PHILEAS CONDEMINE

Lead Data Scientist

Covéa Data Science Team : NLP, ML/Deep-Learning in Production with Azure + DataBricks

3.5 years at Technical Excellence Center AXA Global P&C: Improving **pricing** techniques & using *machine learning* techniques for **claims handling**. Also getting & leveraging **external data** for better pricing, user-experience & claims management.

3.5 years at French Ministry of Health : Modelling **Health Sequences**, handling **Big Data** from the Public Health Insurance (covers 100% of 67M french citizens). Also tailoring tools to gather & visualize data for decision making during CoViD-19 crisis.

EXPERIENCE

2021-Now

Lead Data Scientist NLP

Data Science Internal Consulting Team — Paris, France

- Multiple **NLP** projects **in Production** Tackling Labelling, **Annotation**, **Monitoring**, **MultiLabel Topic Classification**, **Sentiment Analysis**, Anonymization using fine-tuned RoBERTa-like **Transformers with HuggingFace** with **Azure + Databricks**.

2018-2021

Senior Data Scientist & (EIG)

Statistical Departement at Ministry of Health — Paris, France

- **Modelling Health Sequences** to predict diseases outcome and detect disruption in the treatment course.
- Classification with **active-learning** & **General Public WebApp** to find Health-focused statistics.
- Tools for regional health agencies : **Interactive Decision Making Tool** to help experts elaborate Zoning for health professionals.
- Tech Lead at **CoViD-19 crisis center** : develop webapps to gather critical information from hospitals - ventilators and BioLabs - supplies, tests results, screening centers location & general info. Share data to stakeholders through dashboards & make advanced statistics from full hospital (SIVIC) & screening (SIDEP) data.
- Produce *open-data* on hospital admissions data (PMSI) involving **privacy constraints** k-**anonymity** & hierarchical l-diversity.

2014-2017

Actuarial Data Scientist

AXA Global P&C — Paris, France

- **P&C pricing innovation** for both housing & car insurance through zoning, vehicle classification, severity/frequency/propensity modelling with **gradient boosting** techniques transferred to GLM using **ML-interpretation** techniques.
- Build a **Claim Cost Analyzer** by predicting the theoretical cost of a claim - for a given vehicle & crash - to **score** a car repairer given their own case-mix. PoC with AXA-Spain, deployed in Spain then adapted to Italy & France with local **Data Engineering teams**.
- Leveraging French Court Decision Open-Data through **Natural Language Processing** to better handle bodily injury cases and assess contentious risk.
- Roads own-risk assessement based on GPS telematics data.

2014

Actuarial Thesis

AXA Belgium — Remote

Handle 1M contracts pricing-database to measure *ceteris paribus* impact of eldering on car crash severity & frequency using econometrics. Elderly drivers own-risk assessment Final selection for SCOR prize.

Download the PDF on Github

CONTACT

phileas.condemine
github.com
gitlab.com
hackerrank.com
twitter.com
linkedin.com
+33 643 549 576

SKILLS

R Expert, **Advanced knowledge of Python** and **Good knowledge of SQL & NoSQL - Spark in particular**.

Machine-learning **GLM, XGBoost/GBM, SVM** for insurance and general problem solving : pricing, claims handling, fraud, churn, attrition.

Handling Big Data with **Python+PySpark** but also **R-data.table** when it's enough.

Deep-learning for NLP & Health-sequences-modelling with **Python** : **Pytorch, Keras, HuggingFace Transformers**

Dashboard, annotation, cluster interpretation and other tailored tools with **R-Shiny**, leaflet, plotly, DT…

2012-2013

Long Internships
SCOR P&C, then Exane Derivatives. Paris, France
6 month pricing CAT-Bonds with MCMC techniques 6 month building a synthetic index as a dynamic basket of stocks & bonds

TRAINING

2017-2020

Deep learning
Advanced training & hands-on projects fast.ai, datascientest, deeplearning.ai

deep learning training : Computer Vision and Natural Language Understanding. Mainly use *transfer-learning* / *fine-tuning*. But also train models *from scratch* for Health Sequences Modelling using pseudo-NLP techniques: LSTM & 🤗 transformers.

2019

Spark & Scala
Scala Programming + Spark applications
Coursera by Martin Odersky & Heather Miller

While following this MOOC, I used pySpark on a daily basis on a *High Performance Computer* at work to handle National Health Claims Data (SNDS).

2016

Web Development
Introduction to HTML, CSS, Javascript & JQuery. CodeSchool.com

Training + application using *MEAN-Stack* : Mongo, Express, Angular & Node. Develop a fast-quote API for Housing Insurance with Express. This training has been very helpful to develop advanced Data-WebApps with R-Shiny.

2014

Introduction to data science
Main techniques of supervised & unsupervised learning.
Coursera by Bill Howe

Support Vector Machine, *Gradient Boosting*, *Random Forests*, *k-means* & hierarchical clustering.

2010-2014

ENSAE Paris - IP Paris
MSc Actuarial Science & Data Science Paris, France

Learning both **Data-Science** + **Big-Data** techniques & applications to **Actuarial Sciences**

TEACHING

2014-2020

Data Science Teacher

Data-Science for Actuaries & Data Science Certification
Paris, France

- **Natural Language Processing** & Text-mining techniques
- **Machine Learning** for structured data : Gradient Boosting & Support Vector Machine
- **Data Science Hands-On** with R, data.table, xgboost, glmnet & liblinear…
- **Build Interactive Apps** with R + Shiny
- **Data Science for Actuaries** (DS4A): 5-days training with hands-on & 1-day hackathon to teach AXA actuaries Data-Science Techniques that can help them better solve Insurance-related problems. The Theoretical Training was given by Arthur Charpentier.

OTHER PROJECTS

I love to hack new data and therefore participated in many hackathons & Kaggle competitions

Kaggle : **AXA** “Telematics”, **Otto** “Product Classification”, **Quora** “Deduplication”, West Nile Virus “Mosquito detection”.

Hackathons : **AXA** “Chatbot for fast-quote with API.AI”, **APHP** “Night-time-Treatment Outliers Detection in Intensive Care”, **ARS-IdF** “Environmental Factors of ER admissions”.

Programming a 4-wheel autonomous mini-car with Arduino.

Contribution to Open Project Bulloterie - a Low-Tech Tool that helps link potential teachers & learners in a given community.

*This resume was built with 💞 using R **pagedown**.*

Last update 2022-02-02.