Laura Galera

Data Engineer (AWS | Pyspark | SQL)

Contact
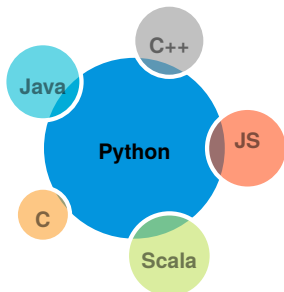
in ✉ ⦿

Date of birth

March 2nd, 2000

Location

Girona - Spain

Phone

(+34) 675 91 50 24

Languages

Catalan ★★★★★
Spanish ★★★★★
English ★★★★☆

Programming

Presentation

I am a Data Engineer at Edpuzzle, focused on designing, maintaining, and optimizing our AWS-based data system. Beyond constructing ETL pipelines, my role extends to orchestrating our data lakehouse and ensuring integration with multiple databases, utilizing AWS services such as S3, Glue, and Redshift.

Prior to joining Edpuzzle, I earned a M.Sc. in Artificial Intelligence from Stockholm University and collaborated with Spotify on Explainability in AI for ranking models. I also obtained a B.Sc. in Computer Science with a specialization in Computing, from the University of Girona, and spent six months at the Institute for Genomics and Bioinformatics at the University of California, Irvine conducting my final thesis on Convolutional Neural Networks.

My interests lie in the fields of databases and machine learning systems. They include cloud data-intensive applications, scalable data analysis, query processing, and optimization for machine learning systems.

Education

2022 - 2023	**M.S. in Artificial Intelligence** **Grade:** 9/10. **Activities:** Member of Student Union DISK.	**Stockholm University (SU)** | Sweden
2018 - 2022	**B.S. in Computer Science** *International Competence* **Grade:** 9.37/10; top 1%. **Activities:** Computer Science Degree mentor. Member of FEMenGIN.	**University of Girona (UdG)** | Spain

Experience

11/23 - Today	**Data Engineer** **Skills:** Pyspark, AWS, Redshift, Glue, S3, Docker, IaC. • Developed a data ingestion pipeline that consolidates data into our data lake from various sources including Salesforce, Mixpanel, and MongoDB, utilizing AWS Glue Jobs, Appflow, Lambdas, Step Functions. • Programmed Pyspark scripts to transform, process and move data to Amazon Redshift, designing the logical schemas and tables of our data warehouse, and implementing recurrent data analysis tasks in SQL. • Led the migration of our Data System to Infrastructure as Code (IaC) model employing AWS CDK library in JavaScript.	**Edpuzzle** | Spain
02/23 - 06/23	**Research Assistant (Master Thesis)** **Skills:** Python, TensorFlow, XAI, Generalized Additive Models, Local Interpretable Model-agnostic Explanations. • Conducted research in Explainable Artificial Intelligence based on interpretable ranking generalized additive models. • Explored techniques for adapting Local Interpretable Model-agnostic Explanations to ranking problems. • Compared different approximations of LIME for ranking to ascertain their accuracy and applicability.	**Spotify** | Sweden

02/22 - 07/22	**Research Assistant**	**University of California, Irvine (UCI)**	U.S.

Research Assistant — University of California, Irvine (UCI) | U.S. (02/22 - 07/22)

Skills: Python, TensorFlow, Convolutional Neural Networks.

• Developed a pipeline to clean and pre-process a data set of colonoscopy images from UCI's center containing polyps and non-polyps.

• Built a convolutional neural network to classify polyp images; evaluated its performance with two optimizers: Gradient Descent (SGD), and Adam.

• Created an ensemble model to detect polyps in colonoscopy images combining three pre-trained architectures –VGG, ResNet, DenseNet– with an accuracy of 96%.

Junior Data Scientist — Additio | Spain (06/21 - 09/21)

Skills: Python, Flask, ElasticSearch, Recommender Systems, Clustering.

• Performed clustering methods to report groups of students with similar learning weaknesses.

• Analyzed features within students in the same cluster to provide an explanation via text.

• Designed a website using Flask to retrieve and filter data from an Elasticsearch database and visualize the model's output.

Certifications

AWS Certified Developer - Associate — AWS (2025)
Credential ID: 41951d72-d841-4a7f-8927-d64a4428f162.
Expires January 2028.

AWS Certified Data Engineer - Associate — AWS (2024)
Credential ID: 41951d72-d841-4a7f-8927-d64a4428f162.
Expires November 2027.

AWS Certified Solutions Architect - Associate — AWS (2024)
Credential ID: 41951d72-d841-4a7f-8927-d64a4428f162.
Expires September 2027.

Neo4j Certified Professional — Neo4j (2023)
Credential ID: 4bbe6414-788d-4ca7-854c-0c938f80a26f.
No Expiration Date.

C1 Advanced in English — University of Cambridge (2021)
Credential ID: B6541238.
No Expiration Date.

Publications

Local Interpretable Model-Agnostic Explanations for Neural Ranking Models (2024)
Co-Author | Swedish Artificial Intelligence Society
DOI: https://doi.org/10.3384/ecp208017

Honors & Awards

Best Thesis in Informatics — Patronat UdG (2023)
Awarded to the best bachelor's thesis in Informatics.

2023	**Bachelor's Degree Extraordinary Award**	University of Girona

2023 · **Bachelor's Degree Extraordinary Award** · University of Girona
Awarded to the highest GPA in Computer Science. 56 graduated students.

2022 · **Beca Josep Maria Ginés i Pous** · Patronat UdG
Awarded to 2 engineering students to pursue a graduate degree in a foreign university.

2022 · **Balsells Mobility Award** · Balsells Fellowship
Awarded to 7 Catalan engineering students to conduct their thesis at UCI.