Emily Natasha Diaz Badilla

Boston, MA • 857.370.2514 • diazbadilla.e@northeastern.edu • https://www.linkedin.com/in/emily-diaz-badilla/
• https://github.com/emsdiaz • https://emsdiaz.github.io/

Education

MS, Data Analytics Engineering
Northeastern University, Boston, USA **May 2023**

- GPA: 4.0
- Fulbright Scholarship from the U.S. Department of State
- Courses taken: Foundations of Data Analytics, Deterministic Operations Research, Computation and Visualization for Analytics, Data Management for Analytics

Bachelor of Science, Statistics
University of Costa Rica, San Jose, Costa Rica **February 2017**

- Average score 9.3 out of 10
- Graduated with honors

Work experience

Walmart, CA, USA

Senior Data Scientist Summer Intern **May 2022 - Present**

- Developing a Proof-of-Concept optimization approach for products in End-of-Life (EOL) stage to optimally place the inventory for the last weeks from the Distribution Center to stores with a multi-site newsvendor model coded in python. This will improve the inventory placement for the products that transition to EOL stage every year, which is ~20% of total products

Northeastern University, MA, USA

Graduate Teacher Assistant for ENGR5964 Projects for Professionals **October 2021 - December 2021**

- This course is centered on a electric consumption dataset provided by Sense, a company that works on energy efficiency and home automation by using several sensors at home and providing feedback to their users.
- Developed LSTM, gradient boosting and traditional time series models to compare accuracy to predict daily home consumption for Sense clients by leveraging their historical consumption data and public weather, charging stations datasets.
- Created and instructed weekly tutorials on time series analysis to ~15 students

Mckinsey & Company, Costa Rica

Specialist, Data Scientist **June 2020 - June 2021**

- Co-designed and executed end-to-end analytics pipelines for top players in the Oil & Energy industry within North America using Python and Kedro as well as storing and collaborating to the code with Github
- Improved demand forecasting models at a client by implementing Prophet time series algorithm in Python
- Managed a team of 2 data scientists over a predictive maintenance project for a key player in the coal energy industry. We developed 30+ models for different components of the energy generation process and different plants.
- Built a team of 5 data scientists from scratch as part of creating an Innovation Center in one of the main Banks in Ecuador. On charge of on-boarding process and served as a technical leader at their first 2 projects: Credit scoring model and Best product to Buy algorithm

Senior Fellow, Data Scientist **December 2018 - May 2020**

- Built tree-based models such as XGBoost and Random Forest in Python for income estimation at a Bank in Latin America
- Co-developed pricing models for a CPG client using LightGBM deployed on Microsoft Azure through. Coded a proof-of-concept R shiny dashboard to monitor the model results and adoption

Analytics Fellow **December 2017 - December 2018**

- Computed Machine Learning models for clients in Latin America, focused on Banking commercial analytics
- Leveraged R for supervised and unsupervised analysis as well as Tableau for data visualization

Junior Research Analyst **January 2017 - December 2017**

- Designed analytical solutions to improve CPG and Banking businesses across Latin America with techniques such as linear regressions, clustering analysis and tree-based models

Analytics Intern **March 2016 - December 2016**

- Reengineered recurrent reporting processes for the Mckinsey Global Institutes Team reducing updating time

Publications

- Devanga, A and Diaz Badilla, E and Dehghanimohammadabadi, M. (2022). Applied Reinforcement Learning for Decision Making in Industrial Simulation Environments. Unpublished manuscript, Department of Mechanical and Industrial Engineering, Northeastern University, Boston, Massachusetts, USA.

- Belsare, S and Diaz Badilla, E and Dehghanimohammadabadi, M. (2022). Reinforcement learning with discrete event simulation: the premise, reality, and promise. Unpublished manuscript, Department of Mechanical and Industrial Engineering, Northeastern University, Boston, Massachusetts, USA.

Technical Skills

- Programming languages: Python, R, SQL, Pyspark
- Analytics and Data Science knowledge: Probability and Statistics, Tree-based models, Generalized Linear Models, Neural Networks, Reinforcement Learning
- ML libraries: scikit-learn, keras, pytorch, pandas, numpy, kedro, prophet, xgboost, simpy, matplotlib, seaborn, nltk, gym, raylib
- Software and platforms: Tableau, Git, Microsoft Azure, Databricks, Alteryx, IBM Watson, Excel, Jira, Confluence, Trello, MongoDB, Neo4j
- Microsoft Certification - Azure Fundamentals (2020)

Volunteering

Co-founder of "*Data Mujer*" – Costa Rica **July 2022 – Present**

- Co-founded a NGO to increase women participation on the data community of Costa Rica

Project Coordinator at "*Costa Rican female data scientists*" – Costa Rica **June 2020 – June 2022**

- Co-developed mentoring program to pair 30 young women with experienced mentors with the objective of helping mentees on STEM career development
- Implemented free preparation course for university admissions tests in Costa Rica with over 500 students from all over the country benefiting from it yearly