

AARON IVAN ORTIZ URZUA

Data Scientist | Business Intelligence y Analítica

Querétaro, México | Remoto (EUA / LATAM) o híbrido en Querétaro y Ciudad de México

aaron@aaronurzua.com | linkedin.com/in/aaron-urzua- | aaronurzua.com

RESUMEN PROFESIONAL

Data Scientist e Ingeniero Industrial Estadístico con más de 5 años construyendo la capa de reporte y predicción de operaciones en manufactura, tecnología para retail y estadística pública. Sustituí reportes manuales recurrentes por data pipelines automáticos y dashboards de autoservicio en Power BI, R Shiny y Python; construí un pipeline de NLP que codificó el 83% de 148,310 respuestas de texto libre sin revisión humana; y soy responsable del programa de mantenimiento predictivo que ha reducido hasta un 9% los incidentes reactivos. Trabajo de punta a punta: extracción SQL, ETL, modelado, despliegue y entrega a los equipos que lo usan.

COMPETENCIAS CLAVE

Análisis de Datos · Business Intelligence · Visualización de Datos · Desarrollo de Dashboards · Automatización de Reportes · ETL y Preparación de Datos · Limpieza de Datos · Analítica Predictiva · Series de Tiempo · Forecasting · Detección de Anomalías · Mantenimiento Predictivo · Análisis de Causa Raíz · Reporte de KPI · Procesamiento de Lenguaje Natural · Análisis Estadístico · Reporte a Stakeholders · Data Storytelling

HABILIDADES TÉCNICAS

Lenguajes y consulta: Python, R, SQL (MySQL, SQL Server), DBeaver, JavaScript

Análisis y Machine Learning: pandas, NumPy, scikit-learn, XGBoost, Gradient Boosting, Random Forest, Decision Trees, ARIMA, time-series decomposition and forecasting, TensorFlow, Hugging Face Transformers

Visualización y BI: Power BI, Plotly, Highcharts, Leaflet, ggplot2, Matplotlib, Seaborn, DataTables, data storytelling, Microsoft Excel (advanced reporting)

Apps analíticas y entrega: R Shiny, flexdashboard, Streamlit, Flask, FastAPI, Docker, REST APIs, Git, GitHub

Ingeniería de datos y plataformas: ETL and data pipeline design, data quality validation, SQL optimization, ServiceNow incident data, Microsoft Azure (AZ-900 certified), Databricks

Métodos: Mantenimiento predictivo, detección de anomalías, ingeniería de variables, análisis de Pareto, análisis de causa raíz, análisis estadístico de proceso

EXPERIENCIA PROFESIONAL

Toshiba Global Commerce Solutions | Senior Data Scientist

Ciudad de México, México | May 2023 - Actualidad

- Soy responsable del programa de mantenimiento predictivo del hardware en tienda de una cadena minorista nacional (terminales punto de venta, cajas de autoservicio, impresoras de ticket, servidores de tienda), **reduciendo los incidentes reactivos hasta un 9% sobre una base de unos 14,000 al mes.**
- Entreno y ajusto los modelos de predicción de falla con XGBoost como único data scientist del programa, sobre el histórico de eventos de la plataforma de monitoreo PAS (plataforma, tipo de falla, número de serie), para programar el mantenimiento de forma preventiva en vez de reaccionar a una caída.
- Pronostico las tasas de falla del equipo recién desplegado que aún no tiene histórico de servicio, con ARIMA y descomposición de series de tiempo, para fijar una referencia inicial y proyectar los reemplazos esperados por mes.
- Desplegué un transformer de Hugging Face que clasificaba las notas de solicitud de servicio en texto libre como falla o no-falla, mientras el programa dependía de la organización de Estados Unidos, **automatizando el 60% de una revisión hasta entonces completamente manual.**
- Diseño y optimizo las extracciones SQL sobre las tablas de eventos y alertas de equipo, y construyo variables por equipo a partir de un flujo de alertas aisladas, generando los conjuntos de datos sobre los que corren los modelos predictivos.
- Construyo pipelines ETL en Python sobre el historial de incidentes de ServiceNow del cliente, con validaciones automáticas de calidad (nulos, duplicados, tipo y rango), eliminando el preprocesamiento manual recurrente del ciclo mensual de análisis.
- Realizo análisis de Pareto y de causa raíz sobre las fallas por tienda y por caja para aislar los modos de falla dominantes, e impulso las acciones de confiabilidad resultantes con los equipos internos de operación y de campo.

- Entrego los dashboards de Power BI y R Shiny y el reporte ejecutivo del programa, dando seguimiento a las tendencias de falla por plataforma y por equipo para el área de operación.

Optimus Steel USA | Junior Data Scientist

Houston, Texas, EUA | Mar 2022 - May 2023

- Construí un modelo de predicción de temperatura de colada para el horno de afino de olla (árboles de decisión, random forest, gradient boosting) que **redujo las muestras requeridas por colada de 6 a 3 (-50%)**, ahorrando cerca de \$30 USD por colada.
- Publiqué el modelo como una app web interactiva donde el operador elige el grado de acero e ingresa las variables de proceso, para que la predicción corra en piso sin depender de un analista.
- Sustituí el reporte mensual de scrap, **3 días de trabajo de un analista cada mes**, por un pipeline automático de MySQL a HTML en R, entregando un dashboard de autoservicio con análisis de Pareto de códigos de no conformidad, mapas de calor y desgloses apilados.
- Implementé un sistema automático de detección y alertamiento de fugas de aceite en equipo hidráulico que envía la alerta a un grupo de Telegram y por correo en cuanto se cruza un umbral, acortando la ventana entre la anomalía y la respuesta de mantenimiento.
- Contenericé las aplicaciones analíticas con Docker y desplegué herramientas web sobre los modelos (Streamlit, Flask) para que el personal de planta pudiera correr predicciones sin depender de un analista.

INEGI - Instituto Nacional de Estadística y Geografía | Analista de Datos

Aguascalientes, México | Mar 2021 - Mar 2022

- Desarrollé el sistema de codificación automática SINCO/SCIAN de la Encuesta Nacional de Ocupación y Empleo (ENOE), alcanzando **82.84% de codificación automática en ocupación y 84.02% en actividad económica sobre 148,310 registros de texto libre**.
- Diseñé el data pipeline de NLP de cinco etapas detrás del sistema: normalización de texto, sustitución por diccionario de sinónimos, coincidencia contra tablas especiales y consulta a tablas de acceso con reglas INCLUYE/EXCLUYE, cada etapa asignando un indicador de confianza al código resultante.
- Entregué el sistema como una aplicación navegable en R Shiny que documenta metodología, ejecución y resultados, con análisis de frecuencia de palabras, nubes de palabras, análisis de sentimiento y tablas de resultados interactivas para el equipo de encuesta.
- Produje el reporte nacional y por entidad de la modalidad de captación de la EEN (internet, medios digitales, papel) para las 32 entidades, como reportes HTML interactivos con visualizaciones de Highcharts, mapas coropléticos de Leaflet y tablas de datos consultables.
- Limpié y preparé los conjuntos de datos de las encuestas ECIS 2021 y ENAFIN 2021 y procesé las variables de ocupación y actividad económica usadas en análisis y publicaciones institucionales.

FORMACIÓN ACADÉMICA

Universidad Autónoma de Aguascalientes (UAA) | Ingeniería Industrial Estadística

Aguascalientes, México | 2021

- Testimonio de Desempeño Sobresaliente en el EGEL-CENEVAL, evaluación nacional de egreso.
- Tutor Par y Reconocimiento Institucional (CCB), Universidad Autónoma de Aguascalientes.

CERTIFICACIONES

Nube: Microsoft Certified: Azure Fundamentals (AZ-900)

Python (IBM): Python for Data Science; Machine Learning with Python; Deep Learning with TensorFlow; Data Visualization using Python

R (IBM): Data Visualization with R; Machine Learning with R; Using R with Databases

Estándar de competencia (A2 Capacitación, 2021): Python (ECPY), R (ECRS), MySQL (ECSQL)

Control de versiones: Curso profesional de Git; Curso a fondo de GitHub

IDIOMAS

Español - Nativo | Inglés - Competencia profesional