Jingbin Cao

Phone: (765) 720-7123
Address: 21 W End Ave, New York, NY, 10023
Email: Jingbin.cao@columbia.edu
LinkedIn: https://www.linkedin.com/in/jingbin-cao/
GitHub: https://github.com/JingbinCao

EDUCATION

Columbia University, Graduate School of Arts and Science *New York, NY*
Master of Arts with major in Statistics – **GPA 4.0/4.0** *Sep 2020 – May 2021*

DePauw University *Greencastle, IN*
Bachelor of Music Arts in Mathematics & Music (Double Major) – **GPA 3.73/4.0** *Aug 2016 – May 2020*

- Honor/Awards: Recipient of School of Music Dean's Scholarship & Music Performance Award
- Dean's List: Every Semester from Fall 2016 to Spring 2020

Coursework: Applied Data Science, Machine Learning, Advanced Data Analysis, Intro to Database, Elements of Data Science

SKILLS

Programming Skills: R (advanced), Python (intermediate), SQL (intermediate), GCP (intermediate), Tableau (intermediate)
Hard Skills: Data Science, Statistical Model, Machine Learning, Algorithm Development, Data Manipulation & Visualization
Soft Skills: Written & Verbal Communication, Public Speaking, Detail Orientation, Cross-functional Teamwork, Curiosity

WORK EXPERIENCE

Percolata *San Francisco, CA*
Machine Learning Intern *May 2021 – Current*

- Designed and implemented Reinforcement Learning Algorithm for simulating portfolio trading in stock market using Pytorch; evaluated trading strategies and made reports through Google Data Studio
- Built both machine learning and time series models including Ridge Regression, KNN, Decision Tree, Random Forest, ARIMA, and Exponential Smoothing; evaluated models by Testing and Training R^2, MSE, and Cross Validation
- Used Selenium (Web Scrawler) with Python to log in and extract recommendation stocks every 5 minutes from MotleyFool; checked the corresponding recommended call option prices and purchased immediately after the release by Interactive and Alpaka Brokers with their APIs; placed a trailing stop loss trigger to sell the option and gain profit
- Implemented google cloud functions and cloud run services to aid the stock trading evaluations and the interactions with Alpaca API, Interactive Brokers API, Stripe API, and google cloud API.
- Designed and implement cloud-based CI/CD workflow to automate the testing and deployment of source code
- Created and managed 3 databases including customer information, stock screening, and experiment tracking by BigQuery

DePauw University *Greencastle, IN*
Admission Intern *Aug 2019 – May 2020*

- Cleaned and manipulated the data by Tidyverse in R, figured out seven possible predictors for students who might accept the offer; conducted a PCA analysis and found three major principal components from the dataset
- Inputted more than 500 prospective students by SLATE database, managed multiple databases by SQL
- Extracted data and visualized the components of prospective, admitted, and current students by Tableau

PROJECT EXPERIENCE

Shiny-App for COVID-19 Resilience Ranking (github.com/TZstatsADS/ Spring2021-Project2-group1) *Spring 2021*

- Developed a website application to show the interactive map of daily covid deaths and cases, interactive plots of vaccinations and mobility, and interactive table of performance of each state in the U.S. for COVID-19
- Gathered and cleaned daily updated online raw data; created data structures; constructed the map and plots for displaying cases and death for each state by Leaflet with Shiny; wrote "About" and "Methodology" page by HTML
- Wrote the big structure of the App for both user interface and server parts; gathered and debugged most of the code; wrote unit tests for data gathering functions and ranking calculus max-min method
- Deployed the app to the online server (virtual machine instances) by Google Cloud Platform and Shinyapps.io

Recognized the Emotions (github.com/TZstatsADS/Spring2021-Project3-group-5) *Spring 2021*

- Developed a new method of facial emotion recognition using Image Transpose, Modified Feature, and Random Forest; improved the classification accuracy for image of emotions on faces to 0.90 AUC and 91.2% accuracy
- Built four types of Random Forest Model with ROSE balanced method, SMOTE balanced methods, the imbalanced training data, and old-feature functions
- Constructed, tuned, and trained Random Forest Model; tested baseline model (GBM) and other advanced models (SVM, LDA, Random Forest, and Ridge Regression); predicted emotions by giving new testing set

ACTIVITIES AND INTERESTS

Clubs: President of Columbia Statistics Club (2020-2021), President of International Student Association (2019-2020), President of Association of Student Interested in Asia (ASIA) (2019-2020)
Music Performance: Professional Violinist, Columbia University Orchestra, DePauw Symphony Orchestra