Kaiyu He

Tell: (917) 238 2465 E-mail : kh3074@cumc.columbia.edu GitHub: github.com/MeanStudent Website:

EDUCATION

Columbia University in the City of New York, New York, US Sep 2021– Jun 2023(expected)

M.S in Biostatistics, GPA:3.9/4.4

Renmin University of China, Beijing, China Sep 2016 – Jun 2020

B.S in Applied Statistics, GPA:3.2/4.0

RESEARCH

A Decoded EEG Neurofeedback Platform Using Muse2, Prof. XiaoFu He Apr 2022 - Present

Research Assistant, Data Science Institute, Columbia University

- Design, train, and tune Deep Neural based Classifier for EEG data; Collect EEG data using Muse2.
- Select model structure including LSTM, and Conv1D model.
- Preprocess raw EEG data for Classifier, including filtering noise by Fourier transform and ICA.
- Successfully increased the accuracy of the classifier by 30% on the test set.
- (Scheduled to submit to rtFIN in October 2022)

New York Covid Case Investigation Analysis, Prof. Sen Pei Mar 2022 – Present

Research Assistant, Mailman School of Public Health, Columbia University

- Analyze key features affecting the complete rate of case investigation and contact intake.
- Using GLM, random forest, and Linear Mixed effect model to evaluate the significance of features.
- Train NLP models using messages in voice mails to predict completeness for each phone interview.
- Devise the best strategies for Covid tracers to make phone calls to potentially exposed people.
- (Scheduled to submit in Nov2022)

PROJECTS

Latent Space Discovery of DCGAN Model **Independent Project** Jul 2022– June 2022

- Train DCGAN models using TensorFlow to generate images.
- Train classifiers for different types of generated images.
- Visualize images by their latent representations to visualize their clusters and patterns.

Abstract text summarization using Sim-CLS. **Independent Project** **Project Link** Oct 2022– Present

- Try SimCLS tuning in more recent models and see how SimCLS boost different types of models.
- Try SimCLS on different datasets and see if SimCLS preferred to work with certain types of data.
- Using medical datasets to train a SimCLS model to auto titling medical papers.

"Microblogging in Museum Marketing: A Clustering Approach with LDA" Apr 2018 - Apr 2019

Team Leader, Renmin University

- Use LDA(Latent Dirichlet Allocation) analysis and regression analysis on microblog topics information to evaluate the effects of microblogging on museum marketing.
- In charge of data scraping, LDA analysis, and linear regression analysis with Python.
- Received the **National Excellent Project Award** for the Undergraduate Innovative Test Program in 2019.

2022 Columbia University Innovation Award (VR) Project Link Video Link Mar - May 2022

Teachers College, Columbia University

- Built an educational game for teaching physics using Unity Engine.
- Designed the whole game, and built the demo video.

WORKING EXPERIENCE

89 Trillion, Beijing, China Aug 2020 – Apr 2021

Game Numeric Designer

- Worked with the AI team to select the in-game numeric features used for Recommend System.
- Conducted A-B tests to determine the performance of new systems in the game.
- Set numeric values for game content according to the statistic model.

CERTIFICATE

- Tensorflow Developer Certificate Apr 2022

SKILLS

- **Programming language:** Python, C#, R
- **ML/DL:** Tensorflow, PyTorch, Numpy, Pandas, Matplotlib, Caret(R), SciKit-Learning
- **Others:** MySQL, Git, Latex, Unity