Wenyi (Wesley) Tao

354W 110 St. New York, NY 10025 Email: wt2271@columbia.edu Portfolio: http://statisticspower.ml/

EDUCATION

Columbia University *M.A. in Statistics, New York* GPA: 4.00 Expected Dec 2018

Fudan University *B.A. in Economics, Shanghai* GPA: 3.53/4.00 Sept.2012 - July.2017

Model: Lasso Ridge, SVM, AdaBoost, Bagging, K-means, Logistic, Hierarchy Bayesian, EM algorithm, lightGBM

Tools: SQL, Python, R, Git, OpenCV, Scikit-learn, D3,Tensorflow, Tableau

WORK EXPERIENCE

Adatos A.I. , New York, *Machine learning engineer Intern* Jun.2018 – Present

- Implemented a deep-learning-powered model for palm tree detection and localization, yield estimation on high-resolution satellite images
- Designed the automated end-to-end data pipeline for the satellite image segmentation, classification, feature importance via python.
- Build a dashboard for internal use to help geospatial analysts learn and use Deep Learning models and Computer Vision algorithms

Fudan Institute of Data Science: Analysis of Electricity User Behavior Under Different Pricing Policies in Shanghai Aug.2017

- Analyzed the electricity meter data from Shanghai Electric Power Company's database and found a gap between two group of users
- Found the non-flat-rate users are developing a habit of consuming more electricity during the non-peak hour period
- Quantified the price elasticity using Hausman Taylor Model and proposed a new pricing policy for the Electric Power Company

Guotai Junan Securities, Shanghai *Intern, Commodities Research* Mar.2016

- Completed a historical correlation analysis of NYMEX crude-oil and domestic palm tree oil price for investment recommendations
- Performed the technical analysis covering three commodity price on a daily bases with data streaming Bloomberg API
- Wrote consistently 78 daily investment reports some of which were published on the Guotai Junan Group's website

PROJECT EXPERIENCE

Data Mining for Target Marketing (1st Place Award) – 2018 American Statistical Association DataFest April.2018

- Discovered the dominant features like license-required, job description that can boost job-matching efficiency using Elastic Net
- Analyzed the current market for Indeed leveraging external data like Social Capital Index to evaluate the potential market

Sentiment Analysis and Topic Modeling with twitter comments on US-China Trade War: April. 2018

- Built a web-base dashboard which visualized twitter users interaction network and filtered the influential opinion leaders on the topic
- Performed a sentiment analysis of millions of tweets and found the majority opinion's shift across time after a series of major events
- Performed the topic analysis using Latent Dirichlet Allocation and generalized 5 different topics for both sides

Image Classification Project on Dog, Muffins and Fried Chickens **Top 1 out of 8 teams** Mar.2018

- Improved a classification baseline model (GBM with decision stumps) with lower running time cost and higher prediction accuracy
- Used three image feature selection models, including SIFT, RGB, ORB with eight classification models
- Implemented GBM, SVM, XgBoost, Random Forest, Neural Network, AdaBoost, Logistic Regression and Classification Trees

Recommendation System: Collaborative Filtering Algorithm with MS Website and Movie Score Dataset April.2018

- Implemented from scratches a recommendation algorithm Bayesian Clustering with a users' website browsing history dataset
- Evaluated different algorithms based on a customers' utility metrics which was also implemented from scratches

Colloquial Trend in President Inauguration Speeches **https://bit.ly/2KDlxv8** Feb.2018

- Designed two scoring metrics to quantify the semantic complexity based on average word choice and its grammatic difficulty
- Performed in-depth analysis of 45 inauguration speeches and found a steady decline using 'educated' words and 'formal' language

Alibaba Cloud Algorithm Competition: Future Challenge Helping Balloons Navigate the Weather **Top 5%** Jan.2018

- Planned safe and fastest flight routes for ten unmanned balloons to deliver packages to their destination
- Creatively implemented from scratches the A-Star algorithm in 3-Dimension space to solve the obstacle changing problem

INTEREST

- Swimming (2000m nonstop with medley)