David Kuan-Yu Chen

Professional Summary

- Graduate student specializing in data analytics with hands-on software development experience in Scala
- Strong academic background in programming and data science, specifically in data processing/pipelines with Hadoop and HBase
- Experience with data mining in Python and Perl and building predictive models involving machine learning in R and Python

Education

MS, Information Management, Data Analytics Specialization **Anticipated May 2016**
University of Maryland, College Park, Maryland GPA 3.9
Bachelor of Science, Biochemical Science & Technology **June 2008**
National Taiwan University, Taipei, Taiwan

Relevant Professional Experience

Software Engineering/Data Analytics Intern **May 2015-August 2015**

PFP Cybersecurity, Vienna, Virginia

- Implemented a new scalable, web-based, large service, and data analytics architecture involving Hadoop, HBase and HDFS
- Installed and set up Cloudera CDH Hadoop cluster for big data development environment
- Designed HBase schema and implemented the backend interface in Scala
- Built front-end pages in Scala/Play framework
- Automated application deployment and virtual machine image importing on AWS with Python script
- Developed beta product for trade shows and demonstrations successfully

Financial Data Analyst/Futures Trader **September 2009-January 2016**

Self employed

- Analyzed public transaction data involving statistical analysis and patterns through visualizing data with Excel
- Constructed hypotheses and examined hypotheses with trading experiments on Taiwan Weighted Index Futures
- Developed trading strategies based on quantitative findings and qualitative analysis

Selected Projects

Data Anomaly Detection (Capstone Project with PFP Cybersecurity) **August 2015-December 2015**

- Gathered project requirements from a business problem and transformed requirements to an analysis plan
- Performed feature engineering on the raw data
- Ran and evaluated several predictive models, such as SVM, K-mean clustering, regression, for flagging anomaly data

Apply Benford's Law to Detecting Twitter Fake Accounts **August 2015-December 2015**

- Designed and developed a Perl-based application to interact with Twitter public user account information through Twitter API
- Designed and implemented an acquisition process to perform statistical analysis
- Automated the inspection of anomaly data for further analysis with Perl script and R

Predicting the Success of Local Events and Local Groups from MeetUp.com **August 2015-December 2015**

- Wrangled the data and performed descriptive statistical analysis in R
- Extracted more features of interest (feature engineering) in R
- Performed natural language processing on several text-based attributes
- Evaluated the predictive performance of different machine learning models, including SVM, decision tree, random forest, and regression, clustering in R

Restaurant Menu Management Database Design **January 2015-May 2015**

- Designed the schema of a restaurant menu management database in MySQL Workbench, including logical design and physical design
- Wrote several testing queries to examine the design of the schema in MySQL Workbench

Americans' Opinion of Controversial Social Topics **January 2015-May 2015**

- Performed statistical analysis to explore the dataset, including chi-square test, in SPSS
- Analyzed trend change of Americans' opinion on controversial topics, such as homosexual marriage, abortion, and marijuana legalization

Skills

Scala (Play), Python (Pandas, Scikit-learn), Java, R, Perl, MySQL, HBase, Hadoop, JavaScript, NodeJS, AngularJS, PHP, Html/CSS