Llorenç Capó Torres

Mathematician — AI Researcher — Deep Learning and Reinforcement Learning Enthusiast

(+34) 674 287 529 | llorenscapo2001@gmail.com | Llorenç Capó | llorenscapo

Professional Profile

Mathematics graduate with a Master's in Big Data Analytics and **hands-on experience in data analysis** and backend development. **Passionate about artificial intelligence**, especially deep learning and reinforcement learning. **Self-driven learner**, currently studying advanced resources such as Deep Learning, Reinforcement Learning and AWS Certifications. **Interested in research**, intelligent systems development, and contributing to impactful AI projects.

Professional Experience

NTT DATA **Mar 2025 - Aug 2025:** *Data Scientist Intern*

- Built an **interactive dashboard** to display medical insights about kidney health using Streamlit, pandas, and matplotlib.
- Performed an extensive **exploratory data analysis** on the characteristics of the Spanish population to derive features and insights related to healthcare center attendance by location, sociodemographic context, age, and other personal variables, leveraging pandas and NumPy.
- Conducted a **data quality and completeness analysis** on two different healthcare databases (BDCAP and RAE-CMBD) to identify problematic variables for future developments, using pandas and ydata-profiling.
- Analyzed **bias and population representativeness** in BDCAP and RAE-CMBD compared to the overall population, ensuring suitability for integration into a unified OMOP-format database.
- Developed the entire **ETL pipeline** from scratch to extract, transform, and load both BDCAP and RAE-CMBD databases into a common OMOP structure, using pymssql, pandas, and Spark.
- Managed the **creation, loading, and maintenance** of the final OMOP-format database in Microsoft SQL Server.

NTT DATA **Feb 2024 - Aug 2024:** *Software Developer Intern*

- Worked in a **microservices environment**, building several RESTful API implementations with JPA Hibernate to handle database queries, implement business logic following best practices, and deliver responses to the frontend using Java SpringBoot.
- Fixed multiple issues related to incorrect or missing frontend-to-backend requests, debugging with Postman and Java SpringBoot.
- Resolved logging issues by configuring **Log4j**, which improved error tracing and debugging efficiency across the project.
- Fixed a bug in the project's **CI/CD pipeline**, specifically in the Jenkinsfile during merge request acceptance.
- Developed and executed **unit tests with Mockito**, identifying and resolving several logic errors in the codebase.
- Worked under the **Agile methodology**, participating in Scrum ceremonies (sprints, story points, daily meetings) and using Jira and Confluence to track progress and collaborate with the team.

Education

Universitat de les Illes Balears (UIB) **Completed:** *BSc in Mathematics*

- ✔ Strong foundation in **Algebra, Probability, and Statistics**, enabling rigorous reasoning and research capabilities.
- ✔ **Problem-solving and resilience**: trained to face complex and abstract problems with confidence.
- ✔ Hands-on experience applying mathematical concepts through **Python programming**.
- ✔ Analytical mindset, logical thinking, and ability to design solutions from first principles.

Universitat de les Illes Balears (UIB) **In progress:** *MSc in Big Data Analytics*

- ✔ Practical experience in **data analysis and machine learning** using Python and R.
- ✔ Exposure to **Big Data frameworks**: Apache **Spark** for distributed data processing.
- ✔ Basic knowledge of **Cloud Computing** and **Docker** for scalable ML workflows.
- ✔ Team-based projects applying ML to real-world datasets, bridging theory with industry practice.

Projects

Master's Thesis: Multivariate Time Series Forecasting Comparison – Ongoing:

- Compared multiple **time series forecasting techniques** across diverse datasets (gold price, household electricity consumption, and retail sales), including Naive, ARIMA, SARIMAX, LightGBM, LSTM, TCN, N-BEATS, Prophet, PatchTST, GCN, GAT, FEAT, and GPT4TS (subject to change).
- Combined widely used models (e.g., SARIMAX, LightGBM) with **recent state-of-the-art methods** from research papers, tested on datasets with different characteristics such as noisy signals (gold prices) and highly seasonal patterns (electricity consumption).
- Focused on **programming, state-of-the-art deep learning** (CNN, GNN, RNN, Transformers, and foundational models), while evaluating key trade-offs: predictive performance (metrics), model interpretability, implementation complexity, and computational efficiency.

[GitHub Time Series Comparision Link]

TensorFlow **May 2025 - Present:**

- Built a **Feed-Forward Neural Network** from scratch with TensorFlow to predict insurance charges on a public dataset.
- Developed a **Convolutional Neural Network** from scratch to classify food images (sushi, steak, pizza), applying multiple **data augmentation techniques** to improve prediction accuracy.
- Applied **Transfer Learning and Fine-Tuning** on three well-known pretrained models (ResNet50, EfficientNetB0, MobileNetV2) on ImageNet for image classification tasks.
- Implemented both a basic **FNN architecture** and a **state-of-the-art N-BEATS model** from scratch in TensorFlow to forecast Bitcoin price.

[GitHub TensorFlow Project Link]

Reinforcement Learning **May 2025 - Present:**

- Implemented from scratch several **Reinforcement Learning algorithms** in Python using only NumPy, including Bandit Algorithms, Dynamic Programming, Monte Carlo Simulations, SARSA, Q-Learning, and Double Q-Learning.

- Integrated these algorithms into multiple learning environments for agents, such as K-Bandit, Static Gridworld, Windy Gridworld, Icy Gridworld, Racetrack, Jack's Car Rental, Blackjack, and Hill Climbing.

[GitHub Reinforcement Learning Link]

Machine Learning Basics with Python **Sep 2022 - Jan 2023:**

- Trained multiple **supervised regression models** (Linear Regression, LASSO, Ridge, SVM for regression, Decision Trees) on structured datasets (Iris, Boston Housing) using pandas and scikit-learn.
- Applied **supervised classification techniques** such as K-Nearest Neighbors, Logistic Regression, SVM, Decision Trees, Random Forests, and XGBoost on the same datasets.
- Implemented **unsupervised learning methods**, including Principal Component Analysis (PCA), LDA, K-Means, DBSCAN, and Hierarchical Clustering.
- Performed comprehensive **data preprocessing and cleaning**, including SMOTE for imbalanced datasets, MICE for missing value imputation, one-hot encoding for categorical variables, scaling and normalization (MinMax, StandardScaler), train-test splitting, and cross-validation strategies.

[GitHub Machine Learning Project Link]

Skills

- **Programming Languages:** Python, Java, R
- **ML/DL Libraries: TensorFlow**, **Scikit-learn**, **Pandas**, Matplotlib, NumPy, Seaborn, **Spark**
- **Databases: SQL**, JPA Hibernate
- **Tools: Git**, Jira, Confluence
- **Cloud Computing:** AWS SageMaker, AWS Bedrock, AWS CloudWatch, AWS Glue

Courses

- Generative AI for Start-ups and Scale-ups (Certification)
- AWS Certified Cloud Practicioner (Udemy)
- AWS Certified Machine Learning Engineer (Udemy)
- Full Java Developer Bootcamp 2024 (Udemy)

Languages

- Spanish – Native
- Catalan – Native (C1 Certified)
- English – Academic IELTS (B2 CEFR)