BOWEN ZHANG

808 Columbus Ave, New York NY 10025 ● (917) 547 3523 ● bz2461@columbia.edu ● LinkedIn

EDUCATION

Columbia University   New York, NY
M.S in Business Analytics   Expected Dec 2023
- **GPA**: 3.92/4.0
- **Relevant Coursework**: Artificial Intelligence, Data Analytics, Business Analytics, Optimization

University of British Columbia   Vancouver, BC
B.Comm in Business and Computer Science Honors   May 2022
- **GPA**: 3.85/4.0, Dean's List
- **Relevant Coursework**: Machine Learning, Statistics, Advanced Database, Algorithm Design

WORK & RESEARCH EXPERIENCE

PIMCO   New York, NY
Quantitative Research Analyst Intern   Jun 2023 – Present
- Implememted **LLM** workflow using **BERTopic** and **openai** to extract topics from financial news and predicted M&A activities with 91% accuracy; facilitated the quant portfolio management team to make investment decisions
- Implemented **incremental learning** leveraging **IncrementalPCA**, **MinibatchKMeans**, and **onlineCountVectorizer** to process 23 years of data; conducted **time series analysis** to showcase the topic trends over time

Noumena   New York, NY
Data Science Intern   Jan 2023 – May 2023
- Productionized the **regression** model to find intrinsic values for self-employed workers based on their personal, credit, and business information provided; leveraged the model to estimate clients' potential and achieved **RMSE** of 3.23
- Retrieved the clients' data from **Snowflake** by writing **SQL** queries; Reported the most important questions generated by the model by drawing out **SHAP force plot** and **summary plot**

China International Capital Corporation (CICC)   Beijing, CN
Data Science Intern   Jul 2021 – Oct 2021
- Developed ML models to assess the likelihood of loan default based on financial metrices from the WIND database and customer demographic data; achieved **roc_auc** score of 78% and contributed to $0.2M potential annual savings
- Implemented and compared algorithms namely **Random Forest**, **XGBoost**, and **LGBM** to make predictions; built pipeline for data preprocessing, conducted hyperparameter tuning using **RandomizedCV** with 5-fold **cross-validation**

TECHNICAL PROJECTS

Deep Learning: Loan default prediction   Apr 2023 – May 2023
- Built a credit scoring model by predicting loan default rates with 3-layered **feedforward neural networks** for LendingClub in **Keras**; optimized the model with **stochastic gradient descent** and achieved accuracy of 90%
- Implemented **EarlyStopping** and **dropout** to avoid overfitting; added **batch normalization** for data scaling; tested the model across 500 **epochs** with different batch sizes to improve accuracy

Natural Language Processing: Yelp Review   Apr 2023 – May 2023
- Predicted Yelp business reviews based on user comments by comparing models such as **DistilBERT** and **MultinomialNB** to improve the baseline accuracy by 11%; conducted **LDA** and **sentiment analysis**
- Applied various preprocessing methods to machine learning models namely **tf-idf**, **tokenization** and **stemming**

Big Data: Movie Recommendation   May 2020 – Jun 2020
- Built a **collaborative-filtering** movie recommender system with **PySpark ALS**, leveraging the computing power of **AWS EC2**; predicted movie ratings with RMSE of 1.36

SKILLS

- **Quantitative Skills:** Python (Numpy, Pandas, Seaborn, Scikit-Learn, MlLib, NLTK), R, SQL, LLM, Transformer
- **Database:** SQL*Plus, MySQL, Snowflake, Microsoft Access
- **Statistical Modeling:** UMAP, HDBSCAN, ANOVA, Factor Analysis, Ensemble Methods