Osvaldo Velazquez Gonzalez

+52 961 192 6030 | osvaldodvego@gmail.com | linkedin.com/in/osvaldovg | github.com/valdolab

Languages: English (professional proficiency), Spanish

Summary

Dedicated Data Scientist, ML, and Software engineer with over 5 years of experience. Expertise in **predictive model development, deployment, and implementation** of artificial intelligence algorithms and methods. I have used **Python, SQL, Matlab**, and **R** throughout my career. Experienced in **NLP**, **Large Language Models (LLM)**, time series analysis, ETL process, and tabular data for classic classification and regression. I have used cloud computing for production ML models, such as **AWS** and **Azure**.

Experience

Senior Data Scientist Feb 2023 – Present
EnhanceIT *Atlanta, USA*

- Development end-to-end of a **churn prediction model** for insurance service applying XGBoost and AutoPilot in AWS SageMaker using demographic information about the customer.
- Development of a **sentiment analysis model**, and NLP solutions applying transformers, and transfer learning, development using **DataBricks** and **Python** in **Azure Cloud**.
- Development of a **text Generative model** to assist in the generation of product descriptions applying LLM, such as GPT3.5 Turbo. Development with **Python**, and **TensorFlow**.
- Collaborate with ML engineers, data engineers, and stakeholders to develop data and models in an accurate manner.
- Deploying and monitoring ML models in order to report and track model performance during the time.

Data Scientist and MLOps Engineer Aug 2022 – Feb 2023
Tata Consultancy services (Goldman Sachs Group, Inc. Client) *Guadalajara, Mexico*

- Migrated to AWS, and deployed ML classification models to identify high-risk clients and opportunities using **XGBoost**, and **Random Forest**.
- Development of pipelines of machine learning functions, such as pre-processing techniques, adapting and processing the data to call the endpoint of the model, ensuring reliable **model deployment**, and monitoring.
- Development of automated **pipelines** of the re-training process of machine learning models in production using SageMaker.

Machine Learning Research Assistant Sep 2020 – Aug 2022
Centro de Investigación en Computación, IPN *Mexico City, Mexico*

- Designed a **novel Machine Learning algorithm** to deal efficiently with class imbalance, with higher performance of various state-of-the-art models.
- Implemented several ML models, like **k-NN, RandomForest, C4.5, MLP, Naive Bayes, and SVM** in order to predict and pre-diagnose several respiratory diseases.
- Proposed and conducted ML research experiments applying post-hoc methods like **Friedman and Holm** to evaluate designed models against state-of-the-art models, applying data cleaning algorithms, feature selection, feature extraction, and feature engineering techniques.

Data Analyst Oct 2019 – Sep 2020
Subsecretaría de Educación Federalizada de Chiapas (SEF) *Tuxtla Gutiérrez, Mexico*

- Cleansed and prepared datasets, ensuring data integrity and reliability. Conducted intricate analyses, employing a combination of **SQL queries**, **Tableau** visualizations, **Power BI** insights, **VBA in Excel**, and using **R**. Transformed raw data into actionable insights, producing comprehensive reports that facilitated informed decision-making processes.
- Played in gathering **ETL requirements** and adeptly translated them into functional needs. Additionally, designed and developed analytical dashboards to visually represent complex data and insights, effectively communicating information to stakeholders.

Undergraduate AI Research Assistant Jun 2019 – Oct 2019
Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) *Tonantzintla, Mexico*

- Built an infant cry translator and pre-diagnosis disease interface, applied ML classification algorithms, such as kNN, and **neuro-fuzzy models** in order to **predict infant hearing loss-related diseases**, or ascertain the cause behind the infant's cry, handling classes like infant pain, infant hungry and infant boring.
- Handled with audio data, applied some spectrogram techniques, preprocessing algorithms, and feature extraction methods in order to convert audio files to easy-to-handle data.
- Developed and deployed the classification models proposed into a medical software tool to **support the early detection** of infant diseases using bioacoustic signals.

Software Engineer Jan 2018 – May 2019
Freelancer *Tuxtla, Mexico*

- Developed and maintained scalable software solutions using PHP, JavaScript, and Python.
- Collaborated in cross-functional teams to define and implement technical requirements of a web system software.
- Managed codebase using Git, facilitating seamless teamwork and version control.
- Implemented and optimized databases with SQL queries in MYSQL for web systems designed.
- Collaborated with stakeholders and transformed non-technical requirements into technical needs.

Education

MSc in Computer Science (Cum Laude)
Centro de Investigación en Computación, Instituto Politécnico Nacional (CIC-IPN) *Mexico City, Mexico*
Thesis: Computational Intelligence Algorithms for Medical Pre-diagnosis of Respiratory Diseases

BEng in Computational Systems
Instituto Tecnológico de Tuxtla Gutiérrez (TecNM) *Tuxtla Gutiérrez, Mexico*

Technical Skills

Languages: proficiency in Python, SQL, Matlab. Good understanding of R, PHP, JavaScript, Bash.
Frameworks and ML Tools: Flask, FastAPI, Git, Docker, GCP, AWS, VS Code, Databricks, PySpark, LangChain, AWS SageMaker, Google VertexAI, DataBricks, OpenIA libs, TensorFlow, Keras, scikit-learn, NLTK.
Libraries: Pandas, NumPy, Matplotlib, Seaborn, SciPy, Scikit-learn, TensorFlow, Keras, NLTK, Stanza.
Data Science and ML Specialities. *Natural Language Processing (NLP):* transfer learning, text embedding, universal sentence encoders, BERT, roBERTa. *Computer Vision:* EfficientNet, ResNet50, YOLO. *classic classifiers:* kNN, SVM, Neïve Bayes, MLP, Random Forest, XGBoost, linear regression. *Time Series:* ARIMAs families, NAR, LSTM. *Clustering:* KMeans.
Analytics: feature selection, and engineering, Bayesian modeling, PCA, L1 and L2 regularization, and hyperparameter tuning.
Optimization: metaheuristic, evolutionary, and genetic algorithms.

Certifications and Courses

Machine Learning Dec 2023
Imparted by Stanford University

Machine Learning with PySpark Oct 2023
Imparted by DataCamp

Introduction to MLflow Sep 2023
Imparted by DataCamp

Introduction to PySpark Sep 2023
Imparted by DataCamp

Database Desing and Programming with SQL Jun 2017
Imparted by Oracle