

JAVIER GARCÍA

+34 695718903 · edjagalo@gmail.com · [Linkedin](#) · [GitHub](#)
Barcelona CP 08508

Data Science MS - Business Intelligence MSc

A seasoned Data Analytics Professional with over 8 years of diverse experience encompassing roles from Engineer to Analyst to Data Scientist. Holds an M.S. in Data Science and an M.Sc. in Business Intelligence.

Skilled in utilizing Cloud Services, object-oriented programming, Machine Learning, and Deep Learning to execute comprehensive data workflows. Proficient in data extraction, transformation, prediction, streaming, analysis, and visualization.

Demonstrated success across multiple sectors including Technology, Pharmaceutical, Oil & Gas, Retail, Manufacturing, and Brew & Beer.

Key skills include critical thinking, analytical prowess, effective communication, and adeptness at balancing detail-oriented tasks with big-picture strategy.

EDUCATION

Data Science Master (2024)

Grade: Master of Science

Project: Development and Analysis of Machine Learning Models for Text Classification in Projects from the CORDIS Public Repository of the Science European Commission.

Keywords: Machine Learning, Text Classification, Data Analysis, Deep Learning, LLM.

Business Intelligence Master (2018)

Grade: Computational Master of Science

Project: Development of a clustering model for taxi drivers to optimize the corporate segment of Smart Taxi.

Keywords: Machine Learning, PCA, Data Analysis, Deep Learning, BI.

Industrial Engineer (2015)

Grade: Engineering

Project: Design and applied automatic forecasting system to improve supply chain in SYGLA.

PROFESSIONAL EXPERIENCE

NTT Data Data scientist

Oct 2023 - May 2024

Generative AI Development: Developed applications of generative AI for the health sector to support medical research for the Galician local government.

Machine Learning development: Developed and maintained machine learning algorithms on Google Cloud Platform for creating recommendation applications in the retail sector.

Data Architecture Improvement: Improved and refactored data architectures to ensure data quality across volume, value, veracity, variety, and velocity.

Bug Resolution: Resolved daily bugs in various data processes to ensure seamless business process streams.

Collaborative Development: Worked with mixed technical teams to develop and enhance applications.

AI Agent Evaluation: Evaluated AI agents using different metrics to enhance application performance.

Scalable Solutions: Utilized various technologies to create scalable solutions.

POC Creation: Created proof of concepts (POCs) for text classification using machine learning and generative AI for the European Union Commission of Science.

Achievements:

- **Text Classification Applications:** Created applications to scale text classification solutions using machine learning, deep learning, and generative artificial intelligence.
- **Alerting System Design:** Designed an alerting system to secure applications.
- **Synthetic Data Generation:** Generated synthetic data to evaluate AI agents.
- **Tool Development:** Developed tools to improve applications.

Senior Data Specialist and software developer

Data Warehouse Management: Created, maintained, and monitored data warehouses and data containers using various technologies and processes.

Pipeline Development: Developed, tested, and maintained pipelines that integrated multiple data sources, utilizing GCP, Python, and DAG to ensure efficient data ingestion and storage in data warehouses.

Cloud Data Solutions: Managed data storage solutions in Google Cloud Services, leveraging BigQuery, Cloud Storage, Dataflow, and REST API to process large volumes of data efficiently, adhering to business-defined timelines and formats while following software development best practices.

Achievements:

Architectural Design & Implementation: Developed the architecture and pipelines to ensure the entire process met stakeholder requirements effectively.

Mercado Libre

Jan 2022 - Aug 2023

Senior Data Engineer

Data Warehouse Implementation: Created and implemented data warehouses in Google Cloud Services and Amazon Web Services, leveraging S3, BigQuery, Data Flow, and Lambda to manage large data volumes efficiently.

Microservices Development: Developed and maintained microservices for transmitting structured, semi-structured, and unstructured data using AWS, Go, Python, and PySpark, ensuring scalable and efficient data processing in both batch and streaming modes.

ETL & Data Orchestration: Developed ETL processes and managed JOBS and DAGS using orchestrators like Dataflow, Airflow, and custom tools to ensure smooth data operations.

Programming Expertise: Utilized SQL (BigQuery, MySQL, Teradata), Python, Go, and Bash Shell to create effective data solutions tailored to business needs.

Emergency Watch Process: Established an engineering emergency watch process to oversee approximately 3,000 processes, ensuring data integrity across volume, value, veracity, variety, and velocity dimensions.

Achievements

- **Data Quality Improvement:** Developed new microservices to enhance data quality.
 - **Cost Minimization:** Refactored processes to significantly reduce operational costs.
-

Colombo American Center

Mar 2021 - Dec 2021

Business Intelligence Lead

Strategic Leadership: Led the BI team in developing and implementing data-driven strategies in collaboration with senior management, fostering a data-centric culture within the organization.

Data Architecture & Flow: Designed robust data architecture and established efficient data flow processes across functional areas, ensuring seamless integration and movement of data from various sources, including CRM systems and custom software APIs.

Technology Evaluation: Assessed and selected appropriate technology solutions (AWS, Azure, Google Cloud) to optimize data storage, management, and analysis, ensuring scalability and efficiency.

Best Practices Implementation: Implemented best practices for data storage and analysis, building a scalable infrastructure to support business needs.

KPI & Dashboard Development: Developed KPIs and dashboards using Tableau and Power BI, enhancing strategic planning and decision-making.

ETL Pipeline Management: Developed, tested, and maintained ETL pipelines to ensure smooth data flow and support decision-making across various departments.

Achievements

- **Team Consolidation:** Successfully consolidated the BI team, enhancing collaboration and productivity.
- **Process Improvement:** Improved data quality and established a new data-driven decision-making process.
- **Efficiency Gains:** Designed and developed applications that improved business processes, reducing process times by 85%.

KPI & Dashboard Development: Created key performance indicators and dashboards in Power BI to facilitate business decision-making, establish insights, and plan for improvements.

ETL Process Creation: Utilized Python (Pandas, NumPy) to develop ETL processes for transforming and analyzing historical sales data, aiding in sales, supply, and production planning at the regional level.

Data Prediction: Developed time series algorithms and other predictive methodologies to enhance forecast accuracy.

Data Pipeline Design: Designed and implemented new data extraction, transformation, and transmission processes, improving decision-making across management areas.

Digital Transformation Leadership: Led digital transformation projects with a focus on data-driven approaches, empowering teams and driving organizational change.

Achievements

- **BI Strategy Implementation:** Established the BI strategy for the department, significantly enhancing decision-making capabilities.
- **Cost Optimization:** Reduced purchasing costs by **26%** through the automation of business rules and monthly requirements.
- **Decision Support:** Reduced stock levels by **30%** with automated alerting dashboards.
- **Improved Prediction Accuracy:** Enhanced forecast accuracy and reduced RMSE metrics by over **15%** by upgrading algorithm quality.

STRENGTHS AND EXPERTISE		
Object-Oriented Programming (Python, R)	Cloud Computing (AWS, GCP, Azure)	Kafka
SQL - NoSQL	Database Management (ETL, Pipeline)	Git-GitHub
Machine Learning & Deep Learning	Data Architecture	Big Data