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
- Courses: Optimization, Deep Learning, Artificial Intelligence, Probability and Statistics, Data Analysis, Topology, Linear Algebra, Abstract Algebra

Experience

Undergraduate Researcher Scholar — August 2024 – Present
Purdue University — *Supervisor: Dr. Michael Gribskov*

- Working on a high-throughput computational **pipeline** for secondary structure prediction of hundreds of RNAs using **high-performance computing** resources.
- Managing ambiguity in data analysis and heuristic applications developed by **Dr. Gribskov** to the **NP-hard problem** of detecting isomorphic structures in graph representations of RNA secondary structures.
- Engaging in **feature engineering** to improve clustering of RNA sequences into structurally similar groups, aiding the prediction of functional RNA roles based on structural motifs.

Volunteer Experience

Data Scientist — August 2023 – June 2024
GeoStats

- **Helped build** a map that will help lawmakers, government officials, and Pro Bono organizations **make informed decisions** to choose which sectors of the state need the most help.
- Worked on a **predictive model** of femicides given a zone, using GeoPandas to work with the geospatial data, using multivariate statistical tools to justify and verify our predictions.
- Worked with the Government of San Pedro Garza García to **decide optimal locations** for pollution detectors based on **geospatial analysis**, on-site inspections, and local expertise.

Projects

Gravitational Wave Signal Classification with Topological Data Analysis |
Python, Scikit-learn, Giotto, PyTorch — June 2024

- Created a **pipeline** to extract **topological features** from gravitational wave simulations using Takens Embedding and Vietoris-Rips Persistence.
- Applied **PCA** for dimensionality reduction and calculated **persistence entropy** to distinguish signal complexity.
- Utilized **logistic regression** and **CNNs** to classify gravitational wave signals versus noise, enhancing detection accuracy.

Air Quality Prediction and Classification: A Study in San Nicolás | *Python, R* — 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 | *GAMS* — 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 | *Python* — 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*

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
Methodologies: Agile Development, Feature Engineering, Heuristic Algorithms, Computational Optimization