MARIE ELYSE BASSIL

Pavia, Italy | +39 3334207210 | mbassil97@gmail.com
https://github.com/Mariee03

Education

09/2025 – Present M.Sc. in **Artificial Intelligence**

- Joint program: University of Pavia, University of Milan (Statale), and University of Milan (Bicocca)

10/2021 – 02/2025 B.A. in **Artificial Intelligence**

- Joint program: University of Pavia, University of Milan (Statale), and University of Milan (Bicocca)
- **Thesis**: 'Developing an automated dashboard for syndromic surveillance: A data-driven approach to public health threat detection'

Work Experience

Jan 2024 – Oct 2024 **Data & AI Intern** – **Accenture**

- Managed data pipelines, models, and computing engines to assess business and operational risks.
- Collaborated with cross-functional teams to implement predictive analytics solutions.
- Developed dashboards and data visualizations to present KPIs and trends.
- Built data pipelines and managed cloud-based datasets using Python, SQL, GCP, Docker.

Dec 2020 – Present **Founder - *Fiber X, Skiniti - Remote***

- Launched and run three e-commerce brands in fashion and wellness.
- Run multi-channel paid acquisition (Meta, TikTok) and customer segmentation, applying the same analytical methods used in my academic work.
- Designed 20+ Shopify storefronts with custom UX/UI and optimised checkout flows, as a freelancer and for own brands.

Jun 2021 – Aug 2021 ***IT*** **Intern** - **Notre-Dame University technology center**

- Supported system maintenance, hardware troubleshooting, and software updates; documented bugs and irregular system behaviour.

Certifications

- **IBM Data Science Professional Certificate** - Coursera (12-course specialisation with capstone)
- **Advanced Algorithms & Data Structures (Python)** - Codecademy

Projects **Additional projects available on my GitHub**

City Mobility Platform – Multi-Paradigm Data Management (GitHub Link)
Python, PostgreSQL, MongoDB, Neo4j, PySpark, GraphFrames, Docker

- Designed and benchmarked a city mobility system across relational, document, and graph database paradigms over 36 dataset combinations (up to 50k users and 100k trips).
- Implemented and compared query performance across PostgreSQL, MongoDB, and Neo4j, identifying trade-offs between data models for different access patterns.
- Applied distributed graph analytics using Spark GraphFrames (PageRank, Connected Components) and analysed partitioning and replication strategies for scalability.

Pediatric Chest X-Ray Pneumonia Classification – Deep Learning & Explainable AI (GitHub Link)
PyTorch, ResNet18, Albumentations, Grad-CAM, Scikit-learn, Streamlit

- Trained and compared a from-scratch CNN against an ImageNet-pretrained ResNet18 on ~5,800 paediatric chest X-rays, addressing ~1:3 class imbalance with inverse-frequency weighting and selecting the best checkpoint using validation ROC-AUC.
- Evaluated performance and analysed the decision-threshold trade-off between missed pneumonia cases and false alarms.
- Implemented Grad-CAM from scratch using forward and backward hooks, and used it to inspect correct predictions and confident errors, revealing possible reliance on peripheral or non-lung image regions in some misclassifications.

Reddit Sentiment Analysis of ChatGPT – NLP & Social Media Analytics (Github Link)

Python, HuggingFace Transformers (RoBERTa), VADER, BeautifulSoup, NetworkX, Pandas

- Scraped Reddit threads to evaluate public sentiment about ChatGPT.
- Performed community detection and applied both rule-based (VADER) and transformer-based (RoBERTa) sentiment analysis.
- Provided key user insights for product feedback and perception monitoring.

Clinical Trials Search Engine – NLP & Information Retrieval (Github Link)

Python, TF-IDF, Cosine Similarity

- Developed a text-based retrieval system to match patients to clinical trials.
- Processed free-text summaries using NLP to return ranked, relevant trials.
- Reduced a manual search task to an automated ranked query.

Breast Cancer Classification – ML for Healthcare (Github Link)

Python, Scikit-learn, Pandas, Jupyter

- Built a supervised learning model to classify breast cancer cases using clinical data.
- Applied feature selection and hyperparameter tuning (GridSearchCV) to optimize performance.
- Achieved 96% accuracy on held-out test data.

Causal Inference in E-commerce - Discount Impact (Github Link)

Python, DoWhy, EconML, Pandas, Seaborn

- Used causal inference to measure how discounts influence conversion rates.
- Identified customer segments with the highest treatment effect.
- Findings informed smarter targeting strategies for sales campaigns.

Technical Skills

- Languages: Python, SQL, R, MATLAB, HTML/CSS
- ML & Data: Scikit-learn, XGBoost, LightGBM, Pandas, NumPy, StatsModels, SHAP, LIME
- Deep Learning & Computer Vision: PyTorch, TensorFlow, Keras, Albumentations
- NLP: HuggingFace Transformers, spaCy, NLTK, VADER
- Databases: PostgreSQL, MongoDB, Neo4j, SQLite, BigQuery, PySpark, GraphFrames
- Cloud & DevOps: Google Cloud Platform, AWS, Docker, Git
- Visualisation: Tableau, Streamlit, Matplotlib, Seaborn

Languages

Arabic (native) · Italian (C1) · English (C1, IELTS 7.5) · French (B2, DELF) · German(A1 beginner)