

Pedro M. Duarte

Data Engineer

Summary

I embrace challenging and largely unspecified problems, break them down into smaller tasks, and effectively communicate the results of my work.

Contact Information

✉ pmd323@gmail.com
🌐 <https://pedromduarte.github.io>
☎ 8326719721
📍 Berkeley, CA
🐦 pmd323
in <https://www.linkedin.com/in/pmd323>
🔗 pedromduarte

Skills

PROGRAMMING LANGUAGES:

Python, C#, C/C++, SQL, HTML, JavaScript, R

FRAMEWORKS/LIBRARIES/TOOLS:

NumPy, pandas, scikit-learn, GNU Scientific Library, Git, Linux, bash, awk/sed, Airflow, Hive, Spark, MongoDB, D3.js

Publications

Compressibility of a fermionic Mott insulator of ultracold atoms 2015
Physical Review Letters 114 (7), 070403

Observation of anti-ferromagnetic correlations in the Hubbard model with ultracold atoms. 2015
Nature 519, 211-214

All-Optical Production of a Lithium Quantum Gas Using Narrow-Line Laser Cooling 2011
Physical Review A 84, 061406(R)

Projects

Prudential Life Insurance Assessment - Kaggle [link]

Technologies used: Python, scikit-learn, xgboost, R.

- Built model to predict life insurance risk category using customer information. Scored in the top 15%.

Dask: Open source parallel computing library for data analytics [link]

Technologies used: Python

- Contributed in the early stages of the project. Helped with pickling bugs on Windows and generic tests for different schedulers.

Predicting Boston House Prices [link]

Technologies used: Python, scikit-learn.

- Implemented a decision tree regression to predict selling price of new homes using the Boston Housing data set.

Build a Student Intervention System [link]

Technologies used: Python, scikit-learn.

- Use classification to predict how likely a student is to pass their high school final exam.

Experience

AltX

Data Engineer

San Francisco, CA

May 2016 to Current

- Built Apache Airflow based ETL pipeline for AltX's hedge fund market intelligence platform.

Intel Corporation

Process Engineer

Hillsboro, OR

Jan 2015 to May 2016

- Developed analysis and visualization application in C# to help engineers disposition lots. Improved average decision time by 2x.
- Wrote visualization of overlay distortion data in D3.js and wrote a C# application to calculate distortion control adjustments to be applied to the processing equipment.
- Assisted and trained more than 20 new hires in the use of open source software tools (Anaconda, Jupyter Notebook, matplotlib, etc.) for use in ad-hoc queries, data analysis and visualization.

Rice University

Graduate Research Assistant

Houston, TX

Aug 2007 to Dec 2014

- Implemented control, data acquisition and data analysis framework for an ultracold atom experimental apparatus using Python, LabView, and C++.
- Wrote data analysis routines in C/C++ that allowed replacing nightly batch jobs with real time analysis for immediate feedback.
- Implemented principal component analysis algorithm to remove background noise in fluorescence microscopy images of ultracold atom clouds.

University of Texas at Arlington

Graduate Research Assistant

Arlington, TX

Aug 2005 to Aug 2007

- Designed and built a state-of-the-art proton timing detector. Published results in peer reviewed journal.
- Designed and implemented data acquisition and data analysis framework used for test runs at Fermi National Accelerator Lab.
- Wrote Monte Carlo simulation code in Python and C++ to understand detector performance.

Education

Udacity - Machine Learning Engineer Nanodegree - In progress

Rice University - Ph.D. Physics - 2015

H. A. Wilson Research Award, for best Ph.D. thesis in the Physics Department.

University of Texas at Arlington- M.Sc. Physics - 2007

Universidad de Los Andes - B.Sc. Physics - 2005