Weihan Li

517 West 113th Street, 41, New York, NY 10025 | 929-273-9389 | E-mail: wl2525@columbia.edu
https://www.linkedin.com/in/weihan-li-296255105 | https://github.com/Saintat1

EDUCATION

Columbia University, Graduate School of Arts and Sciences **New York, NY**
MA in Statistics (GPA: 3.5/4.0) December 2016

- Relevant Coursework: Data Mining, Algorithms for Data Science, Applied Data Science, Linear Regression Models, Probability and Statistical Inference, Bayesian Statistics, Time Series

University of Science and Technology of China **Hefei, China**
BS in Physics (GPA: 3.6/4.0) June 2015

- Relevant Coursework: Quantum Physics, Thermodynamics and Statistical Physics, Computational Methods, Computational Physics, Computer Programing, Probability and Statistics

WORK EXPERIENCE

Guzman Energy, LLC **Miami, FL**
Intern, Quantitative Analyst Aug 2016 – Dec 2016
Full time, Quantitative Analyst Dec 2016– Now

- Took part in the building of ETRM (Energy Trade and Risk Management) system with JAVA and MySQL. Developed arbitrage models and tools, built P&L panel to show the performance of traders and different strategies
- Developed report system with R and task scheduler, sending daily email reports (like daily risk report, margin call report) automatically to management team and traders, helping to make better decisions
- Built DNP3 (Distributed Network Protocol) in C++ to receive real-time data from outstation, helped trader to trade power product

FXCM **New York, NY**
Intern, Foreign Exchange Trader Aug 2015 – Sept 2015

- Employed technical indicators, risk management and basic trading strategies of forecast trends
- Researched macroeconomic trends for currency pairs, built daily reports on market conditions
- Designed a high-frequency trading program to catch instant changes of the prices in FX market and obtain profits, achieved approximately 30% profit within a month with a $50,000 mocking account

SKILLS

Programming: Python, R, Java, MySQL, JavaScript, HTML, C, C++, Linux
Data Analysis: Statistical models, Bayesian Statistics, Data Mining, Machine Learning, Database, NLP, Data Visualization

PROJECT EXPERIENCE

Amazon Movie Recommendation System **New York, NY**
Columbia University Apr 2016
Project Website: https://weihan.shinyapps.io/movieShiny/

- Scraped data from Amazon website with bs4 package in Python to get review, review helpfulness, score, and time for each movie, cleaned the data and made it into a format of rating matrix using Pandas in Python for later use
- Introduced item-based and user-based collective filtering algorithm into our recommendation system
- Created a movie recommendation system with Shiny App in R Based on collective filtering and co-occurrence matrix, make recommendation for users according to their preference

Image Recognition: Cats and Dogs **New York, NY**
Columbia University Mar 2016

- Extracted features from 7000 pictures of Cats and Dogs with OpenCV-Python, used Edge Detection to find high frequency information within the image to allow SIFT to focus on keypoints of the face of the animal
- Used boxed face detection from original dataset to crop images, utilized SIFT to create Image Vocabulary
- Used Cross Validation with 5 K-fold to test multiple ML algorithms on image vocabulary (Random Forests, Gaussian Naïve Bayes, Quadratic Discriminant Analysis, Neural Networks and SVM) with sklearn in Python
- Found Random Forest to be the best algorithms with an accuracy of 74%, which has an improvement of 18%

President Election Analysis Based on Twitter **New York, NY**
Voxgov Company and Columbia University Jan 2016
Project Website: http://geekelection.netai.net

- Collected raw data with Twitter API in R, cleaned it with tm package to get rid of useless content such as stop words, employed TF-IDF method (with textir package in R) to choose key words
- Conducted sentimental analysis based on the corpus built, summarized characteristics of every candidates' tweets
- Used D3.js to visualize our interesting result, built a website with HTML and CSS to show our article and chart