Kunru Lu

kl3743@nyu.edu | (347)405-0620 | Brooklyn, NY
LinkedIn: linkedin.com/in/lukunru | Github: github.com/carajumpshigh

EDUCATION

New York University, MSc in Urban Informatics (Graduating in 12/2020) — **New York, Present**
Tongji University, BEng in Civil Engineering & Mathematics — **Shanghai, 09/2015**
Shanghai University of Finance and Economics, Accounting — **Shanghai, 09/2016**

CORE CURRICULUM

Applied Data Science, Machine Learning, Fundamental Algorithm, Big Data Management, NLP, Deep Learning, Urban Decision Models, Civic Analytics & Urban Intelligence, Principal of Urban Informatics, Statistics

TECHNICAL SKILLS

Machine Learning: Python (Scikit-Learn, Tensorflow), MATLAB | Predictive Modeling, Regression Analysis, GAN, NLP
Data Engineering: SQL, Spark, Hadoop, Hive, MapReduce, Neo4j
Data Analytics: SAS, SPSS, Tableau, QGIS, A/B Testing, Microsoft Skills (Excel, Pivot Tables)

PROFESSIONAL EXPERIENCE

GE Digital| *Data Analyst Intern, GDS*| Shanghai, China — **08/2018-02/2019**

- Manipulated the Financial Data Lake and worked on technical documentation
- Rewrote the python files in PySpark for real time, large-scale data processing
- Generated weekly monitoring of large-scale invoices' data with typically including over 100,000 items by SAS

Project: Global Project of Account-Receivable (AR) Scoring and Monitoring

- Translated between SAS and python code to meet the need of code standardization for the global team
- Accomplished data cleaning and implemented the score card model to evaluate the performance of the invoices
- Conducted automatic monitoring and visualization system using python

PwC LLP| *Junior Data Analyst Intern, Advisory* |Shanghai, China — **07/2018-08/2018**

Project: Mapping Knowledge Domains Development

- Parsed XML with Go and established graph database on Neo4j with Cypher for knowledge mapping& querying
- Used Stanford CoreNLP to carry out Chinese word segmentation, labeled the importance of vocabulary entry based on PageRank algorithm and reduced over 50% of the previous workload

Assurance Intern, PwC Assurance — **01/2018-03/2018**

- Wrote the draft of enterprise annual report and used Aura and Excel for processing and analyzing financial data
- Performed data integrity test and sampling inspection

China Times Asset Management Co., Ltd |*Quantitative Analyst Intern*| Shanghai, China — **10/2017-01/2018**

- Preprocessed data cleaning, used PCA and other algorithms to generate main factors that affected the market
- Constructed multi-factor stock selection strategy framework based on MATLAB
- Implemented the optimization application of attention mechanisms based RNN in FOF quantitative investment, conducted back-test, and wrote algorithm evaluation reports

Shanghai Hua Rui Bank| *Data Mining Intern, Big Data Center*| Shanghai, China — **07/2017-10/2017**

- Created new risk management model based on iForest algorithm and LOF algorithm and implemented by MATLAB
- Cooperated with data scientist on establishing new anti-fraud system based on SNA algorithm
- Trained non-tech colleagues in the algorithms that used in our products, and introduced new algorithms and related application scenarios to tech colleagues on weekly base

iSIG Lab, Fudan University| Research Assistant| Shanghai, China — **03/2017-07/2017**

Project: Quantitative Evaluation of Corporate Social Responsibility (CSR)

- Generated desk research, analyzed the listed companies by dealing with more than 700 CSR related indicators that across 6 dimensions & 14 attributes, and incorporated new indicators into the evaluation system of Fair Wealth Lab

PROJECTS

Applying Multi-Agent RL to SLAM with Graph Pose for Autonomous Drone Swarms, NYU Courant — **03/2020-08/2020**

- Wrote web crawler for bulk download of paired disaster images and used GAN to generate synthetic data

- Compared the performance of SIFT/SURF and ORB algorithms to solve the change detection problem, and did image enhancement to improve spectral and spatial resolution of paired images
- Implemented Deep Q-Learning algorithm for path planning, job assignment and scheduling of the drones based on uncertainty in each raster
- Generated damage equation as export knowledge for calibrating the synthetic images generated by GAN

Analysis of Airbnb house pricing in NYC & LA, NYU CUSP **03/2019-05/2020**

- Conducted time series analysis to compare the long-term trend and seasonal changes of Airbnb price
- Conducted sentiment analysis on the reviews for the customers' general impressions on each listing house to filter out the houses that are not reasonably priced
- Spatial joint the pricing data with demographic data and crime data, and implemented multiple regression model and random forest classification to get the influential factors of NYC and LA for comparison

Google Q&A Labeling, Kaggle Competition **12/2019-01/2020**

- Conducted the baseline model by using BERT pre-tuned models to process natural language and Tensorflow to solve the classification problem
- Implemented the SMOTE algorisms and multilabel stratified KFolds methods to solve the imbalanced data problems and improved mean AUC of the pre-tuned Bert model by 0.32

Analysis of NYC Noise Complaints, NYU CUSP **03/2019-05/2020**

- Conducted EDA, time series analysis and spatial analysis on NYC 311 open data and census data
- Accomplished feature engineering, and generated regression to study the correlation between noise complaints and demographic, economic and socio attributes

Pipeline Leakage Diagnosis, Shanghai Institute of Disaster Prevention and Relief **03/2019-07/2019**

- Proposed a double-layer clustering model for anomaly diagnosis on over 16,000 time series based on DTW and spectral clustering algorithm, saving 98% of the total computation
- Optimized the initial model based on Page Rank algorithm and risk cluster analysis, considering the geographic distribution of accidents