Valeria Vera

• Washington, D.C. • valveralagos@gmail.com • LinkedIn • Portfolio • (+1) 2027181340

EDUCATION

Master of Science in Data Science and Analytics   08/2022 – 05/2024
Georgetown University   USA
- Elective courses: Natural Language Processing, Large Language Models, Reinforcement Learning, Optimization, Deep Learning
- Award: Fulbright-García Robles grant to pursue a graduate degree in the US

Bachelor of Engineering in Information Technology - **Cum Laude**   08/2015 – 11/2020
Benemerita Universidad Autonoma de Puebla (BUAP) - Computer Science department   Mexico
- Thesis: Misogyny detection through supervised classifiers, results presented at GCURS Symposium at *Rice University*
- Research visit: Natural Language Processing using Java at the *University of Texas*
- Award: Nation's best career development in Information Technology Award *ANIEI*

WORK EXPERIENCE

Graduate Technical Intern – Machine Learning   05/2023 – 08/2023
Intel Corporation   Oregon, USA
- Automated data extraction from an Azure SQL database using Python and Kusto, lessening extraction time by 80%
- Performed data reduction, modeling, and correlation analysis for a 400,000 rows by 50,000 columns dataset
- Implemented unsupervised clustering techniques, evaluated similarity of clusters, and cut duplicated data by 60%
- Coded a Python-based website allowing users to obtain automatic clustering results decreasing analysis time by 90%

Research Assistant   01/2023 – 05/2023
Center for Security and Emerging Technology   D.C., USA
- Conducted an in-depth review of approximately 100 papers focused on "explainability" in AI systems
- Summarized and categorized the evaluation methodologies, benchmarks, and models adopted in the papers

Software Development Engineer   01/2021 – 08/2022
Intel Corporation   Jalisco, Mexico
- Collaborated on a test generation framework using Reinforcement Learning, decreasing test time by 40%
- Refactored a Python project legacy code reducing new costumer tools integration time by 60%
- Generated bash scripts for collecting hardware signals and communicating across more than 30 Unix cluster systems
- Debugged software and hardware memory issues in Linux systems in more than 20 Intel products
- Trained a supervised model that predicts electrical conditions variability, cutting analysis time by 92%
- Produced a Natural Language Processing solution that automates test content generation, reducing engineering time by 96%
- Published research results in the Design and Test Technology Conference (DTTC) 2022 and 2021

Software Validation Engineer Intern   12/2019–11/2020
Intel Corporation   Jalisco, Mexico
- Automated test content creation through Python scripts that reduced development time by 85%
- Developed a Client-Server communication model on Python that reduced hardware expenses by 80%
- Designed a supervised model (SVM) for software bugs classification using Natural Language Processing on Python

PROJECTS

OpenDCDiag - Intel   02/2024
Conducted a statistical analysis of CPU defects and developed data collection scripts in C++

Toxicity analysis of large corpora using transformers   01/2024
Analyzed the impact of toxicity levels in large corpora across NLP tasks

TPU performance prediction using Graph Neural Networks   08/2023
Optimized performance of Graph Neural Networks for TPU runtime prediction with Keras

Walkability impact in D.C.   01/2023
Collected data with the Reddit API, classified sentiment using transformers, and created visualizations using Plotly and Javascript

Revictimization: a misogyny detection problem   08/2022
Web scraped news articles with Python and finetuned a Large Language Model to detect revictimization with Pytorch

Math learning support system for kids with ADHD   11/2019
Developed a WebApp using Unity based on serious games research. Presented at National Congress of Education Technologies CONTE

SKILLS

Programming skills:	Object-Oriented Programming, Algorithms design, Design Patterns, Intel architecture, Git
Programming languages:	Python, Unix shell language, R, Java, C++, C, SQL
Data packages:	scikit-learn, numpy, pandas, matplotlib, plotly, keras, pytorch, spacy, huggingface
Big data frameworks:	Apache Spark, Hadoop
Languages:	Spanish: Native speaker, English: Full professional fluency