YONGHAO YU

+1 3476766821 | yy3564@cumc.columbia.edu

EDUCATION

Columbia University, *New York City, NY* Sept 2024 - Present

Master of Science in Biostatistics (Public Health Data Science Track)

City University of Hong Kong, *Hong Kong, HK* Sept 2020 - June 2024

Bachelor of Science in Data Science | Minor in Mathematics, GPA:3.56

Major courses**:** *Data Visualization, Data Structures, Database Systems, Single-Variable Calculus, Multivariate Calculus, Linear Algebra, Probability and Statistics, Convex Optimization, Machine Learning, Reinforcement Learning, Operation Research*

Minor courses**:** *Mathematical Analysis, Numerical Methods, Real analysis, Functional Analysis, Advanced Statistics, Bayesian Statistics, Stochastic Processes, Mathematical Finance*

RESEARCH EXPERIENCE

Research Assistant, Department of Mathematics, City University of Hong Kong Mar 2023 - June 2023

Supervisor: Dr.Xiaosheng ZHUANG

- Conducted literature reviews to support the research project
- Reimplemented Matlab image processing scripts in Python using OpenCV, imageio, and other libraries
- Utilized grayscaling and geometric transformations to analyze images and audio related data, enhancing research progression

Research Assistant, Department of Mathematics, City University of Hong Kong June 2023 - Sept 2023

Supervisor: Dr.Yukun HE

- Investigated convergence rate of Central Limit Theorem in a novel perspective, focusing on the behavior of random variables as sample sizes increase
- Employed advanced mathematical tools including statistical properties, cumulant expansion, Euler's equations, first-order inhomogeneous ordinary differential equations, and various inequalities for research purposes
- Proved a weakened conclusion of the Berry-Esseen theorem regarding the convergence rate, showcasing a novel perspective on its behavior

PROFESSIONAL EXPERIENCE

ByteDance Ltd. May 2022 - July 2022

Product Manager Intern

- Designed a simplified platform to assist merchants in advertising with precision and efficiency by developing a statistical questionnaire especially with eight problem-solving questions, specializing in the e-commerce platform of TikTok
- Introduced two distinct platform solutions tailored for merchant attraction, and conduct A/B test to evaluate the performance of both strategies
- Acquired in-depth knowledge of live-streaming e-commerce and advertising, offering valuable insights into commercial product
- Created a comprehensive product document and report for the simplified platform project, covering background information, business details, project plan, future outlook, and expected timeline

PROJECT EXPERIENCE

Credit Card Fraud Detection Oct 2021 - Nov 2021

- Conducted data preprocessing, including checking for missing and outlier values, data dimension reduction, and visualization
- Performed undersampling, used machine learning algorithms including logistic regression, support vector machine, AdaBoost, and random forest to evaluate prediction performance under five-fold cross-validation

Telecom Customer Churn Prediction Analysis Mar 2021 - Apr 2021

- Calculated the AUC value of the machine learning models and plotted ROC curves to select the best model
- Utilized logistic regression model to obtain importance scores for different features, identifying key factors affecting customer churn and calculating the probability of churn for each customer based on their unique features

Deep Reinforcement Learning on OpenAI Gym Games Oct 2022 - Dec 2022

- Trained agents in a car racing game environment using Proximal Policy Optimization(PPO) and Deep Q-Learning Network(DQN)
- Compared the advantages and disadvantages of both the two algorithms, studied and further investigated advanced algorithms for network enhancements

Online Social Media Data Analysis of the Seoul Halloween Crowd Crush Event Oct 2022 - Nov 2022

- Collected extensive data on Instagram, manually categorized the data into five categories, analyzed the proportion of comment categories, and used hashtags, comments, and other relationships to analyze the degree and speed of the event's spread
- Used Long Short Term Memory(LSTM) network to predict the approximate time when the event will subside on the Internet

High-Tech Companies Stock Return Prediction Analysis Jan 2024 - May 2024

- Did exploratory data analysis on a series of high tech companies stock prices, then analyzed their trends in conjunction with news events over the course of five years by adopting causal inference techniques.
- Enlightened by Chaos Theory and tried to model the financial market and corresponding return by modifying the Lorentz Equation to customize the financial field use, then generate the stock returns
- Fitted five years of stock data using GARCH model, LSTM model, GRU model, and made predictions for future stock returns, then compared four methods to select the best model

PROFESSIONAL SKILLS

Language: Chinese Mandarin(Native), English(TOEFL: 104)(GRE: 328), Cantonese(Beginner)

Programming Language: Python, MATLAB, R, C++, SQL,

Technical Skills: Gephi, Tableau, Microsoft Office(Word, Excel, PowerPoint), Video Editing(PR)