

# Wenjia (Ivy) Zhang

+1 (412) 818-1847 | [ivy.wenjiazhang@gmail.com](mailto:ivy.wenjiazhang@gmail.com) | <https://www.linkedin.com/in/wenjia-zhang-cmu/>

## EDUCATION

### Carnegie Mellon University

Pittsburgh, PA

**Master of Information Systems Management** | GPA: 4.02/4.0

12/2022

- Coursework: Advanced Deep Learning, ML with Large Datasets, NLP, Data Structures, Distributed Systems, Big Data and Large-scale Computing, ML with Structured Data, Computer Systems

### Tongji University

Shanghai, China

**Bachelor of Mathematical Finance** | GPA: 3.7/4.0

06/2020

## SKILLS

**Programming:** Python (Pandas, Scikit-Learn, PyTorch, Keras, Tensorflow, PySpark, Matplotlib, Plotly) | R | SQL | Java

**Database & Tools:** Snowflake | Spark | Redshift | MongoDB | InfluxDB | Databricks | Git | Docker | Tableau

**ML:** Anomaly Detection | Time Series Forecasting | A/B Testing | Recommendation Systems | Survival Analysis | NLP

## WORK EXPERIENCE

### Capital One

New York, NY

**Principal Data Scientist, Global Finance**

03/2025 – Present

- Modernized prepayment forecast by replacing traditional regression with a **Cox Hazard** survival model, improving macro-sensitivity tuning and reducing prediction error by 10%, with extended adoption across multiple teams.
- Designed interactive analytical tools using Plotly and Tableau to enhance cross-team adoption and decision-making.
- Improved model interpretability and monitoring via automated drift detection, feature diagnostics, and reporting.

**Senior Data Scientist, Risk Management**

03/2023 – 03/2025

- Built a default balance prediction model for a \$100B+ commercial portfolio, boosting forecast robustness under COVID-like stress scenarios by 20% via data resampling, feature engineering, and synthetic default augmentation.
- Productionized and parallelized six enterprise risk models into an integrated loss forecasting system, reducing runtime by 50% through **parallel computing** and improving CI/CD deployment readiness.
- Collaborated with DA, SDE, and BA for model governance, documentation and performance monitoring.

**Data Science Intern, Risk Management**

06/2022 – 08/2022

- Created a scalable ETL pipeline handling 50M+ time-series records in **SQL** and **Python**, plus feature engineering and hypothesis testing (BP Test, Ramsey Test) for model assumption validation.

### Zhiyi Technology

Shenzhen, China

**Data Scientist**

07/2020 – 07/2021

- Built anomaly detection models (**DNN**, **LSTM**) for power plant IoT time series data using **TensorFlow**, increasing model recall rate from 35% to 89% through advanced imbalanced data algorithms (Focal Loss, cGAN).
- Processed TB-scale data via **PySpark**, enhancing pipeline reliability, scalability, and latency for real-time monitoring.
- Developed a promotion **recommendation system** using sentiment derived user features; **A/B Tests** increased retention by 21%.
- Collaborated cross-functionally with SDE, UI team, and PMs throughout R&D, UAT, deployment and maintenance.

### Amazon

Shanghai, China

**Business Intelligence Engineer Intern, Amazon Web Services**

06/2019 – 06/2020

- Optimized data aggregation by architecting an ETL pipeline to extract 5.5M+ AWS accounts from Redshift and Salesforce using **SQL** and **Python web crawling (Pandas, Selenium)**.

## PROJECTS

### Exposure at Default (EAD) Model

2023 – 2024

- Built and deployed a staged **logistic regression** model forecasting expected loss for a \$100B+ portfolio.
- Designed a two-stage framework to address the imbalanced default samples by first learning stable patterns from large healthy datasets and then modeling rare default behavior separately, reducing MAPE by 20% than prior model.
- Implemented feature selection, cross-validation, model evaluation, productionization, and automated monitoring.

### Promotion Recommendation System

2021 – 2022

- Developed and deployed a deep learning **recommendation model** for a 5k+ DAU food ordering platform.
- Conducted **A/B testing** with sentiment-adjusted weights, uplift modeling, and user-behavior segmentation; increased revenue by 10%.