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, GPA 3.6/4.0 — *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, GPA 3.8/4.0 — *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
- **Data Analysis & Visualization:** MySQL, Tableau, Plotly | PowerPoint, Excel
- **Big Data:** Google Cloud Platform, AWS (S3, Sagemaker, Gateway)
- **Machine Learning:** Logistic Regression, SVM, XGBoost, Random Forest, K-Means, 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 using BERT model in Python, to predict key skills match with model accuracy achieved of 90%, delivered personalized career guidance to job seekers
- Scraped software websites and clustered ~500 text reviews utilizing unsupervised Machine Learning algorithm (Topic Modeling), identified 3 user experience opportunities for digital team's marketing campaign plans
- Collaborated with software engineers to cut Web App's run time by 85%; partnered with product team and created weekly Tableau reports on user behavior insights

CreditX (Fintech Company) — New York, NY
Data Science Student Intern (Credit Risk Monitor) — *Feb 2020 – May 2020*
- Reduced loan default risk by 10% from online-lending user behaviors by deploying predictive models with Random Forest, XGBoost and LSTM in cloud environment (GCP), monitored potential fraud and gave alerts to business
- Mined 200K operations data with 50+ categorical variables and integrated NLP techniques to extract text features; conducted exploratory data analysis (EDA) in Tableau and data manipulation in SQL for modeling
- Devised and productionized feature extraction flows for company's unstructured datasets enabling weak-indication data features effective for business, presented findings to financial analysts and managers biweekly

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

Machine Learning: Improving Auto Insurance Assigning Strategy — New York, NY
- Formulated new insurance designation strategy by combining more customers' features, facilitated 2% increase of expected profit by adjusting model feature weights, in collaboration with teammates
- Built a two-layer logistic regression model using decision factors analyzed from classification models, simulated 2 processes in Python to obtain expected insurance acceptance rates under different policies

Facebook: Interpreting Mathematical Reasoning Abilities of Sequential Models — New York, NY
- Constructed deep learning models in Keras for solving algebra problems with probability of correct answer achieved of 0.57, synthesizing calculation rules to make model inputs mathematically reasonable
- Evaluated and visualized contribution of each feature to prediction result in LIME, interpreted models' ability to generalize knowledge by understanding internal learning mechanism