours de **data science** et **big data** en parallèle



ENSEIGNEMENT

2018-2020

- **Formation au text-mining**
Techniques "classiques" de traitement du texte : tokenisation, lemmatisation, vectorisation, *embedding*.
📍 CEPE ENSAE/ENSAI

2018-2020

- **Formation au machine learning**
Techniques de classification pour des données structurées : *SVM, Gradient Boosting, Random Forests*
📍 CEPE ENSAE/ENSAI

2014-2017

- **Data Science for Actuaries DS4A**
Formation d'actuaire du groupe AXA aux techniques de data science dont [text mining](#) à travers des Bootcamp d'une semaine organisés en interne. Formation théorique avec la contribution d'Arthur Charpentier.
📍 DS4A AGPC

AUTRES PROJETS

Participation à des compétitions [Kaggle](#) : AXA "Telematics", Otto "Product Classification", Quora "Deduplication", West Nile Virus "Mosquito detection".

Création d'un robot 4-roues autonome programmé avec Arduino.

Participation à des hackathons AXA "Chatbot de fast-quote API.AI", [APHP](#) "Anomalies de prise en charge nocturne pour soins intensifs", ARS-IdF "Facteurs environnementaux d'augmentation des passages aux urgences".

Contribution à la [Buloterie](#) - un outil communautaire d'identification d'experts et curieux dans toutes sortes de domaines.

Ce CV a été généré grâce au package R [pagedown](#).

Mis à jour le 2022-02-28.