

# MATT SHADISH

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

---

## EDUCATION

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

**July 2014 – 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

*Data Analytics Intern*

**November 2014 – March 2015**

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

*Data Scientist (full-time)*

**December 2013 – August 2014**

- Prototyped Apache Spark submit scripts to model price frequency distributions using Python
  - Designed/implemented the data ingestion model for storing auditor data using Python and Oracle PL/SQL
  - 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, SVN, and Assembla
- 

## PROJECTS

**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 ratings from text data using word vectorization and SVM, k-NN in Python with scikit-learn

**Financial Markets Sentiment Extraction**

**September 2014 – October 2014**

- Scraped market opinion sites to understand market sentiment using Python and BeautifulSoup
- 

## TECHNICAL SKILLS

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