



Pablo Moreira

INGENIERO DE DATOS

Ingeniero de Datos con más de dos años de experiencia en el diseño, desarrollo y orquestación de pipelines de datos y plataformas cloud. Experiencia en entornos empresariales e investigación, trabajando con Python, SQL, Spark y tecnologías de contenedores y orquestación.

+34 601154746

Pablo.moreira.data@gmail.com

<https://pablomoreiragarcia.github.io>

EXPERIENCIA

2024-2025 (INGENIERO DE DATOS)

POSSIBLE INC.

- Diseño y mantenimiento de pipelines ETL/ELT en Python y Spark (PySpark) para proyectos de alto impacto, procesando datos de múltiples fuentes internas y externas.
- Implementación y gestión de flujos de trabajo con Apache Airflow, reduciendo incidencias en producción y mejorando los tiempos de respuesta de los procesos críticos.
- Desarrollo de asistentes inteligentes basados en RAG (Retrieval-Augmented Generation) sobre documentación técnica, utilizando LangChain, bases de datos vectoriales y LLMs.
- Configuración y despliegue de microservicios y componentes de datos con Docker y Kubernetes en Azure, integrados en pipelines de CI/CD.
- Colaboración con equipos multidisciplinares para entender necesidades de negocio y traducirlas en soluciones técnicas escalables y mantenibles.

2023-2024 (PERSONAL DE APOYO A LA INVESTIGACIÓN)

I3A - UNIVERSIDAD DE CASTILLA - LA MANCHA

- Investigación aplicada en computación paralela de redes bayesianas utilizando Java y técnicas de multihilo (JavaThreads).
- Desarrollo de scripts automatizados para la ejecución de experimentos a gran escala, trabajando con grandes volúmenes de datos.
- Participación en la redacción y revisión de artículos científicos basados en los resultados experimentales.

2021-2021 (DESARROLLADOR DE SOFTWARE)

OPINATOR

- Prácticas de la universidad, desarrollo backend

HABILIDADES

- Lenguajes de Programación: Python, Java, Spark (PySpark), Scala, SQL.
- Sistemas Operativos: Linux, Windows.
- Frameworks y Herramientas de Desarrollo: Django, Flask, FastAPI, Pandas, Numpy, Scikit-Learn, LangChain, LangGraph.
- Orquestación y Contenedores: Docker, Kubernetes, Apache Airflow.
- Cloud y DevOps: AWS, Azure, Git.
- Entornos: Jupyter Notebooks, Visual Studio Code.

IDIOMAS

- Castellano | Nativo
- Inglés | Avanzado (B2)

FORMACIÓN	2025 - LANGCHAIN, LANGGRAPH Y AGENTES IA CON PYTHON Udemy
	2024 - TABLEAU DESKTOP Y TABLEAU PREP DESDE CERO Udemy
	2024 - BIG DATA MASTERY: PYSPARK, AWS, SCALA & DATA SCRAPPING Udemy
	2023 - MASTER EN BIG DATA Y COMPUTACIÓN EN LA NUBE Universidad de Castilla-La Mancha
	2022 - GRADO EN INGENIERÍA INFORMÁTICA Universidad de Castilla-La Mancha
PROYECTOS	TFG - PASARELA Y PROGRAMACIÓN EN PYTHON DE LAS APIs DE APRENDIZAJE AUTOMÁTICO WEKA + SCIKIT-LEARN + SHAPVALUES Desarrollo de un módulo en Python que integra modelos de WEKA (Java) bajo el esquema de Scikit-Learn, permitiendo aprovechar algoritmos avanzados ya validados en Java desde entornos Python. El proyecto se centra en la aplicación de técnicas de aprendizaje automático explicable (XAI), utilizando SHAP para generar explicaciones comprensibles de las predicciones. Como caso de estudio, se analizó una base de datos real de jugadores profesionales de fútbol sala, con el objetivo de predecir la aparición de lesiones musculares y entender los factores que más influyen en ellas.
	TFM - LA VIÑA WINE PREDICTION Desarrollo de un proyecto presentado al Datathon de CajaMar 2023, en colaboración con la cooperativa vinícola La Viña y The Weather Company (IBM). El objetivo fue construir modelos de aprendizaje automático capaces de predecir con antelación la producción de uva de la campaña a partir de históricos de cosecha y variables meteorológicas. Esto permite estimar la producción meses antes de la recolección y optimizar la planificación de recolección, procesado, almacenamiento y distribución. El proyecto integró técnicas avanzadas de modelado predictivo y análisis de datos a gran escala, reforzando competencias en sistemas inteligentes y machine learning aplicado al sector agroalimentario.
	LANGCHAIN, LANGGRAPH Y AGENTES IA CON PYTHON Asistente RAG multimodal que permite preguntar en lenguaje natural sobre documentos PDF combinando texto, tablas e imágenes. El backend (FastAPI + Qdrant + MinIO) ingesta los PDFs, extrae texto, tablas (Camelot) e imágenes, los vectoriza y los almacena para búsqueda semántica. El frontend en Streamlit ofrece carga de documentos, visualización de tablas/imágenes relevantes y respuestas generadas por un modelo LLM multimodal de OpenAI, con manejo de errores y healthchecks para todos los servicios.