

JUAN HERNANDEZ VARGAS

Científico de Datos | Especialista en ML & Modelado Predictivo | Fintech, Seguros & Biotech

+57 315 457 0159 | juandanihv@gmail.com | Bogotá, Colombia | <https://www.linkedin.com/in/juan-hernandez-vargas/>

GitHub: <https://github.com/Juanhv24>

RESUMEN PROFESIONAL

Científico de Datos con sólida formación en análisis estadístico y una certificación avanzada en Data Science. Experto en la construcción de modelos de Machine Learning (Boosting, Random Forest) y automatización de procesos complejos mediante Python y SQL. Especializado en la evaluación de riesgos financieros, predicción de comportamiento de usuarios y análisis de datos biotecnológicos complejos, transformando datos crudos en ventajas competitivas y estrategias de negocio de alto impacto

HABILIDADES TÉCNICAS

- **Ciencia de Datos & ML:** Modelos de Clasificación y Regresión, Gradient Boosting (CatBoost, LightGBM), Random Forest, Scikit-learn, Pruebas de Hipótesis.
- **Análisis & Procesamiento:** Python (Pandas, NumPy), SQL (PostgreSQL), Limpieza de datos y ETL.
- **Estadística Avanzada:** Técnicas de Bootstrapping, Análisis de Varianza, Intervalos de Confianza y Regresión Lineal.
- **Visualización & Herramientas:** Power BI (Nivel funcional), Streamlit (Despliegue de modelos), Excel Avanzado, Matplotlib y Seaborn.

PROYECTOS

Interconnect — Predicción de Fuga de Clientes

- Desarrollé una solución *End-to-End* para identificar usuarios en riesgo de cancelación, optimizando estrategias de retención.
- Entrené y comparé modelos de **LightGBM, XGBoost y Random Forest**, seleccionando LightGBM por su eficiencia superior.
- Entrené Alcancé un **AUC-ROC de 0.9054** y una precisión global del ~85% en el conjunto de prueba.
- Creé un Dashboard interactivo en **Streamlit** que incluye un simulador de riesgo en tiempo real y visualización de KPIs técnicos (SHAP, Matriz de Confusión).
- Identifiqué que la antigüedad menor a 2 meses y los contratos mensuales son los principales factores de riesgo.

OilyGiant — Modelado de Riesgo Operacional y Rentabilidad | Python, Bootstrapping, Scikit-learn

- Realicé una evaluación de riesgos exhaustiva para identificar las ubicaciones más rentables para el desarrollo de nuevos negocios mediante regresión lineal.
- Implementé técnicas de Bootstrapping para simular 1000 escenarios, calculando el intervalo de confianza del 95 % para las ganancias y cuantificando la probabilidad de pérdida.
- Ofrecí recomendaciones basadas en datos para maximizar los márgenes de ganancia, manteniendo el riesgo operativo por debajo del umbral del 2,5 %.
- Impacto: Validación de los procesos de toma de decisiones mediante la cuantificación estadística del riesgo.

Beta Bank — Análisis de la pérdida de clientes y comportamiento | Python, Scikit-learn

- Desarrollé un modelo de aprendizaje automático (Random Forest) para predecir la pérdida de clientes, analizando datos históricos de 10 000 clientes bancarios para identificar cuentas en riesgo.
- Se abordó el desequilibrio de datos mediante técnicas de sobremuestreo, mejorando la capacidad del modelo para detectar eficazmente los patrones de salida de clientes.
- Logré una puntuación F1 de 0,61 y un ROC-AUC de 0,86, demostrando una alta precisión al distinguir entre clientes fieles y clientes que se van.

EDUCACIÓN

- **Especialización en Análisis Estadístico para la Ciencia de Datos** | Universidad De La Salle | En curso (2026).
 - **Data Science Career Track** | TripleTen Bootcamp | Graduado (2026).
 - **Biología** | Universidad de La Salle | Graduado (2023).
-

EXPERIENCIA PROFESIONAL

Operations Customer Expert | Teleperformance **2023-Present**

- Análisis de métricas operativas (CSAT, QA, AHT) y generación de informes de rendimiento precisos para la dirección.

Pasante de Gestión y Planificación de Museos | Universidad de La Salle **2022-2023**

- Administración y categorización de bases de datos técnicas, asegurando la integridad de registros digitales.
-

IDIOMAS

- **Español:** Nativo
- **Inglés:** B2 (Competencia laboral profesional)