MATT SHADISH

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

EDUCATION

M.S. Analytics, University of San Francisco **July 2015**

Relevant coursework: Machine Learning, Linear Regression, Time Series Analysis, Multivariate Statistics, Databases (RDBMS and NoSQL), Distributed Computing, Data Visualization

GPA: 3.88

B.S. Mechanical Engineering, UCLA **December 2013**

Honors**:** Phi Beta Kappa, Tau Beta Pi

GPA: 3.75

WORK EXPERIENCE

General Electric, San Ramon, CA **November 2014 – March 2015**

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 to customer contacts using Levenshtein distance metrics, custom weightings, and multiprocessing in Python

Engage3, Davis, CA **December 2013 – August 2014**

Data Scientist, Category Management Algorithms

- Designed and prototyped Spark submit scripts to model price frequency distributions in Python
- 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 Data Science team members using Git, TortoiseSVN, and Assembla

PROJECTS

Convergence Investment Management **March 2015 – May 2015**

- Supplemented existing trading strategy with machine learning techniques in Python and R

Kaggle: Driver Telematics Analysis **January 2015 – March 2015**

- Established driver fingerprints by extracting and creating features from driver positional data using R
- Performed anomaly detection across driver data using a combination of a bagged logistic regression estimator and a random forest in Python with scikit-learn
- Scored in the top 30% of contestants out of 1500+ with an AUC of 0.85

Kaggle: Sentiment Analysis of Movie Reviews **November 2014 – January 2015**

- Predicted movie review scores from text data using word vectorization and SVM in Python with scikit-learn

Financial Markets Sentiment Extraction **September 2014 – October 2014**

- Scraped market opinion sites to understand market sentiment using Python and Beautiful Soup

TECHNICAL SKILLS

- Programming Languages
  - Proficient: Python, SQL (Oracle PL/SQL, PostgreSQL, MySQL), R
  - Familiar: Bash, JavaScript (D3.js), SAS, LaTeX, C++
- Other technologies: Git, Hadoop Streaming, Spark, Elasticsearch, Apache Solr, MongoDB
- Modeling: Classification, Regression, Ensembles, Clustering, Anomaly detection, Feature extraction, PCA