



Giuseppe Cannizzaro

✉ giucannizzaro@gmail.com

Machine Learning Engineer
Data Scientist

☎ +39 392 327 89 69

«If the design takes too long to bold, it's a bad design»

Programming languages:

- Python
- PySpark
- Java
- C
- C++
- C#
- JavaScript

Other languages:

- SQL
- HTML
- XML/XSD/DTD
- CSS

Experienced in

- AWS
- GCP
- Tableau
- Tensorflow
- Docker
- Pytorch
- Serverless framework
- Git
- GCP
- NodeJS
- Unity3d

Spoken languages:

- Italian: Mother tongue
- English: Proficient
- Spanish: Basic

Nerd of:

- Data
- Astronomy
- Science
- Rockets
- Artificial Intelligence
- Neural Networks
- Music

Education

Bachelor's degree - Computer Science and Engineering (Politecnico of Milan)

Master's degree - Computer Science and Engineering (Politecnico of Milan)

Experience / Projects

2024-Present: Freelance IT Consultant / Teacher:

- Data Scientist for Datalake implementation project at Tannico
- Developing websites for small businesses
- IT Leader for Humaverse: overseeing the entire IT department and developing AI tools to increase productivity of the company
- Teacher at Tech7 E.T.S.: teaching web development using HTML, CSS, JS and Express as a backend to young students with the aim of making them able to independently develop a website for real clients

July 2021 – December 2023: Data Scientist at Tannico (wine e-commerce):

- Overseeing the entire data infrastructure, I took charge of extracting transforming and orchestrating ETL processes to ensure the availability of data for various services, including recommender systems and Tableau dashboards.
- Developed API endpoints for the recommender systems from scratch, allowing for seamless integration with front-end applications.
- Designed and implemented a comprehensive data architecture, serving as the backbone of the company's data ecosystem. This involved leveraging AWS services such as Lambda, Glue, Step Functions, DynamoDB, and S3, in addition to utilizing Docker, VSCode, and GitHub for development.
- Successfully addressed infrastructural and internal process challenges by implementing a robust data lake infrastructure and creating detailed documentation that captures essential company information.
- Achieved a significant improvement in click-through rates (CTR) by developing a custom item-based recommender system using collaborative filtering techniques, resulting in a doubling of the CTR.
- Ensured high performance and reliability by leveraging memory-based recommendations stored on DynamoDB, enabling low-latency responses to user requests.
- Spearheaded reporting activities by developing and maintaining Tableau dashboards, utilizing Tableau Prep and SQL, alongside custom ETL processes that leverage the data lake infrastructure.

May 2020 – July 2021: Data Scientist at Business Integration Partners: I worked for small and large clients in the field of Machine Learning, Software Engineering, Data Analytics and Reporting. I worked on:

- **Traffic flow forecasting (GCP)** for an Italian toll booth operator. I developed an embedded solution (Neural Network + Statistical Models) to predict traffic with a total RMSE of 34 vehicles on an average flow of >400k vehicles. I also developed a dashboard in DataStudio to visualize results
- **Text Entity Extraction (GCP)** from pdf product labels for an Italian labelling company. I extracted all the required information such as «ingredients» or «Cross contamination» using an embedding of GCP-AutoML, GoogleOCR and Regular Expressions. Results reached an accuracy above 90%.
- **Datalake migration (GCP)** for International media company located in UK. Configured ETL for data migration and Planned and realized the enrichment of the DataLake to fit the requirements of an internal tool

2020 – Paper publication (ECML-PKDD) – Automated integration of Genomic Metadata with Sequence-to-Sequence Models: Result of the Master's degree project, I was asked to propose and implement a solution to the problem of integration of information between the NCBI's GEO repository of biological metadata and Polimi GeCo lab database. The major challenge was dealing with unstructured and noisy input text data which prevents from success of the «standard» Information Retrieval techniques (such as Regular expressions or Classification). I proposed a new way to extract information from GEO text using neural net seq2seq models (GPT2). I set up 3 experiments to prove the efficacy of the S2S models against a classifier, deducing that they overcome the SoA with an accuracy around 95% (showing suprising ability in deduction and handling of noisy text input) and do not suffer of other approaches' problems resulting in also easier to train and test. [[paper](#)]