

CONTACT

Phone: +54 1131603769

Email: 96agustin.lehmann@gmail.com

Linkedln: <https://linkedin.com/in/agustin-lehmann-abaab41b1/>

Location: Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina

Portfolio: <https://lemonpi3.github.io/portfolio/>

PROFESSIONAL
EXPERIENCE

Product Owner/ Data Scientist

J.P Morgan | Analyst | 2024–Present

- Researched , evaluated and deployed a Search Engine for Legal Knowledge Management using Hybrid Search which also works as a RAG engine which allowed Google-like AI overviews for some of the queries, being a significant improvement from the previous solution.
- Manage to sort and categorize Legal documents using embeddings and clustering algorithms which allowed us to understand what content we had and what we were missing in the application. This also included a lot of analysis and experimentation using NLP as well as Node Graphs for some of the experiments.
- Helped adapting, using vector databases, a document-duplicate detection algorithm which was incompatible with our application stack and budget which used LSH-Minhash as is core at a cost of slightly more latency.
- Automated and improved Reports using different technologies such as matplotlib, plotly, seaborn in python as well as ggplot2 in R.
- Coordenated and assited MLCOE and Tech Leads in the deployment of AI/Data Science Solutions

Product Owner/ Data Scientist

J.P Morgan | Internship | 2023–2024

- Made a data extraction tool that got deployed internally for Adobe Analytics that improved the reporting process speed drastically and allowed the use of that data on other projects.
- Found Power Users and determined several user personas using formulas and clustering methods like DBSCAN and KMeans, that help the team understand the app users.
- Created a POC for a Hybrid Search system made with Semantic Search and BM25 Keyword Search to improve the Search in the main app. This also included a text extraction process that extracted and normalized text from different file types. Which is now being implemented with the firm technologies.
- Made a POC for a recommendation algorithm using collaborative filtering using usage metrics as well as a formula-based recommendation for users that were new on the app to display the potencial most relevant content.
- Helped cleaning Adobe Analytics from deprecated/broken/duplicated data that was making noise and causing confusion on its users.
- Created Several data pipelines for automating reports freeing time on the team to focus on more complex tasks.

SKILLS &
TECHNOLOGIES

- | | | |
|------------------------------|-----------------------------|--------------|
| • Python | • text extraction libraries | • streamlit |
| • NLTK | (pymupdf, python-pptx, etc) | • numpy |
| • Transformers (huggingface) | • NetworkX | • pandas |
| • tensorflow | • Requests | • NLP |
| • spacy | • Beautiful Soup | • matplotlib |
| • pytorch | • Web Scraping | • seaborn |
| • sklearn | • Adobe Analytics | • plotly |
| • Fast API | • Excel | • ggplot2 |
| • Pydantic | • C# | • Javascript |
| • Langchain | • Java | • React |
| • Flask | • Git Hub | |

EDUCATION

Data–Science | 2022–Present

University of Buenos Aires

Chemist | 2008–2015

Escuela Técnica N°27 Hipólito Yrigoyen, Buenos Aires, Argentina