

# Jiayang Nie

jiayangnie.com | Berkeley, California | 6264281707 | jnie@berkeley.edu

## EDUCATION

University of California, Berkeley

Dec 2022

Master in Statistics; GPA 4.00

*Courses: Linear Modelling, Statistical Inference, Advanced Probability, Deep Learning, Causal Inference, Experimental Design*

University of California, Berkeley

May 2021

Bachelor in Statistics; Minor in Mathematics; GPA 3.97; **Graduated with the Highest Distinction**

*Courses: Time Series, Linear Algebra, Stochastic Process, Game Theory, Bayesian Inference, Data Structure, Real Analysis*

## SKILLS

**Programming:** Python, R, SQL, Java, PySpark, Databrick, Git, Pandas, Pytorch, Sklearn, Seaborn

**Modelling:** AB-Testings, Ensembles, Clusterings, Linear Models, Bayesian Models, CNN, RNN, Transformer, Collaborative Filtering

## WORK EXPERIENCE

Vizio

Jun 2022 – Aug 2022

*Machine Learning Engineer Intern, AdsTech*

San Francisco

Enhanced smart TV ads targeting by detecting user engagement behaviors (what games they are playing and how long they played).

- Architected an ML structure of running XGBT on the nearest neighbors' distances retrieved from Approximate Nearest Neighbor model to classify the games played by users on Vizio's smart TV with the audio fingerprints collected from user end.
- Customized a Transformer to classify games from sessions of audio fingerprints, trained a k-means cluster on the max-pool layer output, and transformed raw data to histogram-valued data for session segmentation by chi-square distance with 90% recall rate.
- Aggregated the Transformer and ANN as two parallel layers to smooth the decision boundary, augmented training data by uniform noise and improved the model latency to be capable of processing 2.5-minute streaming data per second per GPU.

Guotai Junan Securities

May 2021 – Jul 2021 & Dec 2019 – Sep 2020

*Quantitative Researcher Intern, ML*

Remote & Shanghai, China

Designed algorithms and regression models on stock panel data to extract and process signals to profit through high-frequency trading.

- Applied models with cross-validation to obtain signals that maximize correlation with stock price change, such correlation coefficient reached around 50% with hypertuned XGBT model.
- Developed high frequency trading algorithm based on the signals retrieved from order book with back test showing a steady 5% return on a scale of 60 million fund for 2-month period; Naïve version of the algorithm already running on day trading machine.

Ant Group, Alibaba

Jul 2018 – Oct 2018

*Data Analyst Intern, Default Detection*

Shanghai, China

Engineered data pipelines to collect features and enhanced the modelling task of predicting bond default probability of public firms.

- Originated a filtering method for the dataset that improved a GBDT model precision rate by 50% with a fixed 80% recall rate for predicting bond-default possibility; team named this filtering method after my name.

## PROJECTS

**Click-Through Rate Prediction**

*Kaggle Data Challenge*

- Researched on the state-of-the-art models that predict CTR of ads impressions to recommend ranked customized ads to users.
- Applied GBDT+LR and Deep & Wide Factorization Machine model to dataset in Spark, reaching 30% recall and 40% precision.

**Movie Recommender System**

*MovieLens Project*

- Deployed collaborative filtering (ALS) on MovieLens dataset to train latent vectors for each movie and user by Spark.
- Ranked the top 10 movies for each user by the movie-user latent vectors to be recommended on their dashboards to watch.

**Experimentation Under Social Network Interference**

*Graduate Research*

- Researched on the solutions for measuring average treatment effects for experiments under social network spillover effect.
- Filled a minor gap for Aronow and Samii (2017)'s paper about what if the assumptions were violated with simulation study.
- Showed that biased WLS/OLS estimator outperforms unbiased Horvitz-Thompson estimator for complex social networks.

**Airbnb Demand Forecasting**

*Capstone Project*

- Led a team at UC Berkeley to analyze Airbnb listings' demand and revenue in Los Angeles areas on Databrick with MLs.
- Trained a price model with sentiment analysis to give suggestions to Airbnb hosts to maximize their return-on-investment.
- Inferenced the real estates and geographic factors that impact the potential return-on-investment and demand by SHAP to make informative advice to Airbnb hosts in terms of where to rent out, what to rent out and how to optimize description.

## PUBLICATION

Jiayang Nie, Xiao Qiao, Sibor Yan (2020). *COVID-19 Effects on Intraday Stock Market Behavior*. In Sabri Boubaker, Duc Nguyen (Eds.). Financial Transformations beyond the Covid-19 Health Crisis. World Scientific Publishing. (Published in 2022)