AVA (TONG) YANG

ty2418@columbia.edu | (929)386-8805 | Portfolio Website | GitHub | LinkedIn

Data-driven and analytical person with experience of leveraging **Statistical Analysis**, **Machine Learning**, **A/B Testing**, and **Data Visualization** tools to identify insights and suggest business recommendations. Well-versed in communicating key findings to cross-functional groups. Proficient knowledge in **Analytics**, **Statistics**, and **Programming**.

EDUCATION

Columbia University — **New York, NY**
Master of Science in Operations Research, Concentration in Data Science — *Sep 2019 – Dec 2020*
- Coursework: Machine Learning, Deep Learning, Business Analytics, Data Visualization, Stochastic Models

Zhongnan University of Economics and Law — **Wuhan, China**
Bachelor of Science in Information and Computer Science — *Sep 2015 – Jun 2019*
- Honors: National Scholarship (1%), College Academic Scholarship (2016, 2017, 2018)
- Coursework: Statistical Inference, Operations Research, Econometrics, Financial Mathematics

TECHNICAL SKILLS

- **Language & Framework:** Python, SQL, PyTorch, TensorFlow, Keras, Bash
- **Data Analysis & Visualization:** MySQL, Tableau, Plotly | PowerPoint, Excel
- **Big Data:** Google Cloud Platform, AWS (S3, Sagemaker, Gateway), Google Analytics
- **Machine Learning:** Logistic Regression, SVM, XGBoost, Random Forest, Clustering, NLP, Neural Networks

WORK EXPERIENCE

Articence (Intelligent Hiring Platform Startup) — **New York, NY**
Data Science Intern (Text Mining and Natural Language Processing) — *Jun 2020 – Aug 2020*
- Built a Job & Resume Analyzer Web App product using Machine Learning model in Python, to predict key skills match with model accuracy achieved of 90%, delivered personalized career guidance to customers
- Scraped software websites and clustered ~500 text reviews utilizing Topic Modeling algorithm, identified marketing campaign opportunities which improved 7% of Click-Through Rate
- Collaborated with engineers to cut product's runtime by 85%; partnered with product team and communicated user behavior insights with Tableau reports visualization

CreditX (Fintech Company) — **New York, NY**
Data Science Student Intern (Credit Risk Monitor) — *Feb 2020 – May 2020*
- Reduced loan default risk by 10% from user behaviors by deploying predictive models with Random Forest, XGBoost, and LSTM in Python, monitored potential fraud and gave alerts to business
- Mined 200K operations data and integrated Natural Language Processing (NLP) techniques to preprocess text data; conducted exploratory data analysis (EDA) in Tableau and data manipulation in SQL for modeling
- Customized Transformer Model to devise a feature-extraction pipeline in cloud environment (GCP), used by financial analysts to facilitate fraud detection process

Deloitte Consulting — **Beijing, China**
Tech Strategy Intern — *Mar 2019 – Jun 2019*
- Established a three-level analytical framework, covering company's over 100 business processes
- Conducted 6 case studies and explored value-adds for client based on current needs, developed and presented 45 PowerPoint slides to client during strategy discussion

PROJECT EXPERIENCE

A/B Testing: Adding Free Trial Screener Feature — **New York, NY**
- Defined evaluation metrics to track new feature impact and invariant metrics for experiment setup sanity check
- Designed experiment and implemented statistical testing with Python, suggested not to launch new feature through test analysis in terms of improving revenue

Machine Learning: Improving Auto Insurance Assigning Strategy — **New York, NY**
- Created an insurance designation model in Python which boosted 2% increase of expected profit
- Built a two-layer logistic regression model using decision factors analyzed from classification models, simulated expected insurance acceptance rates under different policies

Facebook: Interpreting Mathematical Reasoning Abilities of Sequential Models — **New York, NY**
- Constructed algebra problems solving models in Keras combining mathematical calculation rules
- Evaluated and visualized contribution of each feature to prediction result in LIME, interpreted models' ability to generalize knowledge by understanding internal learning mechanism