Kuan-Lin (Gary) Liu

kuanlin.liu@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: 3.80 / 4.0 — Expected May 2021

- *Related Courses*: Natural Language Understanding, Advanced Machine Learning, Big Data (Hadoop, Spark, Scala), Database Systems, Introduction to Data Science

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

- Dean's List, Fall 2017; Scholarships from Academia Sinica, Summer 2017 and 2018
- *Related Courses*: Data Mining, Dimension Reduction, Data Structures and Algorithms, Object-oriented Programming, Experimental Design, Mathematical Statistics, Time Series, Categorical and Linear Regression, Econometrics

TECHNICAL SKILLS

Programming	Python, Scala, SQL, R, C++, MATLAB, SAS
Software & Tools	Hadoop, Spark, MySQL, MongoDB, RESTful API, Git, Tableau, HTML, CSS
Packages	PyTorch, Scikit-learn, NLTK, Spacy, Flask, Selenium, BeautifulSoup, Pandas, NumPy

SELECTED PROJECTS

Fake Review Detection for Restaurants in New York City — Mar 2020 - May 2020

- Boosted efficiency of the text preprocessing pipeline from 20 hours to minutes by parallel computing and Spark NLP
- Focused on behavioral feature engineering to improve ROC score by 16%, compared to Bag-of-Words text features

Automatic Text Summary Generator: News Application — Mar 2020 - May 2020

- Integrated BertSum and TextRank to rank the importance of each sentence in the 1.4GB CNN/DailyMail dataset
- Designed a novel text summarization pipeline to reduce word repetition by fine-tuning a sequence-to-sequence model

Book Recommendation Systems for Goodreads.com — Mar 2020 - May 2020

- Analyzed 4GB user-item interaction data in PySpark and built collaborative filtering recommendation systems
- Enhanced efficiency of the book-searching system by 10 times and visualized clustering of products with T-SNE

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 and Spark SQL queries for faster cleaning JSON files from HDFS
- Recommended similar products on Amazon by Locality-sensitive Hashing and extracted top words by TF-IDF

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 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 communicated with remote servers and partners using TeamViewer