JOSHUA BONACORSI

DATA SCIENTIST SPECIALIZING IN GENERATIVE AI | EAGER TO EXPLORE OPPORTUNITIES IN USA

France & Luxembourg (ready to relocate) | +33 6 51 77 56 01 | bonacorsijpro@gmail.com | LinkedIn | github

EDUCATION

Master of Science in Engineering, Major in Applied Maths & Data Science 2018 - 2023
CY Cergy Paris University, France

- Relevant coursework : Machine Learning, Advanced Mathematics (Calculus, Statistics, Probability…), Computer Science (Parallel Computing, Graph Theory…)
- GPA : 15.05/20, magna cum laude (in French scale)

WORK EXPERIENCE

Data Scientist, Generative AI Apr, 2024 - Now
Canal+ Group, Kirchberg, Luxembourg 7 months

- Leading the development of a suite of applications leveraging **LLMs** and **Generative AI,** including a chatbot, a recommendation system, and a search engine. Implemented a powerful **Retrieval-Augmented Generation (RAG)** system, significantly enhancing the quality of the user experience across platforms.
- Developing an innovative platform for market and competitor analysis within the media industry using **Streamlit** on **AWS/Zerve**. Established a comprehensive **data ingestion pipeline** to streamline data collection and analysis, enhancing strategic insights and decision-making capabilities.
- Spearheading an analytical project to evaluate the ROI of content, employing **predictive models (XGBoost…)** and data-driven insights to forecast performance. This initiative aimed to provide actionable recommendations for optimizing budget allocations and enhancing decision-making processes in content strategy.

Data Scientist Trainee, Generative AI Oct, 2024 - Mar, 2024
Canal+ Group, Kirchberg, Luxembourg 6 months

- Spearheaded the coding and deployment of a Marketing Automation system that utilizes **Generative AI and LLM** technologies to enhance content recommendation across email campaigns. This innovative approach led to a marked increase in customer engagement with an important raise of viewing time, directly impacting business outcomes.
- Developed and automated an end-to-end **data workflow** using **Zerve/AWS cloud services**, significantly improving the efficiency and reliability of data storage and retrieval processes. This initiative streamlined operations and supported scalable growth.
- Employed **Snowflake** and **SQL** extensively for data wrangling and management tasks, ensuring robust data integrity and optimized accessibility. This role involved transforming raw data into actionable insights, crucial for strategic decision-making.

Data Scientist, NLP Sep, 2022 - Sep, 2023
Groupe Covéa, Paris, France 1 year

- Leveraged **Azure Databricks** to design production-ready Data Science products, set up computing clusters, and deployed interactive notebooks
- Performed **NLP** tasks (Sentiment Analysis, Topic Modeling & NER) with **PyTorch** and **Huggingface**, including data preprocessing with **pySpark** & **Pandas** on large sized dataset (>1M of rows) and incorporating **LLM technologies (BERT, GPT)**.
- Developed a data analysis web-app using **Dash** and **Plotly**, created interactive interfaces, integrated visualizations, and deployed on a web server
- Industrialized ML models, including **code refactorization**, adopted **best development practices**, implemented **continuous integration pipelines**, and trained teams on developed models and processes

TECHNICALS SKILLS

Languages

- **Programming :** Python (mostly), R, SQL
- **Spoken :** French (native), English (fluent)

Technologies & Platforms

- **Cloud & Big Data** : AWS, Microsoft Azure Databricks, Apache Spark… currently learning Google Cloud Platform
- **Databases :** MySQL, MongoDB, MariaDB

Development Tools

- **Python Libraries :** Numpy, Pandas, PySpark, Scikit-Learn, TensorFlow, PyTorch, BeautifulSoup, OpenCV, HuggingFace Tools
- **Data Visualization :** Streamlit, Dash, Plotly, Matplotlib, Seaborn, RShiny
- **Version Control :** Git, Gitlab

SIDE PROJECTS

Horse Injury & Health Monitor, In-going Project 2024 - now
Me & friends studying in SupOptique Graduate School, Paris-Saclay, France

This project develops a predictive system using computer vision and machine learning to assess the risk of injuries in horses.

- Adapting a proven **computer vision** methodology, originally developed for human bone fracture detection, to equine needs using transfer learning techniques. This approach not only maximizes the utility of existing technologies but also enhances predictive accuracy for equine injuries.
- Applying **XGBoost** and explainable machine learning techniques to detect underperformance and analyze health patterns in horses. This integrated approach not only identifies early signs of potential injuries but also provides clear, actionable insights that enable veterinarians and trainers to understand and address issues effectively, enhancing overall equine care.