MATT SHADISH

San Francisco, CA · (925) 487-1921 · m.shadish@engineering.ucla.edu · GitHub: mshadish

EDUCATION

M.S. Analytics, University of San Francisco   **Expected July 2015**
  Coursework: Machine Learning, Linear Regression, Time Series Analysis, NoSQL Databases, Data Acquisition
  GPA: 3.97

B.S. Mechanical Engineering, UCLA   **December 2013**
  Honors**:** Phi Beta Kappa, Tau Beta Pi
  GPA: 3.75

Certifications: Base Programming for SAS 9

PROJECTS

Kaggle: Driver Telematics Analysis   **January 2015 – Present**
  - Establishing driver fingerprints by extracting and creating features from driver positional data using R
  - Performing anomaly detection across driver data using a bagged ensemble of logistic regression estimators in Python and scikit-learn

Kaggle: Sentiment Analysis on Movie Reviews   **October 2014 – January 2015**
  - Classified movie review sentiment using word vectorization and a support vector machine in Python

Financial Markets Sentiment Extraction   **September 2014 – October 2014**
  - Scraped market opinion sites to understand market sentiment using Python and Beautiful Soup

Eccentric Exercise Machine Prototype   **June 2013 – September 2013**
Founder, Lead Engineer
  - Researched, designed, and fabricated a prototype exercise machine made to facilitate eccentric exercise
  - Drafted and submitted a provisional utility patent application

WORK EXPERIENCE

General Electric, San Ramon, CA   **November 2014 – Present**
Data Analytics Intern
  Parts Sourcing Project
  - Improved search experience through parts database with elasticsearch and Python Flask
  - Implemented and deployed fuzzy matching of part descriptions as a MapReduce job with Hadoop Streaming
  Company-Contact Mapping Project
  - Doubled match rate of company names using Levenshtein distance metrics and custom weightings in Python

Engage3, Davis, CA   **December 2013 – August 2014**
Data Scientist, Category Management Algorithms
  - Managed, designed, and implemented Oracle PL/SQL algorithms to categorize 800M+ retail product records
  - Improved runtime of daily Oracle PL/SQL batch processes by 90% by developing incremental features
  - Modeled regular and promotional pricing among retailers using regression analysis of product prices
  - Collaborated with other members of the Data Science team using TortoiseSVN, Git, and Assembla
  - Began migration to Apache Cassandra with EIS team, leveraging my knowledge of pricing algorithms

TECHNICAL SKILLS

  - Programming Languages
    - Proficient: Python, SQL (Oracle PL/SQL, PostgreSQL, MySQL), R, SAS
    - Familiar: Bash, LaTeX, C++
  - Other technologies: Git, elasticsearch, Hadoop Streaming, MongoDB, Apache Solr