Jay Gondin

Data Scientist

+1 (408) 757 7817
gondin@gmail.com
gondin
jgondin

Profile

Description: Data Scientist with an advanced degree in Mathematics, and a background in economic analysis and programming.

Skills: Python (pandas, scikit-learn), hive, spark, R, NoSQL, SQL, Linux, Machine Learning, Stats, Math, Portuguese

Education

2008.02-2010.02 **M.S. in Mathematics**, *Pontifícia Universidade Catôlica do Rio de Janeiro*, (Puc-Rio).
Adviser: Paul Schweitzer, Ph.D. | **Project**: Classification of 3d-geometries.

2002.02-2006.01 **B.S. in Mathematics**, *Universidade Estadual de Montes Claros*, (Unimontes).
Adviser: Antonio Wilson, Ph.D. | **Project**: 3D Visualization software for Curve and Surfaces.

Professional Experience

2016.04-2016.07 **Data Science Fellow**, *Metis*, San Francisco CA.

- Intensive Data Science training on data manipulation, statistics, and machine learning.
- Tools: AWS, SQL, NoSQL, python (numpy, scipy, pandas, scikit-learn), d3.js, spark, hadoop.

2013.09-2015.03 **Economic Analyst**, *Getulio Vargas Foundation*, Rio de Janeiro, Brazil.

- Optimized cost/efficiency of ongoing analysis, and published reports and technical papers.
- Improved seasonal adjustment system and evaluated it against competitor's product.
- Redefined sentiment indexes according with the new Economic Activities classifications.
- Tool: Time Series, Economic/Finance Data, Sentiment Analysis, R, Python, SQL.

2007.08–2013.09 **Mathematics Lecturer**, *Federal University of Rio de Janeiro*, (UFRJ), Courses: Calculus, Linear algebra, Complex Analysis and Real Analysis.

2004.09-2005.12 **Programmer**, *Mathematics & Computer Science Partnership – Unimontes*, Montes Claros.
Developed software to visualize 2D/3D curves and surfaces accompanied by a friendly GUI.
Tools: C/C++, Linear Algebra, Software development, Differential Geometry.

Data Science Projects

Email Sherlock [source]
Analyzed and clustered large email datasets in order to identify most relevant email.
Tools: NLP, spaCy, D3.js, XGboost, Random Forest, MongoDB.

Writing Identifier
Identified email authors in Enron public dataset using only the email body.
Tools: NLP, spaCy, D3.js, XGboost, Random Forest, MongoDB.

Predict Water Pump Failure [source]
Predicted the failure of water points throughout Tanzania and identified main features that contribute to pump failure. Top 0.6% on the competition hosted by drivendata.org.
Tools: Pandas, scikit-learn (Random Forest, SVM, XGboost), Flask and AWS.

Predict Film Industry Revenue by Country
Discovered a novel relationship between movie revenue and countries' characteristics.
Tools: Web-Scraping, Pandas, scikit-learn and Regression.

MTA & Night [source]
Identified busiest subway station at New York at specific time frames.
Tools: Advanced pandas, Makefile, Data Cleansing, Outliers Detection, Tableau.