SHANSHAN DING

(440)463-2990 • New York, NY
shanshan.ding08@gmail.com • shanshanding.github.io

SUMMARY

Data scientist and mathematician with four years of tech industry experience in a wide range of functions including predictive modeling, content discovery, NLP, data warehousing, and product analytics.

EDUCATION

University of Pennsylvania Philadelphia, PA
Ph.D. in Mathematics *2008 - 2014*

- Concentration: Probability and Combinatorics
- Honors: Benjamin Franklin Fellowship, Penn Presidential Fellowship, Award of Excellence in Teaching

Columbia University New York, NY
B.A in Mathematics and Economics *2004 - 2008*

EXPERIENCE

Compass New York, NY
Senior Data Scientist *July 2017 - October 2017*
Data Scientist *January 2016 - June 2017*

- Developed "Similar Listings", a real estate recommender system, as project lead in both data science and back-end engineering (tech stack includes Python, Thrift, and PostgreSQL).
- Created valuation model for real estate listings.
- Using NLP techniques (with NLTK and gensim libraries), prototyped automatic extraction of structured listings data from free-form text descriptions.
- Built the company's reporting and analytics functions almost from scratch, including the schema design and ETL of a data warehouse and an A/B testing platform.
- Collaborated with product team to define KPIs and to assess the progress thereof; also, to propose and execute product analytics projects such as user segmentation.
- Mentored engineering, product, and business development teams in building a data-fluent culture.

Gilt Groupe New York, NY
Data Scientist II *August 2015 - January 2016*

- Built profit-maximizing pricing model by predicting the demand curve of each product.
- Led end-to-end investigation of A/B testing methodology and platform.
- Supported and implemented new features in LookML for inventory optimization tool.

YP Mobile Labs New York, NY
Data Scientist *June 2014 - August 2015*

- Automated pacing control for DSP mobile advertisement campaigns (production code written in R).
- Generated lift in campaign performances by improving conversion rate predictions.
- Designed and implemented anomaly detection to monitor for possible disruptions in the RTB system.
- Analyzed quality of mobile location data obtained from ad exchanges and collaborated with marketing team to communicate findings in white papers.
- Correlated spatial time series derived from mobility data with macroeconomic trends.

EXPERIENCE, continued

Insight Data Science — Mountain View, CA
Data Science Fellow — *August 2013 - October 2013*

- Created Python-based web app to optimize fantasy baseball lineups.
- Applied k-means clustering to predict outcomes of pitcher-batter matchups.
- Built front end with Twitter Bootstrap and JavaScript, and used Flask to deploy app to AWS.

Columbia Business School — New York, NY
Research Intern — *June 2013 - July 2013*

- Analyzed nine years of European and American professional soccer play-by-play data to extract patterns in styles of play, in-game strategies, and player development.
- Debugged and streamlined existing codebase in Python, R, and Stata.

OTHER ACTIVITIES

New York University — New York, NY
Adjunct Instructor — *January 2015 - April 2015*

- Gave a series of guest lectures for DS-GA 1003 (Machine Learning and Computational Statistics) on mathematical foundations of machine learning.

University of Pennsylvania — Philadelphia, PA
Teaching Fellow — *2009 - 2013*

- Piloted pre-freshman math boot camp for incoming undergrads of disadvantaged backgrounds.
- Taught multivariable calculus as primary instructor and served as teaching assistant for eight other math classes ranging from introductory calculus to graduate-level topology.
- Led annual teaching training program for new graduate students and postdocs.

TECHNICAL SKILLS

Programming Languages	Python, R, familiarity with Java and JavaScript
Databases	PostgreSQL, Redshift, MongoDB
Big Data Frameworks	Hadoop streaming, Spark
Other Technologies	HTML, CSS, YAML, Thrift
Tools	Vim, Tmux, Git, Gerrit
BI Softwares	Periscope, Looker, Tableau
Machine Learning	linear regression, hierarchical linear model, logistic regression, random forest, boosted decision trees, naive Bayes, SVM, k-means clustering, Gaussian mixture model, SVD/PCA, tf-idf, LDA, word2vec
Statistics	hypothesis testing, Bayesian inference

PERSONAL

Citizenship	USA
Natural Languages	English (native), Chinese (fluent), French (intermediate), German (basic)
Interests	traveling, photography, distance running, choir, fantasy baseball