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: R, Python (Proficient),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 predicting model of palm tree yield (tree detection, tree counting/density estimation, leaf and soil nutrient analysis, fertilization analysis, age estimation and weather/disasters analysis) on satellite imagery to incorporate the signals in palm oil commodity future trading strategy
- Build a web-app(Flask in Python) for internal use to help geospatial analysts learn and use Deep Learning models and Computer Vision algorithms to analyze satellite images (image segmentation, classification, regression, clustering, feature importance analysis, and etc.)

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

- Study the user's behavior under different pricing policies and build a model to quantify price elasticity.
- Found and then visualized the difference between the flat-rate users and non-flat rate users on an hourly basis

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

- Completed a historical correlation analysis of crude-oil and palm tree oil price for investment recommendations
- Wrote 78 daily investment reports with R markdown in response to client's requests

PROJECT EXPERIENCE

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

- Discovered the dominant features that contribute to job-match efficiency by elastic-net logistics regression
- Introduced the Growth-Eagerness matrix along with Social Capital Index to evaluate and visualize potential market

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

- Use interactive D3 to visualize the retweet relation networks and found those influential opinion leaders on this topic
- Visualize majority opinion's shift across a series of events after performing sentiment analysis on all the comments
- Use LDA for topic modeling and found pro and cons argument for both sides in this debate

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) regarding running time cost and 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.

R shiny App: Visualization For Fire Accidents in NYC **https://jz2891.shinyapps.io/fire__rescue_nyc/** Mar.2018

- Build an interactive map using R shiny to visualize the fire incidents location in the New York City
- Create an animation for distribution changes of fire incident type across time implemented by javascript d3

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

- Implement a memory-based strategy Bayesian Clustering with EM algorithm from scratches
- Evaluating different algorithms based on a rank-scoring system

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

- Quantify the difficulty of each speech by grammar complexity and choice of the word
- Found that the presidents inauguration speeches have a steady decline in using "educated" languages across time

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 a-star shortest path algorithm in 3-D space to solve the obstacle changing problem

INTEREST

- Swimming (2000m nonstop with medley)