

JOSE LUIS ALEJANDRO MUÑOZ ZEVALLOS

E-mail: alejandrozevallos.218@gmail.com | (+51) 947 535 428 | <https://www.linkedin.com/in/alejandromz2/>

AI & ML Engineer con 5+ años llevando IA a producción. Integro LLMs con LangChain (Gemini/OpenAI), diseño APIs seguras y orquesto despliegues en GCP (Cloud Run, Vertex AI, BigQuery, Artifact Registry). Fuerte dominio de Python, ML/DL y nube (GCP y Oracle Cloud). He construido soluciones de RAG y analítica que automatizan decisiones, elevan la calidad de datos y optimizan latencia/costos en entornos reales. Trabajo cercano con data, producto e infraestructura para entregar arquitecturas escalables, medibles y mantenibles, con buenas prácticas de versionado, pruebas y observabilidad.

Experiencia Laboral

Vooxel | Data Scientist (Alicorp) | Abr. 2025 - Actualidad

- Desarrollé un modelo de promociones para canal Tradicional y Moderno, el cual aprende de la data histórica de los productos de Alicorp y recomienda tandas promocionales óptimas y personalizables, dependiendo del objetivo específico de cada canal.
- Desarrollé un reporte utilizando consultas SQL de información de las ventas de la compañía, reduciendo en un 80% el tiempo de consulta de información por parte de negocio.
- Desarrollé una segmentación de proveedores para optimizar la base de gastos y apoyar al área de Procurement en la negociación de precios.

Tecnologías Principales: Python, SQL, Google Cloud, BigQuery, Vertex AI, Git

Anyone AI | Machine Learning Engineer (Part-Time) | Ene. 2025 - Jun - 2025

- Se implementó un pipeline de ingeniería de datos utilizando una base de datos de prueba de ecommerce, realizando un ETL usando funciones en Python y realizando una orquestación con Airflow.
- Se implementaron algoritmos de regresión logística y random forest con la librería de Scikit-learn, realizando un cross validation para identificar qué cliente es apto para recibir un crédito o no, realizando la validación de los modelos mediante una curva ROC y obteniendo un score de 0.81 en el entrenamiento y 0.73 en la validación.
- Se desarrolló una aplicación basada en microservicios, estructurando el frontend, backend, modelo de Machine Learning y base de datos en Postgres como servicios independientes dentro de contenedores Docker, todos implementados en Python. La aplicación está diseñada para la clasificación de imágenes utilizando el modelo ResNet50.
- Se almacenó la imagen de docker en Artifact Registry de GCP para su posterior despliegue en Vertex AI, asegurando una infraestructura escalable y optimizada.

Tecnologías Principales: Python, SQL, PostgreSQL, Docker, AWS, Apache Airflow, Scikit-learn, Google Cloud, Vertex AI

Ferreiros SA | Python Developer | Ago. 2024 - Actualidad

- Desarrollé una automatización en Python para la carga de Nueva Información de Caterpillar a un spreadsheet de Plan de Trabajo, reduciendo el tiempo de selección de los documentos que aplican al parque de máquinas de Ferreyros.
- Desarrollé una aplicación en Python para facilitar la creación de Kits de Compras en el área de ingeniería a partir de datos de una base de datos de entrada de repuestos por Work Order almacenada en BigQuery, reduciendo los tiempos de creación de nuevos kits en un 75%.
- Desarrollé un Stored Procedure en BigQuery mediante el cual se valida en tiempo real si se cuenta con repuestos disponibles para la aplicación de una mejora de Caterpillar. Esta validación se agrega como un mensaje en la descripción de la mejor que se encuentra subida en la plataforma Smarty mediante una implementación en AppScript.
- Desarrollé una funcionalidad para el chatbot Qimi utilizando las APIs de Google Cloud, permitiendo capturar mensajes en los chats de los especialistas de recepción de componentes. El número de serie del componente se envía mediante una solicitud POST a una API en Python, desplegada en Vercel, que lo redirige a una plataforma local donde un Web Scraping en Selenium valida la información. Esta implementación garantiza que los Kits de Compras solicitados sean los correctos.

Tecnologías Principales: Python, SQL, BigQuery, GCP, Apache Kafka, Appscript, AppSheets, Power BI, Flask, Selenium

Fima SAC | Data & Automation Specialist Internship | Feb. 2024 - Jul. 2024

- Diseñé un dashboard de ventas en Power BI, integrando datos procesados desde vistas en SQL mediante una conexión ODBC al ERP Global Shop. Esta implementación optimizó la visualización de la información en las reuniones de Gerencia General y Ventas, mejorando el análisis y la toma de decisiones.
- Aplicación de un modelo de Marketing Mix Modeling en el área de ventas para analizar el impacto de las inversiones por cliente, ventas perdidas por industria y producto, ventas efectivas y presupuestos proyectados a dos años. El modelo permitió identificar que los sectores de Gran Minería e Industrial presentan la mayor cantidad de ventas perdidas, proporcionando información clave para la optimización de estrategias comerciales.

Tecnologías Principales: Power BI, Python, SQL, Bizagi, Arduino

Venkizmet | ML Specialist | En. 2018 – Ago. 2022

- Desarrollé dos proyectos MVP centrados en Machine Learning: una aplicación en nube de Microsoft Azure ML y un modelo de predicción de clientes para un programa de lealtad en Oracle Notebooks.
- Implementé algoritmos de Regresión Lineal y Árboles de Decisión en Azure ML Studio, realizando pruebas y calculando métricas de evaluación para identificar el modelo más preciso.
- Integré Python en Jupyter Notebooks para conectar y extraer datos desde Oracle Notebooks, optimizando los procesos de clasificación de clientes elegibles para préstamos bancarios.
- Lideré sesiones de capacitación para el equipo sobre el uso, desarrollo y diferencias entre algoritmos disponibles en la plataforma, fomentando mejores prácticas en el desarrollo de modelos de ML.

Tecnologías Principales: Oracle Cloud, Azure ML Studio, Python, Oracle APEX, OCI Compute, Oracle DataWarehouse, SQL

Proyectos

Segmentación de prestadores desplegada en Vertex AI | 2025

Proyecto de segmentación en dos clústeres con foco en **despliegue a producción**. Entrenamiento y empaquetado en Vertex AI; artefactos de ambos clústeres almacenados en Cloud Storage (GCS) y orquestación del flujo con KubeFlow Pipelines.

Chatbot Agente SQL para Ecommerce desplegado en Cloud Run | 2025

Este proyecto consistió en el desarrollo de un chatbot con Langchain que funciona mediante Telegram, el cual se conecta mediante un webhook a un backend desplegado en un contenedor de docker almacenado en Artifact Registry y desplegado en Google Cloud Run mediante Terraform. Este chatbot fue diseñado para que el negocio pueda realizar consultas sobre información en diferentes tablas de una base de datos subida en BigQuery.

Arquitectura RAG para estudio de abogados | 2024

Desarrollé un chatbot en Flask para la gestión de documentos judiciales, almacenando archivos en formato PDFs, en Google Cloud Storage. Generé embeddings con un TextEmbeddingModel de Vertex AI, almacenando los vectores en un bucket de GCP y conectándolos a Compute Engine para la creación de un índice de búsqueda de vectores. Implementé Gemini 1.5 como modelo de IA generativa y desarrollé el proyecto en Python, permitiendo la gestión de grandes volúmenes de información legal y consultas judiciales con contexto.

Algoritmo de aprendizaje para la clasificación musical mediante Machine Learning | 2020

Este proyecto consistió en el procesamiento de un conjunto de datos musicales de Kaggle, para posteriormente separarlos por frecuencias usando la Transformada de Fourier. A partir de estos datos se entrenó un modelo de Clasificación Logística Binaria y Clasificación Logística Multiclase tanto en Python como en Matlab.

Educación

Universidad Nacional de Ingeniería | Jun. 2025 - Diciembre 2026

Maestría en Ciencias e Ingeniería Estadística

DMC Perú | Apr. 2024

Especialización en Advanced Data Science

Universidad de Ingeniería y Tecnología | Dic. 2023

Bachiller en Ingeniería Mecatrónica

Certificados

NVIDIA | Deploying RAG Pipelines for Production at Scale | **Mar. 2024**

NVIDIA | Efficient Large Language Model Customization | **Mar. 2024**

NVIDIA | Building AI Agents with Multimodal Models | **Mar. 2024**

IBM | IBM Python for Data Science and AI | **Abr. 2024**

Oracle University | Oracle Cloud Infrastructure 2023 AI Foundations Associate | **Dic. 2023**

Stanford University & DeepLearning.AI | Machine Learning by Stanford University & DeepLearning.AI | **Oct. 2024**

Herramientas

Habilidades Técnicas: Python, NumPy, Scikit-learn, Pandas, Matplotlib, Seaborn, Flask, LLMs (n8n, Langflow, LangChain, Milvus, Qdrant, Pinecone, GPT-4, Ollama), Tecnologías Cloud: Oracle Cloud (Data Warehouse, Notebooks, APEX, VM Instances), GCP (BigQuery, Vertex AI, VM Instances, Cloud Storage, APIs & Services), Redes Neuronales: Hugging Face Transformers, YOLOv8, TensorFlow, PyTorch, Keras, SQL, MATLAB, JavaScript, InfluxDB, Grafana, Docker.

Lenguajes: Fluído en Inglés, Español Nativo.