David Vargas Leos

+52 (449)257-9204 | davale1201@gmail.com |

Education

Instituto Tecnológico y de Estudios Superiores de Monterrey — Monterrey, MX
B.S. in Data Science and Mathematics — *2021 – 2025*
- Scholarship for Academic Excellence. Member of Academic Group IAnswers.
- Courses: Optimization, Deep Learning, AI, Probability and Statistics, Data Analysis, Topology, Linear & Abstract Algebra.

Experience

Undergraduate Researcher Scholar — August 2024 – December 2024
Purdue University — *Supervisor: Dr. Michael Gribskov*
- Worked on a **high-throughput computational pipeline** for RNA secondary structure prediction, leveraging **high-performance computing (HPC) resources** and applying heuristic algorithms to detect isomorphic structures in RNA graphs.
- Engineered features to **improve clustering** of RNA sequences by structural similarity, supporting functional role prediction based on structural motifs.

Data Scientist — August 2023 – June 2024
GeoStats
- Helped to develop a **predictive model** for femicides in a given zone, working in a cross-functional team to integrate geospatial data using GeoPandas and applying multivariate statistical tools to validate predictions.
- Collaborated with the Government of San Pedro Garza García to determine optimal locations for pollution detectors through **geospatial analysis**, inspections, and local expertise, presenting insights via **Looker Studio** for better decision-making.

Web Developer — January 2023 – June 2023
Expertos Mindfulness
- Developed a responsive website for Expertos Mindfulness using Node.js, Express, and EJS.
- Structured and organized site content to ensure clarity and ease of navigation.

Projects

LLM-Powered Automated Scoring System for Public Fund Allocation | *Nuevo León Government* — February 2025
- Developed an **LLM-based automated scoring system** using **fine-tuned GPT-4o Mini**, turning Excel data into **JSON and JSONL** for model training. Achieved an **F1 score of 0.75**, suggesting a promising level of accuracy and scalability.
- Helped improve the efficiency and transparency of allocating **250,000 million MXN (12.5 million USD)** in public funds by automating evaluations, potentially reducing manual workload and aiding time-efficiency.

Gravitational Wave Signal Classification with Topological Data Analysis — June 2024
- Developed a **pipeline** to extract **topological features** from gravitational wave simulations using Takens Embedding and Vietoris-Rips Persistence, applying PCA for dimensionality reduction and **persistence entropy** to analyze signal complexity.
- Utilized **logistic regression** and **CNNs** to classify gravitational wave signals versus noise, improving detection accuracy.

Air Quality Prediction and Classification | *Nuevo León Government* — December 2023
- Collaborated with the **Air Quality Agency of Nuevo León**, using **PCA** and **logistic regression** to analyze pollutant dispersion and classify air quality in San Nicolás, achieving **79% accuracy** for PM10 and **82%** for PM2.5.

Time-dependent Orienteering Problem (OPTW) model — June 2023
- Modified the OPTW model to **optimize tourist routes** based on Points of Interests with time windows and Google Maps ratings, **improving itineraries** within time and budget constraints using Bing Maps API.
- Adapted the model to a scenario in Puebla, Mexico, **factoring in time and budget constraints**, and utilized GAMS for optimization and simulation.

Natural Language Interpreter for Classification | *Ternium* — June 2023
- Implemented an **Unsupervised Machine Learning model** with **NLP** and **K-means clustering** to automate classification of failure reports, addressing challenges like heterogeneous entries and spelling errors.
- Enhanced database standardization for **Ternium**, achieving **reduced analysis time** and improved data organization without manual intervention.

Personal Portfolio | *Node.js, Eleventy, Markdown, Liquid, Backend Development*

Skills

Languages: English(C1/TOEFL), French (B2 courses/not certified), Spanish(Native)
Coding Languages: Python, JavaScript, R, C++, C#, MATLAB
Libraries/Frameworks: React, Node.js, NEXT.js, Flask, Pandas, NumPy, Matplotlib, SciKit-Learn, PyTorch, Giotto
Tools: Git, VS Code, Bash, Linux, Anaconda, QGIS, GAMS, Excel, MongoDB (Distributed Storage), PostMan, LaTex