Contact

www.linkedin.com/in/ptoroisaza (LinkedIn)

Top Skills

Software Development
Discrete Event Simulation
SAS (Programming Language)

Languages

English (Native or Bilingual)
Spanish (Native or Bilingual)

Certifications

Master of Science

Honors-Awards

Mellon-Mays Undergraduate Fellowship
Tutor of the Month
Suzanne Keller Award for Academic Excellence in Sociology - Honorable Mention

Paulina Toro Isaza

Data Science & Machine Learning at IBM Research | Natural Language Processing | AI for Development Operations
White Plains, New York, United States

Summary

I am a data scientist trained in machine learning and statistics with a focus in natural language processing. I am an incredibly curious person who finds wonder and satisfaction in learning about and applying my technical skills to a wide variety of topics. Topics of past work include: development operations, agriculture, computational biology, academic publications, climate, law, education, transportation, astronomy.

2+ years: Data science, machine learning, natural language processing, cloud computing
3+ years: Professional data management
5+ years: Academic data analysis, experimental design, data collection, data cleaning, model selection

Experience

IBM
Research Software Developer
July 2021 - Present (1 year 10 months)
Yorktown Heights, New York, United States

Software Developer at IBM Research focused on machine learning and natural language processing for development operations research.

Intelligent IT Operations
• Designed and wrote production code for new Cloud Pak for AI Ops feature accessing data quality of text input and leveraging data quality metrics to improve NLP model performance.
Narrative Story NLP Bias Toolkit
• Designed toolkit for extracting characters and their events from a narrative story to analyze gender bias of events.

Inari
Machine Learning Engineer Intern

February 2021 - May 2021 (4 months)
Cambridge, Massachusetts, United States

• Adapted, trained, and validated NLP models in PyTorch for use with protein sequences and biological properties.
• Fine-tuned protein sequence specific BERT models for sequence classification, token classification, and regression.
• Extracted BERT protein token embedding spaces for analysis and comparisons.

Atos
Data Science Intern - Digital Transformation
July 2020 - September 2020 (3 months)

• Designed report dashboards, data collection schemes, and data models for Atos Green and related applications that help employees record and reduce their individual carbon footprints
• Collaborated in Agile environment with design and development teams to display complex quantitative relationships
• Researched and established links across data sources and studies, presenting findings to business leads

The Research Alliance for New York City Schools
Co-Data Manager
February 2017 - August 2019 (2 years 7 months)
Greater New York City Area

• Processed, cleaned, manipulated, and restructured structured and unstructured NYC Department of Education administrative and academic records from 1996 to present of over five million students and six hundred thousand teachers (Over 4 TB in size)
• Analyzed data using descriptive statistics and inferential statistics including multiple linear regression and mixed models
• Communicated data needs and concerns, project progress, and evaluation results with stakeholders across non-profits, government, and community

CUNY Institute for Computer Simulation, Stochastic Modeling, and Optimization
Assistant Simulation Developer
August 2016 - January 2017 (6 months)
Greater New York City Area

• Troubleshooted and reworked discrete event-based simulation Java program of NYC public bike system for research proposal on testing system self-awareness on system efficiency and customer satisfaction
• Analyzed and converted Citi Bike data into statistical distributions to randomly simulate customer arrivals, travel paths, and travel time

Carnegie Mellon University
Statistics Research Assistant
June 2016 - July 2016 (2 months)
Greater Pittsburgh Area

• Collaborated with team of students in analyzing more than 13,000 galaxies and applying model selection techniques (multiple regression, nonparametric models, principle components analysis, k-nearest-neighbors) to calibrate telescope software.

Princeton University
Sociology Quantitative Research Assistant
June 2015 - August 2015 (3 months)
Princeton, New Jersey

•Analyzed quantitative data in R to test hypothesis that local and central state capacity vary within provincial regions in Colombia.
• Interpreted and situated findings within geographic, historical, demographic, and conflict contexts.

Hunter College
SEEK Tutor
July 2014 - December 2014 (6 months)

• Tutored students in introductory statistics, calculus, and computer science in about a hundred sessions in six months.
• Awarded Tutor of the Month during first month of work.

Education

Harvard University
Master of Science - MS, Data Science · (2019 - 2021)

Hunter College
Bachelor's degree, Statistics, Sociology, & Computer Science · (2013 - 2016)