Emma Scharfmann

https://emmascharfmann.github.io/ linkedin.com/in/emma-scharfmann/
emma_scharfmann@berkeley.edu 510-984-8337

PROFILE

I'm working as a Data Scientist at UC Berkeley on research projects led by the PIs Lee Fleming and Matt Marx. I'm applying machine learning techniques to research projects at the intersection of science (publications), technology (patents) and business (companies). I hold two master's degrees: one in Operations Research from UC Berkeley, and one in Artificial Intelligence from CentraleSupelec. Driven by my interest in optimization, machine learning and innovation, and my aim is to help researchers, companies and governments to have a better understanding of innovation.

EXPERIENCE

- Data Scientist - UC Berkeley, Haas Business School — August 2022 – Expected July 2024
  - Develop machine learning models and use NLP to create a database which links the worldwide scientific publications (+200 millions) to the US patents (+8 millions) written by the same individuals, in collaboration with Lee Fleming.
  - Create a database of patent-paper-pairs (hosted on Zenodo: https://zenodo.org/records/7996195) using machine learning and NLP, which provides pairs of papers and patents written by the same individuals describing the same invention, in collaboration with Matt Marx.
  - Supervise Berkeley master students (5 groups of 4 students) on machine learning projects (prediction of time series, computer vision applied to patents images, disambiguation of the worldwide patents database using NN with contrastive loss).
  - Develop of a platform hosted on hugging face (https://emmascharfmannberkeley-synapse-project.hf.space/) to facilitate and optimize the research of scientists, inventors, scientific papers, patents and companies.
- Machine Learning and Optimisation research project — August 2022 – November 2022
  - Research project: improving the loss function of Machine Learning models when solving repetitive linear optimisation problems.
- Project with HealthTech startup Organos — Sept 2021 – Expected May 2022
  - Using convolutional neural networks to detect effective treatments within cardiac drugs tested on Organos' testbed.
  - Developed intellectual property-based business models that integrate technology, business, and law approaches and recommended a commercialization strategy for Organos.
- University of California, Berkeley, CA — Jan 2022 – Expected May 2022
  Teaching assistant in Mathematics
  - Taught discussion sections and office hours for 50 undergraduate students (20 hours per week)
- Lycée Lavoisier and Lycée Henri IV, Paris, FR — Sept 2019 – Jun 2021
  Teaching assistant in Mathematics
  - Advised and questioned students enrolled in the intensive mathematics and physics programs.

EDUCATION

- University of California, Berkeley — August 2022 - May 2022
  Master of Engineering, Industrial Engineering & Operations Research
  - Technical classes: Data Analysis, Optimization (linear and non-linear), Applied Stochastic Process, Mathematical programming.
  - Business classes: Intellectual Property and Innovation: Analysis, Strategy and Management; Commercializing Science and Technology Breakthroughs.
  - Concentration: Intellectual Property and Entrepreneurship Strategy
  - GPA: 3.91

- Centrale Paris, Paris-Saclay University — Sept 2021 - June 2022
  Master of Engineering, Artificial Intelligence
  - Top 3 engineering schools in France
  - Technical classes: Statistics and Machine Learning; General Topology; Optimization.
  - GPA: 3.99
- Centrale Paris, Paris-Saclay University — Sept 2019 - June 2021
  Bachelor of Engineering
  - Top 3 engineering schools in France
  - Technical classes: Convergence, Integration, Probability; Partial Differential Equations; Algorithmic.
  - Business classes: Business Management; Corporate Accounting and Finance.
  - GPA: 3.99
- Lycée Henri IV, Paris — Sept 2017 – Jun 2019
  "Classe préparatoire", Intensive mathematics and physics program
  - A two-year post-secondary intensive curriculum in mathematics and physics leading to nationwide competitive entrance examinations to the leading institutions for scientific studies.
  - GPA: 3.99

ADDITIONAL

- Programming languages: Python, SQL, stata
- Strong programming experience in the python data science stack: pandas, scikit-learn, Jupyter, tensorflow, numpy, opencv, scipy, matplotlib, ...
- Experience with MySQL, PostgreSQL, GitHub, A/B testing, causal inference
- Relevant skills in building Machine Learning based solutions, using supervised (regression, classification, random forest, decision trees, boosting) and unsupervised (clustering, K-means) techniques.
- Relevant skills in Natural Language Processing.
- Deep Learning skills: NN and CNN models using Tensorflow and Keras.
- Spoken Languages: Native in French, proficient in English.