VIKTOR SHAUMANN

Berkeley, CA ▪ (818) 455-7038 ▪ viktor.shaumann@gmail.com ▪ GitHub: vshaumann ▪ LinkedIn: viktorshaumann

EDUCATION

UNIVERSITY OF SAN FRANCISCO San Francisco, CA Jul 2016
Master of Science in Analytics

- *Academics*: Machine Learning, Distributed Computing, Relational & NoSQL Databases, Design of Experiments, Web Analytics, Advanced Data Visualization, Linear Regression, Multivariate Statistics, Time Series, Data Acquisition, Business Strategies for Big Data, Computation for Analytics and Exploratory Data Analysis

UNIVERSITY OF CALIFORNIA, SAN DIEGO San Diego, CA Mar 2013
Bachelor of Arts in Economics, Minors in Business and Accounting

- *Academics*: Econometrics, Linear Algebra, Python Programming, Micro & Macro Economics, Calculus for Science & Engineering, Business Strategy, Project Management, Product Marketing, Organizational Leadership and Enterprise Finance

EXPERIENCE

AT&T San Francisco, CA Nov 2015 – Present
Data Scientist Intern

- Designed and implemented a distributed and parallelizable community detection algorithm on the AT&T to AT&T call network
- Developed statistical validation procedures and performed hypothesis testing for graph models
- Processed over 100 billion of raw records using Pig and Python UDFs into directed and undirected network graphs
- Performed EDA and calculated self-defined historical customer churn metrics using Hive SQL
- Utilized Agile methodology to plan deliverables, created documentation and communicated results to the stakeholders

BDO CONSULTING – BUSINESS ANALYTICS GROUP Los Angeles, CA May 2013 – Jun 2015
Analyst

- Designed integrated financial models to value entities, complex financial securities, intangible and fixed assets
- Analyses included discounted cash flow analysis, option pricing model (Black-Scholes) and Monte Carlo simulations
- Identified patterns and trends in financial and operational data, performed company, industry and economic research
- Valued over 50 venture capital backed leading technology start-ups in the Silicon Valley and San Francisco
- Experienced in internet, software, technology, media, semiconductor, consumer products and pharmaceuticals industries

TECHNICAL SKILLS

- *Languages*: Python, R, SQL, NoSQL (PostgreSQL, MongoDB, Apache Drill)
- *Big Data*: pySpark, Pig, Hive
- *Visualization*: ggplot2 (R), matplotlib, seaborn (Python), D3.js (Javascript)
- *Other*: Jupyter/IPython Notebook, UNIX, bash, API (YouTube, Yelp, Facebook and Twitter), AWS (EMR, EC2, RDS)
- *Business Software*: Excel, PowerPoint, Tableau, Rally, Asana

PROJECTS

- *Amazon Review Helpfulness Prediction (2016)*: Predicted usefulness measure and sentiment based on the context of 4.6 million of reviews. Feature extraction with Word2vec, modeling with Logistic Regression and Random Forest. Application built in Flask
- *Kaggle - What's Cooking? (2016)*: Achieved 77.0% accuracy (vs. 82% first place) with a neural network built using Tensor Flow
- *Restful Web Service (2015)*: Built an analytics web server (EC2) with a Flask backend and PostgreSQL database (RDS). Performed ETL of over 2 million Amazon product reviews
- *Time Series Case Study (2015)*: Forecasted national bankruptcy rates in R by utilizing ARIMA, SARIMA and GARCH models
- *Map Reduce Anagrams (2015)*: Developed a MapReduce algorithm in Python to parse text and identify anagrams using a single-node Hadoop cluster
- *YouTube Trend Analysis (2015)*: Scraped YouTube videos trending worldwide using Python and Google API. Utilized Pandas and PostgreSQL to store and query the data. Visualized the data with Plotly to identify trends and relationships
- *Naïve Bayes Sentiment Analysis (2015)*: Achieved 80% accuracy on a test set. Classified 2,000 movie reviews from polarity data set as positive or negative. Computed probabilities of each class using algorithm written in Python
- *Los Angeles Crime Analysis (2014)*: Identified relationship between violent crimes and temperature by analyzing and visualizing LAPD dataset with 234,151 crime records using IPython Notebook and Pandas