Elie Azeraf

(+33)6 12 90 31 88
azerafelie@gmail.com
Websites:

Ph.D. in Statistics and Machine Learning

Competent and motivated data scientist, I am a last-year Ph.D. student at Telecom SudParis and IBM France, holder of two Master of Science degrees in Statistics and Machine Learning. My research project are always focused on business and real-world applications, conscious about production issues. I am looking for a **data scientist position** or an **applied scientist position** from May 2022. I am open to relocation.

Education

Ph.D. in Applied Mathematics, Statistics, and Machine Learning — Paris
Telecom SudParis, Institut Polytechnique de Paris / Paris-Saclay University — 2019 –Current

- Main advisors: Wojciech Pieczynski and Emmanuel Monfrini
- Title: *Generative Probabilistic Models: Discriminative classifiers and Neural Networks*
- Thesis at Hadamard Doctoral School of Mathematics, common to Paris-Saclay University (ranked #1 in Mathematics in the Shanghai Ranking) and Institut Polytechnique de Paris (ranked #31), at the SAMOVAR lab of Telecom SudParis
- Funding: 3 years grant from IBM.

M.Sc. in Data Science and Machine Learning — Paris
ENSAE Paris — 2016 –2018

- ENSAE Paris is considered as one of the top 3 engineering school in Statistics in France

M.Sc. in Statistics — Paris
Telecom SudParis — 2014 –2018

Preparatory School MPSI-PSI* — Marseille
Lycee Thiers — 2012 –2014

Experience

Ph.D. Student — Paris
IBM and Telecom SudParis — April 2019 - Current

- About the ability to any generative probabilistic model to define a discriminative classifier, parametrized with neural networks, with application to Natural Language Processing. These thesis allows to develop neural models based on Probabilistic Graphical Models, combining great modeling potential, performances, and easy to serve.
- Some research projects:
  - Developing extensions of the classifier induced from the Naive Bayes for Text Classification and Sentiment Analysis, dividing the error by 4.5 while keeping the same complexity.
  - Developing new algorithms for Hidden Markov Chain, allowing to consider complex features and achieving relevant results for Part-Of-Speech Tagging, Named-Entity-Recognition, state-of-the-art results for Chunking, and being easy to serve.
  - Showing the equivalence between the linear-chain Conditional Random Field and the Hidden Markov Chain.
- **Tech: Python, PyTorch, Transfomers, Flair**

Data Scientist Consultant — Paris
IBM France — November 2018 - April 2019

- Data scientist developing a long-tail chatbot using state-of-the-art Question Answering models (BERT) in French coupled with a document retriever module (based on Elastic Search)

- Other projects: clustering of unsupervised data, mail classification, data analysis
- **Tech: Python, Tensorflow, Keras, Jupyter, Flask, Scikit-learn, GoogleCP, ElasticSearch**

Data Scientist Intern Paris

IBM France May 2018 - November 2020

- Development of a database enrichment model lying on user questions, based on the sequential neural model BiLSTM, then BERT, query analysis, and word embedding methods GloVe and FastText.

- **Others:** research intern during summer 2016 and summer 2017 at Telecom SudParis - Paris, and developer intern during summer 2015 at Bo Digitize Ltd - Tel Aviv

Publications

1. **E. Azeraf**, E. Monfrini, and W. Pieczynski, "Using the Naive Bayes as a discriminative model," *13th International Conference on Machine Learning and Computing (ICMLC 2021)*, Association for Computing Machinery, 2021, pp. 106110
2. **E. Azeraf**, E. Monfrini, E. Vignon, and W. Pieczynski, "Highly Fast Text Segmentation With Pairwise Markov Chains," *2020 6th IEEE Congress on Information Science and Technology (CiSt)*, IEEE, 2020, pp. 361-366
3. **E. Azeraf**, E. Monfrini, E. Vignon, and W. Pieczynski, "Introducing the Hidden Neural Markov Chain Framework," *13th International Conference on Agents and Artificial Intelligence (ICAART)*, 2021, pp. 1013-1020

Submitted

1. **E. Azeraf**, E. Monfrini, E. Vignon, and W. Pieczynski, "Hidden Markov Chains, Entropic Forward-Backward, and Part-Of-Speech Tagging.", submitted at AISTATS 2022
2. **E. Azeraf**, E. Monfrini, and W. Pieczynski, "Deriving discriminative classifiers from generative models," submitted at Transactions for Pattern Analysis and Machine Intelligence
3. **E. Azeraf**, E. Monfrini, and W. Pieczynski, "Improving usual Naive Bayes classifier performances with Neural Naive Bayes based models," submitted at ICPRAM 2022
4. **E. Azeraf**, E. Monfrini, and W. Pieczynski, "On equivalence between linear-chain conditional random fields and hidden Markov chains," submitted at ICAART 2022

Awards

- Best Ph.D. student award of Telecom SudParis, winning an iPhone 11 2020
- Best internship award of ENSAE Paris, winning a reward of 1000€ 2018

Miscellaneous

- **Languages:** English (fluent), French (native)
- **Teaching:** Head Teacher of *Probabilistic Models and Machine Larning*, M2 course at Telecom Paris, and Assistant Teacher of many courses at ENSAE Paris, ENSTA Paris, and Telecom SudParis (L3, M1 and M2 levels), always highly appreciated by students. All these schools are among the top engineering schools in France.
- **Comics:** Passionate about comics, especially DC and Marvel; *The Dark Knight Returns* by F. Miller, *Spider-Man Blue* by J. Loeb and T. Sale, and *Justice League: Forever Evil* are among my favorite works.
- **Cloud certifications:** Preparing some Azure cloud for Machine Learning certifications
- **References:**
  - Emmannuel Monfrini, Full Professor at Telecom SudParis, emmanuel.monfrini@telecom-sudparis.eu
  - Emmanuel Vignon, ex-Watson Practice Leader at IBM France, CEO of Sicara, emmanuelv@sicara.com
  - Vincent Giraud, Senior Managing Consultant at IBM France, vincent.giraud2@ibm.com