Kuan-Lin (Gary) Liu

kll482@nyu.edu ◇ (929) 332-5669 ◇ linkedin.com/in/kuanlinliu ◇ Jersey City, NJ ◇ garylkl.github.io

EDUCATION

New York University — New York, NY
M.S. in Data Science, GPA: 3.56 / 4.0 — Expected May 2021

- Related Courses: Natural Language Understanding, Machine Learning, Big Data Application Development, Database Systems, Introduction to Data Science

National Taipei University (NTPU) — New Taipei, TW
B.B.A. in Statistics and B.A. in Economics, GPA: 3.83 / 4.0 — Sept 2014 - Jan 2019

- Related Courses: Data Mining, High Dimensional Data Analysis, Data Structures and Algorithms, Object-oriented Programming, Applied Time Series Analysis, Experimental Design, Mathematical Statistics

TECHNICAL SKILLS

Programming	Python, Scala, SQL, R, C++, MATLAB, SAS
Software & Tools	Hadoop, Spark, MySQL, MongoDB, Flask, Git, Tableau, HTML, CSS, Microsoft Excel
Packages	PyTorch, Scikit-learn, NLTK, TextBlob, Pandas, NumPy, Matplotlib, TensorFlow, Dash

SELECTED PROJECTS

Twitter Sentiment Analysis: COVID-19 Outbreak — Present

- Collecting realtime tweets related to COVID-19 using Spark Streaming and stored JSON data into MongoDB
- Analyzing people's reactions and comparing the results between TextBlob and transfer learning on 1.6 million tweets

Predicting Kickstarter Success and Recommending Products from Amazon — Oct 2019 - Dec 2019

- Built machine learning pipeline with Spark's MLlib and BigDL, and visualized patterns with a Tableau dashboard
- Implemented Spark using Scala queries and Spark SQL for faster cleaning JSON files from HDFS
- Recommended similar products on Amazon by Locality-sensitive Hashing and extracted top words by TF-IDF

Predicting NBA Game Outcomes Using Machine Learning — Oct 2019 - Dec 2019

- Improved AUC by 18% with Ridge Logistic Regression and SVM, compared to Decision Tree
- Used Scikit-learn to build machine learning models with blocked time series cross validation
- Cleaned player statistics by Pandas and Numpy, and exploring play-by-play data by Matplotlib and Seaborn

Automatic NBA Game Information Messaging Bot — Sept 2019 - Oct 2019

- Scheduled Python scripts on AWS EC2 and retrieved data from Amazon RDS for MySQL
- Constructed a web dashboard by Dash, Plotly, and HTML with a web scraper (Requests, Beautiful Soup)

PROFESSIONAL EXPERIENCE

Research Center for Humanities and Social Sciences, Academia Sinica — Taipei, TW
Research Assistant — Mar 2018 - July 2018

- Wrangled energy usage data from an economics experiment and visualized abnormal records (dplyr, ggplot2)
- Grouped 300+ rooms by usage habits on panel data (Hierarchical Clustering, Dynamic Time Warping)
- Instructed teammates in machine learning methods and presented data insights using R Markdown
- Extracted extra 10+ factors by scraping meteorological data (RSelenium)

Department of Economics, National Taipei University — New Taipei, TW
Teaching Assistant (Courses: Programming for Data Science) — Sept 2018 - Jan 2019

- Instructed 80+ students in collaborating with Git and R programming data wrangling using real-world data
- Designed an interactive tutorial web app by learnr and R Shiny to assess students' learning progress

Department of Statistics, National Taipei University — New Taipei, TW
Undergraduate Researcher — Dec 2017 - June 2018

- Decreased RMSE by 30% with a baseline SVR model and evaluated dimension reduction techniques (SIR, Isomap)
- Set up R scripts on Google Cloud Platform and controlled remote servers using TeamViewer
- Analyzed the dissimilarity using Hierarchical Clustering among 9 economic factors within various industries