Robertson Wang

Education

2008–2012 **BSc, Economics**, *New York University*.
Magna Cum Laude

2015–2016 **MSc, Economics**, *Universitat Pompeu Fabra*.
Fundación Ramón Arece Scholarship

2017 **Data Science Certificate**, *Georgetown University*.

Summary

My primary stack is comprised of Python, SQL, and R. My main interests are in the applications of machine learning in the prediction of human behavior. In my current position, I have developed an end-to-end machine learning pipeline, written ETLs to source external data, annotated existing data, and written external reports posts on the intersection of data science and charitable giving.

Day to day I use a variety of statistical modeling techniques and machine learning algorithms to assess a wide range of different data sources from customer care to lead generation. I have a background in Finance and Economics with a special interest in time series and regression analysis.

Experience

2018–Present **Senior Data Scientist**, *Classy*, San Diego, CA.
Agile Practitioner mainly using Python, R, and SQL
- Developed text processing pipeline which processes raw text using standard text cleaning methods and produces naive sentiment features
- Worked on models that incorporated Word Embeddings, LDA, and PCA-based dimensionality reduction techniques for unstructured textual data
- Developed a machine learning pipeline that generates predictions in discrete testable steps so that models can be used reliably in production
- Developed a Random Forest lead scoring model to be used by the Sales team
- Developed classification models to predict donor retention and charitable campaign performance
- Developed and maintained ETLs to get data into a Redshift data warehouse
- Wrote annotations for existing data stores and data quality tests
- Wrote research reports and blog posts on online fundraising and charitable giving
- Managed two data science interns, providing guidance on summer projects related to NLP-based clustering and predicting fundraising readiness
- Performed analysis using inferential statistics to guide business decisions in customer care, identifying up-sell opportunities, and forecasting revenue

2016–Present **Data Science Consultant**, *Self*, Remote.

Machine Learning and Data Science Consultant

- Led machine learning projects from model design to implementation
- Developed recommendation system for songs and artists for a music streaming start up
- Implemented MatrixFactorizer and multilayer perceptron networks for a collaborative filtering model
- Maintained and rolled out recommendation engines for mobile application using AWS
- Developed experiments and tests to determine algorithm performance, training time, and to assess user response in product
- Developed data visualizations and tracking dashboards in R Shiny for autism related healthcare research
- Developed interactive Plotly dashboards for healthcare researchers
- Assessed code quality and provided feature engineering guidance for small business credit risk models

2016–2018 **Senior Research Assistant, Risk Analysis**, *Federal Reserve Board of Governors*, Washington, DC.

Python & R Developer

- Co-Authored an academic paper using sentiment analysis to capture settings in which management communications have greater credibility
- Wrote web-scrappers in Python using BeautifulSoup, lxml, and selenium to acquire corpus of financial documents
- Managed large unstructured datasets (1 tb+) on Hadoop, wrote BASH scripts and cron jobs to streamline data ingestion, cleaning, and modeling
- Replicated NLP-based measures, built using LDA and word embeddings, to predict emerging financial risk
- Produced web-scrapping packages in order to perform on-the-fly NLP analysis and classification of financial press releases
- Performed entity resolution and data cleaning on time series CDS positions data (10+gb/day) in order to produce quarterly risk profiles on FRB regulated banks
- Wrote Python and R code to work with the SLURM cluster and Apache Spark in order to parallelize production work, reducing research overhead across teams

2012–2015 **Financial Associate, Valuations & Modeling**, *Morgan Stanley*, New York.

R Developer & Master of Excel Hell

- Crafted various linear regression models for asset pricing. Regression model types included OLS, LASSO, and Ridge
- Developed and maintained forecasting models to price mortgage and credit derivatives
- Wrote code for a high performance cluster in order to price securities using Monte-Carlo simulation techniques
- Created quantitative models to value securitized products (MBS, ABS, MBS Derivatives)
- Led methodology development on a option pricing model, based on Black-Scholes, to price a bespoke REIT IPO
- Designed database schemas, ETL procedures, and standardized data reports using SQL Server and PostgreSQL
- Automated reporting tasks using Python, VBA, and SQL
- Wrote research reports on macroeconomic risk, local job market/housing market conditions, and real estate specific background
- Worked with external auditors and firm counter parties to communicate and defend parameter estimations and model assumptions
- Managed and provided mentorship to a team of financial analysts and summer interns

Blog Publications

Title *"Election Effect" Drives Unprecedented Increase in Recurring Donations*
https://www.classy.org/blog/election-effect-drives-increase-recurring-donations/

Title *Why DAFs Pose Big Problems for Nonprofits and Social Progress*
https://www.classy.org/blog/dafs-pose-problems-nonprofits-social-progress/

Capstone Project

Title *A NLP Restaurant Recommendation System Based On Yelp Reviews*

We used a combination of LDA and tf-idf features trained on user reviews to recommend new restaurants to users. For each user, we test three different classification algorithms and select the best performing machine learning algorithm. We used XGBoost, Random Forest, and linear SVM.

Technical Skills

Statistical Packages	Matlab and Stata	*Data Stores*	Redshift, MySQL, PostgreSQL, and NoSQL
Programming	Python and R	*Big Data*	Apache Spark, Hadoop, SLURM
Markup	LaTeX, XML, MarkDown, YML	*Office*	Excel, PowerPoint, Word

Relevant Coursework

- Machine Learning
- Probability Theory
- Time Series Analysis
- Data Visualization
- Regression Models & Panel Data
- State Space & Bayesian Models

Interests

Rock Climbing, Modern Literature, and Linguistics