

DI MU

400 W 113th Street
New York, NY 10025

(917) 325-0727
dm3686@columbia.edu

EDUCATION

Columbia University

Master of Science in Data Science, 4.0/4.0,

Coursework: Databases, Applied Machine Learning, Big Data Analytics, Practical Deep Learning System Performance, Computer System, Algorithms, Stats Inference

New York, NY, US

Expected Dec 2022

University of Toronto

Bachelor's of Science in Statistics, 3.99/4.0, Dean's List

Minor: Economics

Coursework: Neural Networks and Deep Learning, Statistical Methods for Machine Learning I & II, Data Visualization, Statistical Computation, Time Series Analysis, Data Analysis in Practice, Applied Statistics, Experimental Design, Sampling Methods

Toronto, ON, CA

Jun 2021

RELATED EXPERIENCE

University of Toronto

Research Assistant

Toronto, ON, CA

May 2021

- Established open media databases and a dashboard-based website for demonstration with support of online exploratory analysis
- Implemented classical statistical and machine learning classification models to analyze quality and statistical sophistication for news and articles. Balanced between the statistical interpretation and model accuracy
- Utilized corpus-based methods, latent semantic analysis and neural network-based methods for media text summarization

Gong Dao Technology | Alibaba Group

Data Science Intern

Hangzhou, CN

Jul 2020

- Constructed data dictionary using SQL. Improved the efficiency of data searching and understanding for team members
- Implemented data interaction with Alibaba Maxcompute Data Cloud under both live and off-line environments for more than three databases, including but not limited to data uploading, data extracting, data cleaning, etc
- Developed more than 13 intelligent interactive Data Visualization Screens, with nodes

Asian Infrastructure Investment Bank (AIIB)

Data Science Assistant

Beijing, CN

Apr 2020

- Led the Maintenance and updates of two significant databases. Cooperated with the data analysis team to automate the data extraction, data filtering, and data visualization process. Immensely relieved the manual work
- Directed the whole process of an emergency project on Bangladesh, Pakistan, and Indonesia government loans, including the construction of four databases, PowerPoint Making, and roadshow organizing
- Managed ad hoc quantitative related requests from the management team, e.g., performed data validation for the board of directors' monthly reports, conducted business intelligence analysis, etc

China Securities Co., Ltd.

Data Analytics, Investment Banking Division

Beijing, China

Jul 2021

- Wrote due diligence, financial status report; collected data and established databases for financial analysis
- Performed industry research and structured the financial analytics result for both internal tracking and external consultation

DATA SCIENCE PROJECTS

Deep Learning, Forecast the stock market movement using NLP methods

Mar 2021

- Pre-processed Reuters and Bloomberg financial news dataset utilizing Word2Vec and Bert embeddings and trained it with CNN and RNN based architectures
- Tuned the hyper-parameters and compared performance in the report. Achieved 57.5% accuracies using LSTM model

Big Data Analytics, Hourly Playlist Generator Based on Weather Forecast

Dec 2021

- Deployed LSTM-based model for hourly weather forecast using the 'OpenWeather' API, and 'Spotify' API, where candidate models are selected by varying response encoding method, window size and model type
- Designed Surveys and executed A/B Test on Facebook Group to select the model to be deployed, where the baseline model is set based on the survey result. Achieved statistically significant result under 5% significant level
- Built website where users could login, view 7 hours ahead temperature predictions and play with the recommended playlists

Visualization, Literacy is Power: A R-shiny Dashboard Platform for Analysis

Apr 2021

- Collected, cleaned and pre-processed data using web scraping, executed feature engineer with the initial exploratory data analytics. Embedded simple model results in the interactive plots to guide visualization designs
- Developed an interactive analytical and predictive dashboard portal for international organizations to track literacy rate and make decisions. Led three team members to complete design rationale, project presentation, and product report

LANGUAGE AND IT SKILLS

- Programming: Python, R, SQL, Java, Stata
- Python Packages: Numpy, Pandas, Scikit-Learn, NLTK, Gensim, BeautifulSoup, Pyspark, Pytorch
- Data Visualization: R-shiny, Tableau, Matplotlib, ggplot2, D3.js, Seaborn, highchart
- Other: Google Cloud Platform, LaTeX, Github