Andrew Young

Résumé

Los Angeles, CA
+1 (808) 675 6227
andcy7@gmail.com

Education

12/16 **M.S. Statistics,** *Duke University*

- *Interests*: Machine learning, computer vision, data visualization, predictive modeling
- *Projects*:
  - *Convolution Neural Network:* Yelp photo classification
  - *SVM:* convex hulls of police precincts from NYC parking citation data
  - *Computer Vision:* real-time face and eye detection
  - *Consulting projects:* hypothermia treatment evaluation, Saudi Arabia's nursing system
- *Teaching Experience*: Stat101: Data Analysis/Inference (Head TA); Math342D: Statistics; Top 2 TA 2016
- *Professional Membership*: Society for Industrial and Applied Mathematics (SIAM)

06/11 **B.A. Business Economics,** *University of California, Irvine*

B.A. International Studies with Honors, *University of California, Irvine*

- Honors thesis, *"International Shocks on the U.S. Economy from 1920-1939"*
- Statistics thesis, *"Fitting Generalized Linear Models with Log-Link Functions to NBA Data"*
- Top 2% GRE score (07/2011); awarded two research grants (Spring 2010, Fall 2010) : $1,300.
- Top 2% Best Buy Scholarship Contestant: $1,000
- Student government leader, steward of $6.58MM budget, 86 staff, and school newspaper writer (more)
- Physics Instructor. ODE Econ Honor Society. National Society of Collegiate Scholars

Professional Experience

08/16–Now **R&D Data Scientist, Manager,** *Neustar,* Los Angeles, CA

- Technical
  - Migrate data from internal computing cluster to Amazon Web Services (AWS) S3 and set-up EC2 & EMR clusters.
  - Teach R&D team how to use AWS Athena query service and other AWS services.
  - Build machine learning models like random forest, logistic, GBT, Lloyd's algorithm/K-means||, etc.
  - Engineer features using domain knowledge of the data, which involves brainstorming, constructing & testing.
  - Translate C++ and Java production code into Python for a baseline comparison with ML models.
  - Code in PySpark, SparkR, Python, R and Jupyter notebooks to discover actionable insights.
  - Execute queries on both Spark and Hadoop MapReduce computation engines, depending on the data volume.
  - Create Hive tables from terabytes of raw data and write efficient HiveQL to query them on compute clusters.
  - Design and code interactive data visualizations like geospatial temporal maps, Gephi network graphs, etc.
  - Write internal whitepapers in RStudio, RMarkdown, LaTeXfor reproducibility and knowledge sharing.
  - Extract, transform & load (ETL) data on a daily basis from disparate sources, each with varying challenges.
  - Evaluate big data samples from new data providers in varying domains, for possible acquisition.
  - Design and execute digital factorial experiments (i.e. "A/B testing" + "multivariate testing").
- Growth
  - Search for groundbreaking methods and improvements in recent research papers for possible use.
  - Attend academic machine learning conferences like NIPS and ICML.
  - Self-study computer vision ideas, complete self-initiated coding projects and review old coursework.
  - Avid reader of Stack Overflow and data scientist networking event enthusiast.
- Leadership
  - Teach tricks to LA, SF and NYC data science teams to improve efficiency and promote professional growth.
  - Share new interactive data visualization techniques, web technologies and algorithms with R&D team
  - Train over 10 data analysts in the aforementioned technical skills and maintain internal R&D documentation.
  - Lead data science team scrums and JIRA board planning with a minimalist philosophy.
  - Source research projects from within the company through soft-skills and internal networking.
  - Architect plans of attack for research projects and triage requests from multiple senior-level stakeholders.
  - Develop and introduce best practices for coding and collaboration, e.g. git version control software (VCS)
- Results
  - Improved accuracy of browser cookie to physical residential address linkage from ~50% to 98% based on multiple out-of-sample errors. Coefficient weights remain relevant for at least one month after training.
  - Identified the most trustworthy data sources of hundreds, for each field of information we receive.
  - Improve age and gender information appended to data for a service that generates $5M per annum.

01/14–12/15 **Senior Consultant,** *Duke University Statistical Consulting Center,* Durham, NC

- Led a team of Ph.D, Master and advanced Bachelor students to solve complicated problems within 7 days.
- Wrote reports compiling recommendations for problems from varied domains like medicine, ecology & business.

07/11–05/13 **Financial Analyst,** *Esurance, an Allstate company,* San Francisco, CA

- Built financial models, cost-benefit analyses, and scenario analyses for strategic planning.
- Prepared ad hoc dashboards and analyses for C-level executives using SQL, PivotTables and VBA.
- Assessed China and Canada for expansion and valued potential start-ups for acquisition.
- Led stragetic planning of a Fortune 100 company through financial modeling.
- Improved efficiency of workflow processes for faster turnaround in a high-paced environment.

02/10–07/11 **Research Team Leader,** *UCI Department of Economics,* Irvine, CA

- Trained and managed 25 Research Assistants to compile NYC banking data for the 1900s.
- Completed a 12-week project in 3 weeks and developed simple data validation formulas.
- Research assistant for MIT Press published textbook, *Money, Payments, and Liquidity.*

10/10–12/10 **Jr. Analyst,** *Belgravia Capital Investment Bank,* Costa Mesa, CA

- Due diligence for $2.3 billion sale-and-leaseback with California, Morgan Stanley and Bank of America.

10/10–12/10 **Intern,** *UBS Financial Services,* Irvine, CA

- Produced weekly profitability reports and a 60-page book of small-cap, institutional investors in Orange County

Skills

- Hive, Hadoop, RStudio, PyCharm/Rodeo/Jupyter, Excel (with Python), PowerPoint, Word, Tableau, git

Languages

- R (SparkR), Python (PySpark, numPy, pandas, scipy, scikit-learn, nltk, matplotlib), Stata, SQL, HiveQL, LaTeX