

Summary

- Had 4 years of quantitative analytics experience in Environmental Science
- Changed major to Statistics and graduated with GPA 4.0
- Gained hands-on project experience through various online communities, such as Udacity and Kaggle

Skills

- Analytics: SQL, SAS, Minitab, SPSS, Excel
- Programming: Python, R, Octave
- Visualization: Tableau, ggplot2, Matplotlib

Projects

Openstreetmap Data Munging

- Extracted and parsed 3GB of San Francisco bay area geolocation data using Python ElementTree package
- Implemented scripts to programmatically audit and clean data
- Imported cleaned data into a relational database (sqlite) and conducted data explorations using SQL queries

Person of Interest Identifier

- Cleaned financial and email data belong to 146 senior management of Enron Corp using pandas and numpy
- Built models to identify person of interests in Enron Scandal using three classifiers including Naive Bayes, Decision Tree, and SVM
- Tuned each classifier using GridSearchCV and Pipeline in Python sklearn library
- Evaluated each classifier using various metrics including accuracy, specification, recall and F measurements, and found the best model with Decision Tree classifier

A/B Test on Udacity Click Screener

- Identified invariant metrics and evaluation metrics for testing the effect of adding a click screener on Udacity website on reducing the number of clicks while keeping payments at the similar level
- Conducted effect size test and sign test using Excel after data pass the sanity check
- Both tests suggested launching the click screener

Exploratory Data Analysis of Prosper Loan Dataset

- Performed EDA on a dataset containing 113,937 loans information to find the main factors for borrower rate
- Visualized characteristics of variables and their relationships using ggplot2
- Revealed main factors that impact borrower rate including credit scores, homeowner status et al.

Education

MS, Statistics, California State University - East Bay (GPA 4.0/4.0)	Jun 2017
PhD, Soil and Water Science, University of Florida (quit with MS)	Jun 2015
MS, Natural Resources, University of Nebraska - Lincoln	Aug 2013
BS, Environmental Science, Nanjing Agricultural University	Jun 2011

Certifications: Udacity Data Analyst Nanodegree, SAS Advanced Programmer

School Courses: Bayesian Statistics, Probability Theory, Statistical Method in Research, Advanced Statistical Inference, Environmental Statistics, Stochastic Processes, Regression Theory, R for Statistical Learning, Advanced SAS Programing, Mathematical Statistics, Advanced Probability

MCCO Courses: R programming, The Data Scientist's Toolbox, Getting and Cleaning Data, Exploratory Data Analysis, Learn to Program - the Fundamental, Machine Learning.

Experience

Statistics Tutor, Student Academic Achievement Center, CSUEB	Jan 2017 - Jun 2017
Teaching/Research Assistant, Soil and Water Department, UFL	Aug 2013 - Jun 2015
Research Assistant, School of Natural Resources, UNL	Jan 2012 - Aug 2013