

# AVA (TONG) YANG

[ty2418@columbia.edu](mailto: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

<b>Columbia University</b> <i>Master of Science in Operations Research, GPA 3.6/4.0</i> <ul style="list-style-type: none"><li>Coursework: Machine Learning, Deep Learning, Business Analytics, Data Visualization, Stochastic Models</li></ul>	<b>New York, NY</b> <i>Sep 2019 – Dec 2020</i>
<b>Zhongnan University of Economics and Law</b> <i>Bachelor of Science in Information and Computer Science, GPA 3.8/4.0</i> <ul style="list-style-type: none"><li>Honors: National Scholarship (1%), College Academic Scholarship (2016, 2017, 2018)</li><li>Coursework: Statistical Inference, Operations Research, Econometrics, Financial Mathematics</li></ul>	<b>Wuhan, China</b> <i>Sep 2015 – Jun 2019</i>

## 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)
- Machine Learning:** Logistic Regression, SVM, XGBoost, Random Forest, Clustering, NLP, Neural Networks

## WORK EXPERIENCE

<b>Articence (Intelligent Hiring Platform Startup)</b> <i>Data Science Intern (Text Mining and Natural Language Processing)</i> <ul style="list-style-type: none"><li>Built a Job &amp; 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</li><li>Scraped software websites and clustered ~500 text reviews utilizing unsupervised Machine Learning algorithm (Topic Modeling), identified opportunities for digital team's marketing campaign plans</li><li>Collaborated with software engineers to cut Web App's run time by 85%; partnered with product team and communicated user behavior insights with Tableau reports visualization</li></ul>	<b>New York, NY</b> <i>Jun 2020 – Aug 2020</i>
<b>CreditX (Fintech Company)</b> <i>Data Science Student Intern (Credit Risk Monitor)</i> <ul style="list-style-type: none"><li>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</li><li>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</li><li>Customized Transformer Model to devise a feature-extraction pipeline in cloud environment (GCP), used by financial analysts to facilitate fraud detection process</li></ul>	<b>New York, NY</b> <i>Feb 2020 – May 2020</i>
<b>Deloitte Consulting</b> <i>Tech Strategy Intern</i> <ul style="list-style-type: none"><li>Established a three-level analytical framework, covering company's over 100 business processes</li><li>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</li></ul>	<b>Beijing, China</b> <i>Mar 2019 – Jun 2019</i>

## PROJECT EXPERIENCE

<b>A/B Testing: Adding Free Trial Screener Feature</b> <ul style="list-style-type: none"><li>Defined evaluation metrics to track new feature impact and invariant metrics for experiment setup sanity check</li><li>Designed experiment and implemented statistical testing with Python, suggested not to launch new feature through test analysis in terms of improving revenue</li></ul>	<b>New York, NY</b>
<b>Machine Learning: Improving Auto Insurance Assigning Strategy</b> <ul style="list-style-type: none"><li>Created an insurance designation model in Python which boosted 2% increase of expected profit</li><li>Built a two-layer logistic regression model using decision factors analyzed from classification models, simulated expected insurance acceptance rates under different policies</li></ul>	<b>New York, NY</b>
<b>Facebook: Interpreting Mathematical Reasoning Abilities of Sequential Models</b> <ul style="list-style-type: none"><li>Constructed algebra problems solving models in Keras combining mathematical calculation rules</li><li>Evaluated and visualized contribution of each feature to prediction result in LIME, interpreted models' ability to generalize knowledge by understanding internal learning mechanism</li></ul>	<b>New York, NY</b>