Kevin R. Pan

Data analyst with experience in machine learning.

San Francisco
(860) 941-2128
kevinrpan@gmail.com
kevinrpan.com

EXPERIENCE

The Brattle Group, San Francisco, CA — *Research Analyst*

July 2016 - Present

Worked with Nobel laureate to develop a Poisson regression model, linking prescriptions to doctors who received kickbacks, and sped up training time 10x.

Handled messy data (10s of GBs) for statistical analysis using R and SQL.

Led office-wide seminar on interactive data visualization libraries in R.

Tudor Investment Corporation, Greenwich, CT — *Quantitative Analytics Intern*

June 2015 - August 2015

Optimized bond price curve estimation for 30x speed while maintaining model accuracy using scipy and pandas.

Created flexible classes to calculate swap instrument pricing using Python.

National Bureau of Economic Research / The Wharton School, Philadelphia, PA — *Research Assistant*

March 2014 - August 2014

Scraped and cleaned economic forecast data using Java and R to analyze prediction error trends over time.

EDUCATION

University of Pennsylvania, Philadelphia, PA — *Bachelor of Arts*

September 2012 - May 2016

BA in Mathematical Economics, Minors in Computer Science and Statistics

PROJECTS

Delta Analytics — *Data Fellow and Teaching Fellow*

January 2017 - Current

Provided regression tree analysis of BUILD's program elements to inform data-driven changes to the curriculum.

Visualized and applied machine learning methods to Kiva data using R and Jupyter notebooks for an open machine learning curriculum on GitHub.

Bernoulli One — *Wharton Digital Health Club*

January 2016 - April 2016

Created a model to predict health status drops, with a lead on the industry standard by 12 seconds, using neural networks, PCA and ARMA modeling in R.

Presented an analysis of false positive alerts and health alarm sensitivities to reduce nurse and doctor alarm fatigue to executive team.

SKILLS

Machine Learning:
Classification, clustering, feature engineering

Statistical Methods:
Regression models, time series, hypothesis testing and confidence intervals, PCA and dimensionality reduction

Data Visualization:
Network graphs, custom maps, interactive plots

TECHNOLOGIES

Proficient with
R: tidyverse, data.table, shiny, ggplot2, leaflet, networkD3, plotly, sparklyr

Python: pandas, scikit-learn, numpy

SQL, **Excel**