

Edgar Fragoso

GENAI & MACHINE LEARNING

📞 (+52) 55-211-849-86 | ✉️ edgarfragosogarcia@gmail.com | 🌐 edgar-omar-fragoso-garcia-69006092

Summary

A passionate Data Scientist and Machine Learning Engineer with over 11 years of experience. Recently, I've expanded my expertise into the forefront of generative AI, focusing on the implementation of Optical Character Recognition (OCR) and Large Language Models (LLMs). I have spearheaded the development of virtual assistants powered by Retrieval-Augmented Generation (RAG) technologies.

Work Experience

Minsait.

CDMX, Mexico

SENIOR CONSULTANT GENAI

May. 2024 - Aug. 2024

- **Achievement:** I spearheaded the implementation of GenAI technologies at CI Bank's legal department, enhancing the extraction of key entities from legal documents. This initiative involved developing a robust API that systematically returns these entities in a JSON format, thereby streamlining data accessibility and integration across legal and financial platforms. This strategic enhancement not only optimizes document handling processes but also ensures accuracy and speed in data retrieval and utilization within the bank.
- **Challenge:** Implementing technologies based in GenAI powered by GCP that allow legal department from CI Bank extract entities from legal documents as a power of attorney and charter.
- **Solution:** I implemented OCR using Document AI and Vertex AI processors to accurately extract various entities from documents. The system was designed with rapid few-shot inference engineering, which was meticulously refined and optimized through an iterative process of trial and error. This approach ensured both high accuracy and efficiency in entity extraction.
- **Workflow:** The solution was mounted using FastAPI into Compute Engine powered by GCP.
- **Technologies:** OpenAI, Genimi, VertexAI, DocumentAI, FastAPI, AWS, Langchain, LightRag, Qdrant, Neo4j, AWS, GCP.

Grupo Salinas.

CDMX, Mexico

ARTIFICIAL INTELLIGENCE CONSULTANT

Jan. 2024 - May. 2024

- **Achievement:** Integration of module to AlopsHUB to predict crashes in node servers in bank systems. Implement forecast to Optimization of cash. Bring Python courses to Data Science to collaborators of group.
- **Challenge:** Developed an AutoML tool for accurate crash predictions, distinguishing between normal operational variances and genuine anomalies, and automating retraining processes.
- **Solution:** Leveraged the Prophet forecasting model to train on recent (seven-day) data sets; implemented a dynamic stop-and-retrain mechanism for model accuracy.
- **Workflow:** Designed an automated system for data collection from log systems, utilizing Elastic to manage CPU usage data and trigger model training and alert notifications.
- **Technologies:** Python, Pandas, MLflow, Numpy, Prophet, ElasticSearch.

SomosDNA.

SF, U.S.A

DATA SCIENTIST & MACHINE LEARNING

Oct. 2020 - Sep. 2023

- **Achievement:** Spearheaded a project that significantly increased user engagement, directly contributing to a substantial profit boost in 2022.
- **Challenge:** Give to end users valuable knowledge of how genetic markers are linked with clinical condition.
- **Solution:** Given a data base with the features clinical condition and genetic markers its was implement a *Linear Model with regularization L1 and L2* to determinate what marker are linked to determinate condition furthermore I applied classical Machine Learning like SVM to understand the distribution of population in U.S. through of clinical reports
- **Workflow:** The prediction was automated using ECR, BATCH, EC2, S3, SQS and Cloud Watch all administration through Lambda functions.
- **Technologies:** MongoDB, Docker, Sklearn, Pandas, Numpy, Plotly and custumed snippet of code made in Python.

Código46.

Cuernavaca, México

DATA SCIENTIST & MACHINE LEARNING

Feb. 2017 - Nov. 2020

- **Achievement:** Automated key processes, enabling the company to focus more on sales and growth.
- **Challenge:** Informed users about their genetic ancestry and ethnic group probabilities.
- **Solution:** Implemented *PCA* for clustering analysis, *SVM* for predictions, and *DBSCAN* for outlier detection based on genetic-geographic data.
- **Workflow:** Developed and deployed Python scripts on Linux systems, integrating results with SQL for web visualization.
- **Technologies:** Azure, Docker, SQL, PySpark, Koalas, Sklearn, Pandas, Numpy and custumed snippet of code made in Bash and Python v3.7.

Skills

ML	Linear, Classification, Clustering
GenAI	OpenAI, Gemini, VertexAI, Langchain, LightRAG
Tech. DevOPS	Docker, Git, Python, Linux, Windows.
DB	MongoDB, SQL, ElasticSearch, Qdrant
Languages	Spanish, English
Tech. Management	Monday, Trello, SCRUM