

DAVID KARAPETYAN

davidkarapetyan.com

david.karapetyan@gmail.com

SUMMARY

Machine learning scientist and applied mathematician with four years industry experience in machine learning, algorithms, and model validation. Full-stack experience designing a time series forecasting and analytics engine from scratch using Python and C++. One year managerial experience overseeing a team of three machine learning engineers. Two years client-facing experience on a team of quantitative analysts and machine-learning experts.

EDUCATION

University of Notre Dame

- **Ph.D**, Mathematics.
- Thesis: *Non-uniform dependence and well-posedness for the hyperelastic rod equation*
- Awarded the **Schmitt Fellowship**. Full scholarship.

2007–2012

Notre Dame, IN

University of California, Berkeley

- **B.S.**, Mathematics and **B.A.**, English Literature.
- Awarded the **Regents Scholarship**. Full scholarship.

2000–2004

Berkeley, CA

EXPERIENCE

Principal Data Scientist (Remote Contract)

BigR

July 2017–Present

Concord, MA

- Built deep learning architecture using **Keras** and **TensorFlow** in order to forecast api count data for a client, and to perform **anomaly detection** for a highly oscillatory signal.
- Architecture utilizes an **LSTM recurrent neural network** with a rolling window of **wavelet-denoised** time series samples, and **autoencoder** layers for feature engineering before the LSTM runs.
- Work accommodates batch feeds of real time data, and is coupled with a bokeh server dashboard developed in tandem with colleagues, allowing the client to view results and change engine settings with ease.
- Active research and implementation of an internal **image and video anomaly detection** engine, using gradient-based histograms as features to construct a basis dictionary for the samples, and then reconstructing the samples from a pruned version of the dictionary.
- Currently implementing a relational tagger for sentences, which for two arbitrary words in a sentence, identifies their relationship to one another (for example, “mother” to “son”, or “captain” to “ship”) Have built two approaches—one which uses a **convolutional neural net** that relies only on a single feature of sentence position of a word, and the other utilizing **tree-structured LSTM** that is trained by being fed features from a **NLTK semantic parsing** of a corpus of sentences.
- Made sales presentations to potential and current clients outlining machine learning expertise and deliverables

Machine Learning Consultant (Contract)

MUFG Union Bank

April 2017–August 2017

New York, NY

- Reviewed and correct machine learning model usage and implementations in a variety of credit risk and pricing products, including currency options, customer level PD, and others.
- In case of original model incorrectness or inefficiency, debugged or implemented own version in order to provide proof of original model inefficiency by providing a benchmark.
- Corrections included proper **hyperparameter tuning**, appropriate metric choices for evaluating model predictive power and risk, speed optimizations (parallel runs instead of single core runs), proper parameter choices (via **feature engineering** and **feature pruning**), and others.
- Principal models considered were Random Forests, Gradient Boosted Trees, SVM, and logistic regression at the end of the analysis for interpretability.

Data Science Consultant (Contract)

Good Shepherd

Sep 2016–January 2017

New York, NY

- Implemented an application that ingests data from a client and analyses primary factors contributing to student success and failure in college access and support programs.
- Application extracts the most important features in rolling datasets via recursive feature elimination using a series of Random Forests cross-validated by ROC-AUC score.

- Built interactive visual analytics for the client to help identify the types of students that stay in the program, and who go on to stay in college, after the critical features have been identified.
- Application predicts probability that new students would go on to graduate, given data on the student's background, and identified primary factors of success and failure for each individual student. Models used were **lasso regression**, **random forests**, and **logistic regression**, with confidence intervals constructed either via bootstrapping or tests using **statsmodels**.

Head Machine Learning Engineer

Prescriptive Data

June 2015–Sep 2016

New York, NY

- Designed and implemented a **machine learning forecasting and analytics engine** for Rudin commercial high-rise buildings.
- Improved upon the startup and rampdown classification accuracy of the previous existing model, as well as the regression predictions of electricity usage, water usage, steam usage, and occupancy by reducing the mean generalization errors and variance of the errors.
- Implemented using **parallel programming in Python** with an emphasis on **functional programming**, in order to process and predict states for many buildings at once.
- Set **Dask** and **Pandas** Dataframes as the central data-structures of the suite (almost all functions return a dataframe). Generated using queries to local **SQL**, **MongoDB** databases, and **HDF5** files.
- Applied **Scikit-learn** in order to build an ensemble model, where cross-validated scores between different model predictions was used to prune models from the ensemble over time. Experience with **Random Forests**, **Gradient Boosted Trees**, **Regression**, **SARIMAX**, and **SVM**.
- Results include less heat being used in the winter, and cooling in summer, amounting to roughly \$2,000,000 – \$3,000,000 in savings, per building, for Rudin.

Quantitative Analyst

Ernst & Young

June 2014–June 2015

New York, NY

- Developed Class Model forecasting module in **R**. Used **ARIMA regression on macroeconomic scenarios** (base, adverse, or severely adverse) and position data to forecast and plot any input bank's PPNR, Provision, Capital and other variables with respect to time.
- Provided valuation and advanced financial modeling expertise to institutional clients in regards to complex securities including equity and foreign exchange options, rates swaptions, and related embedded derivative instruments.
- Analyzed **Monte Carlo** and **Finite Difference models** to determine fair value of client instruments for accounting purposes.
- Designed and performed **stress-tests** for investment bank client's pricing models for CCAR purposes. Evaluated the impact on PV and option Greeks of client's portfolio of equity and foreign exchange exotic instruments under adverse and severely adverse market scenarios.
- Provided data analysis of trade desk definitions and descriptions, and reported anomalies to client. Trades included forex USD and G10 pairs, G10 and emerging market pairs, trades with long and short expiry, Asian options, barriers, and a variety of others.

Visiting Assistant Professor

University of Rochester

July 2012–June 2014

Rochester, NY

- Researcher of partial differential equations, with an emphasis on nonlinear evolution equations.
- Taught courses on Numerical Analysis, Linear Algebra, Differential Equations, and Financial Mathematics.
- Developed **numerical simulations in C++** to gain intuition about whether certain equations are well-posed or ill-posed for rough initial data.
- Publications list available at davidkarapetyan.com/mathematics.

Research Associate

Battelle

May 2006–May 2007

Arlington, VA

- Conducted research for Pentagon sponsored projects in biowarfare.
- Specifically, researched the analytic hierarchy process and used it to devise a **classification** system that ranked biochemicals by overall tactical effectiveness, given data on their features.

Research Associate

Institute for Defense Analyses

Aug 2004–Aug 2005

Alexandria, VA

- Conducted research for Pentagon sponsored and privately sponsored projects.
- Applied **k-Nearest Neighbor clustering** and **logistic regression** to analyze existing data on high occupancy toll lanes in the Los Angeles, San Diego, and Chicago metro areas.

TECHNICAL SKILLS

- Python (full SciPy stack, Flask), R, C/C++, SQL, \LaTeX , Git, MongoDB, HDF5, Debian/Ubuntu, FreeBSD

AWARDS AND EXTRAS

- Chess Expert chessdryad.com/articles/mi/article_165.htm
- Music composition davidkarapetyan.com/music
- Awarded the **Regents Scholarship**. Full scholarship at UC Berkeley.
- Awarded the **Schmitt Fellowship**. Full scholarship at Notre Dame.
- Finalist for Shaheen award for top graduate student at Notre Dame.