

Enrico Foglia

MLOPS - BACKEND DEVELOPMENT - DATA ENGINEERING

Passionate about machine learning, data and programming. Experienced in ML modelling, MLOps, development of APIs and containerized microservice architectures in the Cloud. Enthusiastic about learning new tools, programming languages and methodologies.

Reach me at:



Saronno, VA



+39 3534035024



enrico.foglia@hotmail.com



linkedin: /efoglia



github.com/Enrico4444/



www.researchgate.net/profile/Enrico_Foglia



Education

University of Salford, UK, 2018-2019

Master Degree (MSc) Data Science
Distinction

University of Westminster, UK, 2012-2016

Bachelor Degree (BSc), Psychology
First Class Honours



Certifications

- AWS Machine Learning Engineer (in progress)
- Azure Cloud Data Fundamentals (DP 900)
- Azure Cloud Data Scientist (DP 100)
- SAS Student Academy Certification



Selected Courses

- Analisi I - Politecnico di Milano
- Algebra Lineare - Politecnico di Milano
- Principi di Informatica - Politecnico di Milano
- REST APIs with Flask & Python - Udemy
- Machine Learning Specialization - Coursera
- Sequence Models - Coursera



Work Experience

NTTData Italia, Milano, Italy

Data Scientist (consultant) - 01/2021 - currently

Worked with clients to deliver Machine Learning, MLOps and Data Engineering solutions, with strong focus on Cloud and Microservice architectures. Also responsible for the Partnership and Certification programs within the company

LOGIVAN, Ho Chi Minh City, Vietnam

Data Scientist - 06/2019 - 01/2021

Modelling and deployment of two ML models, a price prediction model and a recommender system, that critically impacted the activity of this logistics startup in terms of increased automation and reduced costs. Worked in a fast-paced startup environment.

Opus Energy, London, UK

Data Scientist (intern) - 02/2019 - 05/2019

Developed an energy consumption forecasting POC

University of East London, London, UK

Research Analyst - 06/2016 - 09/2016

Written a published systematic review on dehydration in the elderly

University of Westminster, London, UK

Research Analyst - 01/2016 - 05/2016

Conducted a cognitive study on Theory of Mind in the adolescents

King's College London, London, UK

Research Analyst (intern) - 09/2014 - 12/2015

Published research papers on cannabis and psychosis



Selected Work Projects

	ML Pipelines on AWS - 01/2022 - currently
Role	AWS Specialist, Data Engineer
Company	NTTDATA — client: energy company
Tools	AWS, SageMaker, SQS, Spark, Sqoop, HDFS, Python, Scala, Bitbucket, Bamboo
Description	Data transfer from relational DB to S3 with Sqoop; optimization of Spark processing of millions of records in terms of cost and time reduction; deployment on SageMaker with StepFunctions and Lambdas; publication of predictions through SQS message broker
	MLOps software development - 07/2021 - 01/2022
Role	Backend Developer, MLOps
Company	NTTDATA — client: bank
Tools	Python, Flask, Nginx, UWSGI, SQL, MinIO, Docker, Kubernetes, GitLab
Description	Developed backend in Python of MLOps software that allows automation of ML pipelines through management, catalog, retraining and monitoring of machine learning models. REST API deployed as containerized microservice on Kubernetes cluster.

Publications

- Author of patent for Recommender System for Logistics (owner: LOGIVAN)
- Publications in the field of Psychology: www.researchgate.net/profile/Enrico_Foglia/contributions

Conferences

- TopDev Vietnam — Speaker — 2019: Presented my work on Recommender Systems
- Barcamp Vietnam — Speaker — 2019: Presented my work on Recommender Systems
- Uk-Vietnam Research Link, IOT workshop - 2018
- Schizophrenia International Research Society — Conference Poster Presentation - 2016

Languages

- Italian (native)
- English (fluent)
- Spanish (basic)

		MLOps software development - 04/2021 - 07/2021
Role	Backend Developer, MLOps	
Company	NTTDATA — client: energy company	
Tools	Azure Cloud, Python, RabbitMQ, Docker, MinIO, Postgres	
Description	Developed backend in Python of MLOps software. Implementation tailored to fit client's architecture (deployment on Azure Kubernetes cluster, asynchronous communication among microservice through message broker - RabbitMQ). In addition, created Python packages deployed on Azure Artifacts	
		NLP - Cognitive Search - 01/2021 - 03/2021
Role	Data Scientist & Backend Developer	
Company	NTTDATA — client: gas company	
Tools	Python, Scala, Flask, MongoDB, Docker, GitLab	
Description	Developed a cognitive search engine based on Word Embeddings using Word2Vec. Optimized Search speed through Locality Sensitive Hashing (LSH) implemented in Python and MongoDB aggregation language, and using MongoDB sharding. Flask API deployed within microservice architecture.	
		Price Prediction for Logistics Orders - 06/2020 - 12/2020
Role	Data Scientist & Backend Developer	
Company	LOGIVAN	
Tools	AWS, Python, Keras, Flask, Docker, GitLab, AWS, Redis	
Description	Prediction of winning price for any bidding made by suppliers to win logistics orders, with an average percentage error of 15%. Built a fully-connected neural network with embeddings using Keras. Deployed Flask API on AWS EC2 containerized with Docker.	
		Recommender System for Freight Matching - 07/2019 - 06/2020
Role	Data Scientist & Backend Developer	
Company	LOGIVAN	
Tools	AWS, Python, Keras, Flask, Docker, GitLab, AWS, Redis	
Description	Identification of the most suitable suppliers for logistics orders through a Recommender System. Recommendation task re-framed as binary classification with F1-score of 0.8. Adopted a siamese neural network with embeddings and triplet loss function with triplet mining techniques, as inspired by Google FaceNet (Schroff et al., 2015). Deployed Flask API on AWS EC2 containerized with Docker. The algorithm was patented.	
		Half-hourly Energy Consumption Forecasting - 01/2019 - 03/2019
Role	Data Scientist	
Company	Opus Energy	
Tools	Python, AWS Sagemaker	
Description	Half-hourly consumption profiles of new customers with no historical consumption data were estimated through a regression model based on customer characteristics. This POC slightly outperformed the standard UK Consumption Profile Classes traditionally used to categorize new customers.	
		Big Data analysis for Security Threats Identification - 03/2018 - 04/2018
Role	Data Scientist	
Company	MSc Coursework	
Tools	Hadoop, Spark, Impala, PySpark, Scala, HDFS, Sqoop	
Description	Millions of twitter feeds analysed for timing and amount to identify security threats during football matches, using Hadoop ecosystem (Spark, Impala, HDFS, Sqoop)	

	SQL Database Design for Email Marketing - 01/2018 - 02/2018
Role	SQL Developer
Company	MSc Coursework
Tools	SQL, SSMS, SSIS
Description	Designed database for email marketing using SQL Server Management Studio (SSMS). Followed DB design principles (normalization, lookup tables, constraints, isolation in concurrency control, atomicity and consistency in transaction management). Complex stored procedures and triggers allowed a variety of actions (subscribe, unsubscribe, etc.) and tracking of user metrics (timing of user behaviour, geomapping, etc.). Bulk actions such as bulk subscription enabled using SSIS. Full, differential and transaction-log backups implemented. DB security ensured through authorization and authentication methods.



Selected Personal Projects

Full-stack Web App Development (in progress)

Github: github.com/Enrico4444/DbSearch/

Tool that, given a database schema, allows user to insert data, import tables from excel, search tables through a UI with AND / OR operators, and view combined data from multiple tables. Developed to demonstrate mastery of all aspects of web development.

- Python API (Flask + UWSGI + Nginx) Backend
- Angular Frontend ([to be done](#))
- Postgres DB and MinIO object storage
- Security and User management with JWT
- Kubernetes deployment ([to be done](#))
- AWS deployment ([to be done](#))
- CI/CD in Jenkins ([to be done](#))

ML Algorithms from scratch

Github: github.com/Enrico4444/AlgosFromScratch

ML and Deep Learning algorithms coded from scratch in numpy, to learn and apply mathematics behind each model



Technical Stack

	Experienced	Intermediate - Basic
Programming lang	Python	Scala, C++, Java, R
Cloud	AWS	Azure
Databases	SQL - any dialect, MongoDB	
Shell scripting	Bash	PowerShell
ML frameworks	Tensorflow, Spacy	PyTorch
OS	Windows, Linux	
CI/CD & DevOps	GitLab	Jenkins, Bitbucket, Bamboo
Containers	Docker	Kubernetes
Big Data	Spark, Sqoop	
Others	MinIO	