Mario Prieto, **Ph.D.**

Senior Data Scientist | AI Expert | Tech. Business Advisor

About me

Data Scientist with 10 years of experience, leading complex projects and generating valuable data-driven insights. Proven ability to solve challenging problems through innovative methodologies, with a strong commitment to integrity and positive impact..

Contact

marprigod@gmail.com

+34 625166845

Madrid, Spain

mario-pg

Languages

Spanish - Native Language

English - Prof. Knowledge

Internships

Elsevier Labs Cambridge, MA. EEUU. (Apr. - Jul. 2018)

Leiden University Medical Center, Leiden, Netherlands. (Sept. - Dec. 2017)

Norwegian University of Science and Technology, Trondheim, Norway. (May - Aug. 2015)

WORK EXPERIENCE

2021-Today — **Senior Data Scientist** — Madrid, Spain

Ministry of Defense

Applied expertise in Large Language Models (LLMs), deep learning, and traditional machine learning to develop advanced solutions in NLP, anomaly detection, and computer vision. Implemented Agile methodologies for iterative development and integrated CI/CD pipelines for seamless deployment and continuous optimization. Key tasks include:

- Developed NLP solutions leveraging LLMs for text translation, summarization, question answering, sentiment analysis, and text classification.
- Fine-tuned Large Language Models, such as BERT and Vision Transformer (ViT), to optimize performance and adaptability.
- Improved model efficacy using data augmentation techniques, LSTM neural networks, and the XGBoost algorithm.
- Designed and deployed optimized big data pipelines using Spark/PySpark, focusing on efficient ETL processes to enhance data workflows, ensure integrity, and design the business-oriented data structure.
- Served as a bridge between clients and the technical team, transforming business needs into business solutions. Offered expert guidance to leadership on AI principles, capabilities, and project feasibility. Ensured adherence to data governance best practices, safeguarding data integrity, security, and GDPR compliance.
- Conducted data analysis using unsupervised learning techniques, including clustering, dimensionality reduction, anomaly detection, and association mining.
- Retrieved and processed data via APIs, developed web scrapers, and performed data wrangling to ensure data quality and accessibility.
- Supported strategic decision-making through data mining and interactive visualizations.

2020-2021 — **Data Steward** — Málaga, Spain

Polytechnic University of Madrid

Managed the quality, consistency, and security of rare disease data. Implemented data standards and ensured compliance with FAIR principles. Collaborated with teams to promote data management best practices. Upheld responsible data handling and privacy protection, aligning with GDPR guidelines to safeguard user information.

2020-2021 — **Python Developer** — Málaga, Spain

FAIR Data Systems

In charge of creating ad-hoc solutions to address unique business challenges. This involved designing, coding, and implementing custom Python applications to automate processes.

2015-2019 — **Data Scientist - Researcher** — Madrid, Spain

Polytechnic University of Madrid

Contributed to the 'Quality and Trust from Provenance in E-Science' project, aiming to enhance data and result quality through provenance analysis. Specialized in NLP with expertise in sentiment analysis, text classification, and text mining. Also proficient in clustering techniques and deep learning. Designed and deployed a ML model achieving 86% accuracy in assigning a 'certainty' sentiment field.

EDUCATION

2023 — **M.Sc. Expert in Data Science and Big Data** — Madrid, Spain

Datahack School

2019 — **Ph.D. NLP/Text Mining field** — Madrid, Spain

Polytechnic University of Madrid

Degree: 10/10 Cum Laude

2016 — **M.Sc. Bioinformatics and Computational Biology** — Madrid, Spain

National School of Health ISCIII

2014 — **B.Sc. Biology** — Seville, Spain

University of Seville

Professional Skills

Project management | MLOps | Data quality | Statistical analysis | Programming languages | Machine learning and data modeling | Data visualization | Experiment design and evaluation | Big data technologies | GitLab | JupyterLab

Soft Skills and Strengths

Curiosity | Flexibility | Self Confidence | Ability to Plan and Organize | Autonomy | Adaptability | Eye for Details | Problem Solving | Team Working | Love Learning New Things | Leadership | Good communication | continuous learning | Critical thinking | Business analysis | Patience

Download My CV

Download my CV via the QR below.

</> PROGRAMMING LANGUAGES AND PYTHON LIBRARIES

- **Python**: Advanced
- **Pandas**: Advanced
- **Numpy**: Advanced
- **Scikit-Learn**: Advanced
- **Tensorflow/keras**: Advanced
- **Pytorch**: Intermediate
- **Matplotlib**: Advanced
- **Plotly**: Advanced
- **SQL**: Intermediate
- **PySpark**: Intermediate
- **Elasticsearch - DSL**: Basic

INFORMATION TECHNOLOGY SKILLS

Cloud	**Google Cloud Platform**: *Basic*
Container Management	**Docker**: *Intermediate* **Kubernetes**: *Basic*
Big Data Platforms	**Hadoop**: *Basic* **Apache Spark**: *Basic*
Visualization	**Tableau** : *Basic* **Kibana**: *Basic* **Power BI**: *Basic*

CERTIFICATES

- Deploy Kubernetes Applications on Google Cloud (Google Cloud, 2025)
- Machine Learning Operations (MLOps) for Generative AI (Google Cloud, 2025)
- Implement Load Balancing on Compute Engine (Google Cloud, 2025)
- Google Cloud Fundamentals: Core Infrastructure (Google Cloud, 2025)
- Introduction to Gemini for Google Workspace (Google Cloud, 2024)
- LLMOps (DeepLearning.AI, 2024)
- ChatGPT Prompt Engineering for Developers (DeepLearning.AI, 2024)
- What is Generative AI? (LinkedIn, 2024)
- Fundamentals of Project Management: Risks (LinkedIn, 2024)
- GDPR: Practical Implementation in the Company (LinkedIn, 2023)
- GDPR: European Data Protection Regulation (LinkedIn, 2023)
- Master SQL for Data Science (LinkedIn, 2023)
- Advanced Python for Data Scientists (LinkedIn, 2023)
- Tableau for Data Scientists (LinkedIn, 2023)
- Transformers: Text Classification for NLP Using BERT (LinkedIn, 2023)
- Elastic Stack (Great Learning, 2023)
- Fundamentals of Project Management (LinkedIn, 2023)
- Fundamentals of Project Management: Calendars (LinkedIn, 2023)
- Fundamentals of Project Management: Quality (LinkedIn, 2023)
- Fundamentals of Project Management: Communication (LinkedIn, 2023)
- Fundamentals of Project Management: Teams (LinkedIn, 2023)
- Fundamentals of Project Management: Budgets (LinkedIn, 2023)
- Fundamentals of Project Management: Stakeholders (LinkedIn, 2023)
- Fundamentals of Project Management: Ethics (LinkedIn, 2023)
- Applying Analytics to Business Problems (Great Learning, 2023)
- R Programming (Coursera, 2015)

PUBLICATIONS

Conference Proceedings
2023

Innovative FAIRification solution for a Rare Disease Patient-led Registry: The Duchenne Data Platform, Nawel Lalout, Mario Prieto, Alberto Camara, Eduardo Quemada, Núria Queralt Rosinach, Bruna dos Santos Vieira, Marco Roos, Peter A.C. 't Hoen, Elizabeth Vroom, Mirjam Franken, Rajaram Kaliyaperumal, Mark D. Wilkinson, *11th European Conference on Rare Diseases & Orphan Products (ECRD) 2022; Orphanet Journal of Rare Diseases*, 10.1186/s13326-022-00264-6

Journal Publication
2022

Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data, Rajaram Kaliyaperumal, Mark D Wilkinson, Pablo Alarcón Moreno, Nirupama Benis, Ronald Cornet, Bruna dos Santos Vieira, Michel Dumontier, César Henrique Bernabé, Annika Jacobsen, Clémence Le Cornec, Mario Prieto Godoy, Núria Queralt-Rosinach, Leo J Schultze Kool, Morris A Swertz, Philip van Damme, K Joeri van der Velde, Nawel Lalout, Shuxin Zhang, Marco Roos, *Journal of biomedical semantics*, 10.1186/s13326-022-00264-6

Conference Proceedings
2021

EJP RD FAIRification Stewards to Harmonize FAIR Implementations Across ERN Rare Disease Registries, Shuxin Zhang, César Henrique Bernabé, Mario Prieto Godoy, K. Joeri van der Velde, Céline Angin, Arnaud Sandrin, Bruna dos Santos Vieira, *International Congress Of Research On Rare And Orphan Diseases*,

Journal Publication
2020

Tomographic and aberrometric assessment of first-time diagnosed paediatric keratoconus based on age ranges: a multicentre study, Carlos Rocha-de-Lossada, Mario Prieto-Godoy, José-María Sánchez-González, Vito Romano, Davide Borroni, Rahul Rachwani-Anil, Carmen Alba-Linero, Jorge Peraza-Nieves, Stephen B. Kaye, Marina Rodríguez-Calvo-de-Mora, *Acta Ophthalmologica Scandinavica*, 10.1111/aos.14715

Journal Publication
2020

Data-driven classification of the certainty of scholarly assertions, Mario Prieto, Helena Deus, Anita de Waard, Erik Schultes, Beatriz García-Jiménez, Mark D. Wilkinson, *PeerJ*, https://peerj.com/articles/8871/

Journal Publication
2019

Evaluating FAIR maturity through a scalable, automated, community-governed framework, Mark D. Wilkinson, Michel Dumontier, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Mario Prieto, Dominique Batista, Peter McQuilton, Tobias Kuhn, Philippe Rocca-Serra, Mercè Crosas & Erik Schultes, *Nature Scientific Data*, https://www.nature.com/articles/s41597-019-0184-5

Conference Proceedings
2018

Unraveling Certainty in Bio-Scholarly Statements, Mario Prieto, Helena Deus, Mark D. Wilkinson, *XIV SYMPOSIUM ON BIOINFORMATICS*, 10.5220/doidoidoi

Conference Proceedings
2018

Bacterial diversity in soils from contaminated and uncontaminated areas of a copper mine, Mariela Navas, Mario Prieto Godoy, Alejandro Rodrígue Iglesias, *VIII Congreso Ibérico de las Ciencias del suelo*, 10.13140/RG.2.2.22735.18080

Conference Proceedings
2016

Approaches to formally tracking, representing, and validating scientific argumentation, Mario Prieto, Mark D. Wilkinson, *Workshop 'New Frontiers in Plant Biology'*, 10.13140/RG.2.2.16178.45761