

Yilun Chen

734-834-6827, yilunc0127@gmail.com, yc3214@columbia.edu
LinkedIn: <https://www.linkedin.com/in/yilun-alice-chen/>

EDUCATION

Columbia University, New York, NY

Sep.2016 - Dec.2017

M.S. Operations Research

Overall GPA: 4.0 / 4.0

Core Courses: Machine Learning, Deep Learning, Digital Operating Models, Database Design, Analysis of Algorithm, Stochastic Models, Optimization Models, Bayesian Statistics, Time Series Analysis, Multivariate Statistical Inference, Simulation

University of Michigan, Ann Arbor, MI

Sep.2014 - July.2016

B.S. Honors Mathematics & B.S. Statistics (With High Distinction)

Overall GPA: 3.88 / 4.0

Core Courses: Data Mining, Statistical Computing, Linear Regression, Numerical Method, Probability Theory, Programming and Data Structures(C++), Data Structure and Algorithm(C++), Real Analysis, Combinatorial Theory, Mathematics of Finance

Sichuan University, Sichuan, China

Sep.2012 - July.2014

B.S. Mathematics (Wu Yuzhang Honor School: Top 1.5% selected)

Overall GPA: 88 / 100

National College Students Mathematical Competition - Third Price in Sichuan Province

WORK EXPERIENCE

Air Liquide, Inc.

March.2018 – present

Research Associate in Optimization & Data Science

Newark, DE

- Use internal productions, pricing, distribution data and Moody's macro-economic data to estimate supply and demand, price elasticity of diverse industrial gases on thousands of different market segments for different years; Construct Cournot Game Theory Model for network market and implement the model using Pyomo to solve out the Nash Equilibrium of price and quantity for future ten years
- Estimate USA freight railroad revenue with diverse machine learning algorithms. Pipeline the data-preprocessing, feature engineering and model fine-tuning process on public waybills freight revenue data and validate the prediction using internal transportation contract data. Compare results from XGBoost, LightGBM, Ridge, LASSO, ElasticNet, Neural Network (AWS AutoML as benchmark); predict railroad distribution cost for diverse industrial gases and create a distribution cost matrix for thousands of geo-spatial data combination of plants and counties; estimate production and distribution cost, demand forecast as inputs for game theory model
- Demonstrate the modeling results to other Research Scientists to support inner-group demonstration and communication using Jupyter Notebook. Work with product manager to validate the production process, marginal cost, price elasticity, etc.
- Collaborate with Software Developer to develop a deployable web application using Bokeh. Deploy the functionality of Login and remote-access with Flask and Bokeh server

AIG, Inc.

Sept.2017 - Dec.2017

Research Assistant at AIG Science Group (Part-time, Contract)

New York, NY

- Participate the redesign of AIG's Italian Financial Line service system to reduce turnaround time (increase throughput) of underwriting service systems; develop near optimal data-driven scheduling strategies on underwriting service and report the research results to internal business stakeholders
- Construct process mining with R on insurance claim event logs and figure out the bottleneck of turnaround time; benchmark the operation workflow by giving priority to smaller tasks (rule-based policy); model validation using simulation-based optimization algorithm with AnyLogic

Pactera Technology International Ltd.

July.2017 - Aug.2017

Data Scientist Intern

Beijing, China

- Explore the data warehouse architecture of the Ai Direct Bank; query 1 TB sampled historical transaction data from Citic bank's customers database; pre-process and wrangle data using Apache Hive and SparkSQL for feature engineering that satisfies the requirement of fraud detection and customer image depiction
- Validated feature engineering using exploratory data analysis and pre-tuned logistic regression models with spark installed in Hortonwork and Zeppline

China Merchants Securities Co.,Ltd.

Jun.2015 - Aug.2015

Investment Analyst Intern

Beijing, China

- Collected related report materials in apparel industries; extracted data from news for word-cloud depiction and sentiment analysis
- Calculated VaR and Expected Shortfall for multiple portfolios; produced data analysis reports and presented result to clients

TECHNICAL STRENGTH

Computer Languages: Python (Scikit-learn, Tensorflow, PySpark, Flask), R, SQL, C++, Matlab

Other skills / Knowledge: Spark, Git, Google Cloud, AWS, Natural Language Processing, Hive

PROFESSIONAL PROJECTS

Deep Learning on Anomaly Detection Project

Sept.2017 - Dec.2017

- Implement and develop Time-series Anomaly Detection using LSTM RNN for network security signal data with Tensorflow and attain an accuracy of 97%

Sentiment Analysis project

March.2018 - May.2018

- Utilize NLP package, NLTK and Deep Learning methods, LSTM to explore Air Liquide's customer voice data. Attain an accuracy of 90% with the combination of Word2Vec and LSTM

ACADEMIC WORK EXPERIENCE

- Course Assistant of Optimization Models & Simulation, IEOR Department, Columbia University

Jan.2017 - Dec.2017

- Course Assistant of Probability theory & Mathematics of Finance, University of Michigan

Sept.2015 - July.2016