

Jesús Vélez Santiago

 <https://www.linkedin.com/in/jvelezmagic/>  <http://jvelezmagic.com/>

SUMMARY

Python Machine Learning Engineer with over four years of experience, specializing in developing and deploying innovative AI solutions. Expert in creating dynamic, AI-driven content and optimizing data pipelines for performance and scalability. Proven track record in solving complex problems, enhancing user engagement, and ensuring data integrity across diverse projects.

EXPERIENCE

Python Software Developer and Machine Learning Engineer

uDocz

June 2023 – Present, Remote

- Developed and deployed a Retrieval Augmented Generation (RAG) system, managing almost 1 million documents through integration of Python, Langchain, FastAPI, Postgres, and Redis, along with leveraging open-source embedding models.
- Engineered real-time, AI-powered flashcards and quizzes tailored to the content of documents, enhancing interactive learning experiences.
- Synthesized and optimized document summaries for both SEO and public engagement, ensuring content accessibility and visibility.
- Orchestrated and maintained data engineering pipelines, ensuring continuous data integrity and system observability with Airflow and DBT.

Machine Learning Engineer

Wannabe

September 2022 – June 2023, Remote

- Built automated infrastructure for efficient service deployment, allowing developers to focus on product development 100% and reducing infrastructure cost by 75%.
- Proficient in GraphQL API development using TypeScript, NestJS, and Apollo Tracing for performance optimization, reducing the error rate by 90%.
- Applied containerization and deployment with GitHub Actions, Docker, and Kubernetes and database design and optimization using PostgreSQL, going from a fully manual deployment (0%) to a wholly automatized deployment (100%).
- Data modeling, storage, training, and deployment of machine learning models.

Data Analyst

Platzy

January 2022 – July 2022, Remote

- Utilized MetaBase and Apache Superset to create dashboards with key performance indicators (KPIs) that resulted in a 25% improvement in the accuracy of business decisions based on data insights rather than assumptions.
- Conducted comprehensive modeling and data analysis from various sources such as Amazon Redshift, PostgreSQL, and Tinybird to develop new course evaluation metrics that resulted in a 40% increase in the ability to identify areas for course improvement and intervention.
- Optimizing databases to improve response times, resulting in a 50% decrease in the time required to extract data for analysis and enabling faster data processing for efficient data analysis and decision-making.

Bioinformatics Specialist

La Jolla Institute for Immunology

October 2021 – January 2022, San Diego, California, United States

- Automation of data cleaning, normalization, and preparation processes by 100% for pre-defined single-cell analysis pipelines.
- Automation of data visualization and generating experiment reports by 100% by allowing researchers to perform a self-exploration of the data, reducing the iteration time considerably.
- Creation of documented and tested internal code libraries for single-cell analysis that allow their reuse in new projects.

R Software Developer

Heidelberg University

August 2020 – June 2021, Heidelberg, Baden-Württemberg, Germany

- Developed and tested software across multiple operating systems, resulting in a 90% increase in overall code quality. Measured by code coverage, unit testing, and reduction of bugs by implementing the test-driven development methodology.
- Reduction of the acceptance time (by approximately 83%) of the software in Bioconductor, going from three months to just two weeks.
- Implemented more than ten statistic algorithms to extract biological signatures from prior knowledge within a unified framework.

CERTIFICATIONS

Data Scientist – Professional

DataCamp · 2021

- This certification is highly relevant as it verifies that the candidate possesses the essential knowledge, skills, and abilities required to excel in a Data Scientist role, such as data management, exploratory analysis, statistical experimentation, model development, coding for production environments, communication, and reporting, which were assessed through challenging timed assessments and a comprehensive practical exam graded by experienced data scientist experts

SKILLS

Tools & Technologies: Javascript, Typescript, R, Python, NestJS, NextJS, FastAPI, PyTorch, GraphQL, SQL, NoSQL, PostgreSQL, Amazon Redshift, Redis, Metabase, Apache Superset.

Industry Knowledge: Extract, Transform, Load (ETL), Machine Learning, Natural Language Processing (NLP), Large Language Models (LLMs), Data Science, Software Development, Data Visualization, Data Analysis, Statistics, Bioinformatics, Bilingual Communications, Retrieval Augmented Generation Systems.

Interpersonal Skills: Communication, Collaboration, Problem-solving, Critical thinking, Creativity, Adaptability, and Time management.