Weibo Zheng

Wayland, MA | (617) 515-0668 | zheng.weib@northeastern.edu | LinkedIn | GitHub

EDUCATION

Northeastern University – Boston, MA
Master of Science (MS) in Data Analytics Engineering **May 2027**
Coursework: Data Management, Foundation Data Analytics Engineering, Data Mining, Master's Project

Ohio State University – Columbus, OH
Bachelor of Science (BS) in Mathematics & Economics **December 2024**
Coursework: Probability, Partial Differential Equation, Linear Algebra, Econometrics

WORK EXPERIENCE

Data Architect & Co-Founder | Hoolii Inc **June 2017 - April 2020**

- Architected scalable data infrastructure serving 20+ enterprise clients with role-based access controls and data segregation, ensuring each client's data remained isolated and secure while processing 100K records monthly.
- Developed custom payroll calculation models with automated distribution, handling complex compensation structures and reducing processing time by 80%.
- Built Inventory management system integrating sales data with warehouse records, enabling real-time inventory tracking and automated discrepancy alerts for supply chain optimization.
- Implemented REST APIs for real-time data synchronization between Zoho Creator and multiple 3rd party platforms including e-signature services (DocuSign/AdobeSign) and sales/CRM systems, automating data workflows across the entire business stack.

PROJECTS

LLM Orchestration Project: *Python, Anthropic API, Neural Networks, Chi-Square, PCA*

- Developed dual-methodology machine learning system achieving 79.2% accuracy in predicting executive promotion readiness using statistical approach (chi-square feature selection, PCA dimensionality reduction, cost-sensitive neural networks with 10:1 penalty) and semantic approach (Claude API holistic analysis with LLM-generated reasoning)
- Engineered LLM orchestration pipeline where Claude semantically groups 135 statistically significant keywords into 4 promotion pathways, generates candidate-specific explanations bridging quantitative scores with qualitative insights, and provides complete calculation transparency for business defensibility
- Built end-to-end data processing system handling 24 candidates (5 promoted VPs, 19 matched controls) with 18-month recognition windows, implementing NER-based anonymization for name collision resolution
- Conducted systematic error analysis categorizing false positives and false negatives to distinguish unavoidable data limitations from correctable model focus errors, improving prediction accuracy through iterative refinement

Cross-Category Recommendation System: *Python, Pytorch, HuggingFace Bert, PySpark, Numpy*

- Built recommendation engine analyzing 2.2M Amazon reviews to suggest toys based on baby product purchases, fine-tuning BERT for sentiment analysis achieving 79.5% Recall and 48.6% NDCG
- Integrated collaborative filtering with BERT-based sentiment features using cosine similarity, creating hybrid system handling 520K cross-category users

Crypto Price Prediction: *Python, ARIMA, Random Forest, XGboost, Sentiment Analysis*

- Developed time series forecasting model predicting crypto prices using ARIMA, Random Forest, and XGBoost with technical indicators and sentiment analysis

MySQL Database & Analytics System: *SQL, MySQL, Python, MongoDB*

- Designed normalized database schema for direct sales operations, wrote complex SQL queries for business intelligence including performance tracking and commission calculations