EDUCATION

M.S. in Business Analytics-Data Concentration

Boston University, Questrom School of Business, Boston, MA

B.Sc. in Economics

Guangdong University of Finance & Economics, School of Economics, Guangzhou, China

WORK EXPERIENCE

Management Consulting Intern, Digital Transformation

KPMG, Beijing, China

- Redefined industry-wide data valuation method by redesigning a depreciation-based cost model, correcting a core flaw in traditional purchase-cost methods, enabling more logical and realistic prediction of data asset value.
- Collected and engineered a multi-table regulatory enforcement records dataset (100+ GB) of financial institutions, leveraged cloud-based PySpark to build scalable pipelines for future data ingestion and pre-processing.
- Built a regression-based risk-scoring model using XGBoost and SHAP to uncover high-penalty client features (e.g., low revenue-to-loan ratio) for China Construction Bank, visualized risk profiles across competitors and uncovered that small-sized clients accounted for 36% of total penalties, prompting targeted audit recommendations.

KA Sales Department Intern, Business Analysis

JD Logistics, Guangzhou, China

- Developed a daily-based supply chain performance monitoring dashboard in Tableau, tracking on-time rates, process speeds, and warehouse efficiency to help teams to quickly detect operational blocks and unexpected risks.
- Implemented Python-based anomaly detection models (rolling z-scores) to flag spikes in product damage rates; utilized SQL and heatmaps for root cause analysis, uncovering that glass items accounted for 35% of all breakage.
- Proposed additional foam packaging for fragile goods, reducing monthly glass breakage rate from 35% to 25%.

PROJECT EXPERIENCE

Unsupervised ML and NLP Project: Personalized Review Recommendation System for Yelp Jan 2025 - Mar 2025

- Designed a modular NLP pipeline on GCP for a 7M Yelp review dataset, using TF-IDF and NMF to extract key topics and TextBlob for sentiment scoring; reduced raw data size by 40% via Parquet conversion for efficiency.
- Set up based model KNN and used Logistic Regression, tuned thresholds, reweighted classes to improved negative sentiment recall by 38%; validated model performance via 5-fold cross-validation on labeled dataset (F1 = 0.78).
- Built a question-driven review recommendation system that ranks reviews based on semantic relevance to naturallanguage queries.

Machine Learning Competition: Predictive Engagement Strategies for Humana LPPO Plan Sep 2024 - Oct 2024

2024 Humana-Mays Healthcare Analytics Case Competition (Humana): Top 14 out of 288 Teams (Team Leader)

- Processed 1.5M+ customer records through Python on GCP, creating pipelines to handle missing values (mode/ median), anomalies, and encode categorical features (One-hot encoding). Conducted EDA using histograms, box plots, correlation heatmaps. Used SHAP analysis to identify key behavioral and socioeconomic drivers.
- Created engagement features (e.g., claim frequency). Applied variance thresholds, mutual information, and RFE to select 30 optimal features from 100+ candidates, enhancing model interpretability and dimensional efficiency.
- Optimized Models (e.g., LightGBM) via StandardScaler, SMOTE for imbalanced data, and Bayesian Search for hyperparameter tuning, boosting AUC of CatBoost (best) from 69% to 77%. Validated model robustness using stratified K-fold cross-validation. Deployed in a cost-saving framework, cutting \$53M annually.

University-Enterprise Collaboration Project: Temu Sales Analysis and Optimization Project Apr 2024 - May 2024

- Conducted trend analysis on key metrics such as GMV and ARPU, identifying fluctuations and pinpointing periods
 of business growth stagnation. Developed targeted promotional strategies to address these weak phases.
- Utilized the RMF model to segment user groups and analyze consumption behaviors and payment habits. Provided personalized marketing recommendations for different user clusters and proposed the introduction of new payment methods to enhance conversion rates.
- Analyzed merchants' sales performance and ratings. Conducted Pareto analysis on sales across product categories, recommending strategies such as product bundling and combo sales to increase overall platform revenue.
- Validated the "Weekend Exclusive Promotion" strategy through A/B testing by selecting stratified samples and designing control/test groups; experiment indicated an 11% increase in DAU and an 8% uplift in weekend sales.

TECHNICAL SKILLS

Programming Skills: Python (e.g., NumPy, Pandas, scikit-learn, Keras), Spark, SQL, Tableau, Microsoft Office, AWS, Git **Analytical & Modeling Skills:** Statistics, Experimental Design, A/B Testing, Data Cleaning, Data Visualization, Dimension Reduction, Causal Inference, GCP, Supervised/Unsupervised Machine Learning, Deep Learning, Big Data

Jun 2024 GPA: 3.62

Jan 2023 - Mar 2023

GPA: 3.50

Expected Graduation: Jan 2026

Jul 2022 - Sep 2022