

Juan José Restrepo Rosero

restrepojuanjo@gmail.com

(+57) 316-6173202 • Bogotá, Colombia • [LinkedIn](#) • [GitHub](#) • [Portfolio](#)

PROFILE

Trilingual Electronics Engineer with a major in Mechatronics, a minor in Computer Science, and a completed Master's in Data Science from Pontificia Universidad Javeriana Cali. Experience in data engineering, machine learning, MLOps, and software development, building end-to-end data pipelines and production-ready systems. Skilled in designing scalable ETL/ELT workflows, model validation, and deploying machine learning solutions using modern data and cloud-native tools. Strong analytical and problem-solving skills, with a focus on reliability, reproducibility, and data-driven decision-making.

PROFESSIONAL EXPERIENCE

Veevart – DevOps Engineer: December 2024 – May 2025

- Developed and maintained backend services in **Java** and **Apex** within a production environment, integrated with Salesforce, ensuring scalability, reliability, and maintainability of core application logic.
- Collaborated on **backend architecture** and **debugging processes**, identifying performance issues, improving error handling, and supporting stable integrations between services and external platforms.
- Implemented automation workflows using **Python** and **Bash**, improving **CI/CD** processes and reducing manual intervention in deployment and operational tasks.
- Managed application **releases** and **version control**, applying **static code analysis (PMD)** and enforcing coding standards, contributing to improved code quality, maintainability, and production reliability.

Holcim ABS – Electronic Engineering Intern: July 2023 – January 2024

- Developed and maintained automation solutions using **Google Apps Script (JavaScript ES6+)** and **VBA**, integrating **Google Workspace** (Sheets, Drive) with **SAP** and **BTP** via **REST APIs** to automate data exchange, build reporting workflows, and improve data accuracy and availability for decision-making.
- Diagnosed and resolved issues in existing automations, performing **debugging** and **optimization** to ensure reliability and continuity of processes.
- Collaborated with **cross-functional teams** to analyze and document **business workflows** (O2C, P2P, R2R), identifying bottlenecks and driving data-driven process improvements to standardize operations and reduce manual workload.

EDUCATION

Pontificia Universidad Javeriana Cali,

Department of Electronics and Computer Science.
Master's Degree in Data Science, July 2024 – In progress

Santiago de Cali, CO

Department of Electronics and Computer Science.

Bachelor of Science in Electronics Engineering, January 2019 – April 2024

Cumulative GPA: 3.69/4.00

Southern Lee High School

Sanford, NC, USA

Foreign Exchange Student, High School Diploma, August 2017 - June 2018

Cumulative GPA: 3.70/4.00

San Antonio María Claret

Santiago de Cali, CO

High School Student, September 2011 - June 2017

Cumulative GPA: 3.80/4.00

SKILLS

Programming Languages:

Python, SQL, Java, Apex, JavaScript, TypeScript, C/C++, Bash

Data Engineering & MLOps:

ETL/ELT Pipelines, Apache Airflow, dbt, PostgreSQL, MySQL, Docker, GitHub Actions, Data Validation, Data

Machine Learning & Data Science:

scikit-learn, Pandas, NumPy, TensorFlow, Keras, NLP (spaCy, NLTK), Feature Engineering, Statistical Analysis, Data Visualization (Power BI)

Software Engineering:

Modular Architecture, System Design, Code Quality, Static Analysis (PMD, Ruff, mypy), Testing, Debugging

Tools & Technologies:

Git, Linux, Neo4j, MATLAB, Simulink

Modeling

Backend Development:

Node.js, Express, REST APIs, FastAPI, API Development, Request Validation, Error Handling

Automation & Scripting:

UiPath, VBA, Google Apps Script (JavaScript ES6+)

Languages:

Spanish (Native), English (C1 – TOEFL iBT), German (B1 – Goethe Zertifikat), French (A2)

PROJECTS

World Cup Prediction Engine – End-to-End MLOps System: March 2026

- Designed and deployed a production-grade MLOps system for football match prediction using a segment-aware hybrid ensemble (XGBoost), implementing a Champion/Challenger framework with automated promotion gates based on performance metrics to prevent model regression. Engineered a Medallion data architecture enabling continuous retraining with dynamic features (ELO, team form), integrated CI/CD automation (GitHub Actions), containerized with Docker, exposed via FastAPI, and incorporated SHAP explainability, monitoring, and model validation.

Knowledge Management & Multi-Spectroscopy Data Pipeline: May 2025 – January 2026

- Developed an end-to-end ETL/ELT data pipeline for spectroscopy data using Python, Airflow, PostgreSQL, and dbt, implementing Medallion Architecture (Bronze/Silver/Gold) to process heterogeneous datasets (Raman, FTIR, UV-Vis). Applied data validation, schema enforcement, and reproducible workflows, enabling reliable analytics, pattern detection, and data-driven decision-making for scientific research at iOmicas Institute.

Master's Thesis – Binary MIL for WSI Classification: May 2025 – Jan 2026

- Developed a deep learning pipeline for histopathological Whole Slide Image (WSI) classification using a Multiple Instance Learning (MIL) approach with weak supervision, including patch extraction, deep feature embedding, and attention-based aggregation. Designed patient-level cross-validation to prevent data leakage and evaluated model performance using standard classification metrics in a medical AI context.

Qversity Data Engineering Pipeline: June 2025 – July 2025

- Built a production-grade ELT data pipeline using Apache Airflow, dbt, PostgreSQL, and AWS S3, implementing Medallion Architecture and modular transformations with data quality testing and validation. Containerized with Docker and delivered 21 actionable business insights (ARPU, revenue segmentation, customer behavior), ensuring pipeline reliability, scalability, and maintainability.

Tasks Manager API – Production Backend: March 2026

- Developed a production-ready RESTful API using Node.js, Express, TypeScript, Prisma, and PostgreSQL, implementing JWT authentication, schema validation (AJV), rate limiting, and centralized error handling. Deployed on Render with environment-based configuration, documented with Swagger, and designed with modular architecture and backend best practices for scalability and robustness.

Formula 1 Predictor – Data Analysis & Machine Learning: March 2025

- Developed a data science application analyzing historical Formula 1 data from the Ergast API, performing EDA, statistical analysis, and feature engineering to model driver and constructor performance. Trained a Random Forest Classifier to predict podium finishes, achieving high accuracy and strong cross-validation performance, identifying key predictors such as grid position, driver performance, and constructor strength.

REFERENCES

Carol Villar

Director Network Engineering

Claro Enterprise Solutions

Contact: +1 (954) 494-8514