Yashi Huang

yashih960@gmail.com | www.linkedin.com/in/yashi-huang | (646)-256-9123 | Rochester, NY

Education

University of Rochester, Ph.D. Candidate in Operations Management 09/2021 – 10/2026 (Expected)

- GPA: 3.9/4.0

Columbia University, M.S. in Operations Research 09/2018 – 02/2020

- GPA: 3.9/4.0

Wuhan University, B.A. in Financial Engineering 09/2014 – 06/2018

- GPA: 3.8/4.0

Work Experience

Data Scientist Intern, Dana-Farber Cancer Institute – Boston, MA 05/2025 – 08/2025

- Designed and deployed the institute's first production machine learning model, integrating multiple data sources to predict patient wait times with **75–80% accuracy**, reducing manual estimation by ∼5 minutes per patient.
- Performed exploratory data analysis and engineered a feature set reflecting cancer-care operations (appointment schedule, real-time queue metrics, provider statistics) across **1M+ patient records**.
- Built scalable data pipelines (**8,200+ lines of code, 3,500+ lines of tests**) on Databricks using PySpark, Delta Lake, and MLflow, with automated retraining, drift monitoring, and performance evaluation.
- Defined custom prediction metrics to assess fairness and reliability and built post-launch monitoring, and visualized findings through interactive dashboards for clinical leadership.
- Collaborated with 15+ clinicians, product managers, and software engineers to align predictions with workflow needs, ensuring adoption and long-term maintainability through clear documentation.

Quantitative Researcher, Veta Investment Partners – Topeka, KS 09/2020 – 04/2021

- Developed and backtested valuation-based trading algorithms in SPY using volatility and momentum signals, achieving a **70% cumulative improvement** over benchmarks.
- Implemented a quantitative framework for option valuation and ETF ranking, integrating scenario analysis, stress testing, and statistical methods.
- Automated intraday portfolio attribution and risk reporting via SQL extraction, performance-metric computation, and real-time email delivery, cutting ∼4 hours/day of manual reporting.
- Built web dashboards to visualize real-time portfolio performance and risk KPIs, improving transparency and decision speed for stakeholders.

Quantitative Analyst Intern, Wisdom Capital Asset Management – New York, NY 02/2020 – 09/2020

- Derived option-implied probability distributions of the S&P 500 using non-parametric methods, improving predictive accuracy by calibrating with the Heston stochastic volatility model.
- Conducted risk analysis and scenario testing across short equity-hedged options portfolios, incorporating implied volatility, stress testing, and Value at Risk (VaR).
- Performed statistical analysis on 20+ assets over 30 years to compare return distributions and developed a customized analytical interface using VBA and Excel macros.

Research Experience

Probabilistic Forecasting of Intraday Occupancy in a Large Cancer Center Oct 2024 – Present

- Validated demand forecasting assumptions via hypothesis testing (Kolmogorov-Smirnov test, Chi-square test) plus trend/seasonality diagnostics using R to ensure model robustness.
- Framed intraday occupancy forecasting in a large-scale cancer-care facility as a set function learning problem: input is a set of scheduled sessions; output is the continuous intraday occupancy curve.
- Developed and benchmarked three complementary modeling routes: (i) top-down / bottom-up ML to forecast daily totals (Random Forest, XGBoost, Neural Networks), shape-matched with scaled Beta and Gaussian-process fits; (ii) direct aggregation of session-level curves by Monte Carlo simulations; (iii) PCA on historical curves combined with ARIMAX time-series models.
- Improved forecast accuracy by 14% over Naive average baselines; fed predictions into a Mixed-Integer Program-

ming model (predict-then-optimize) to reschedule physician sessions, raising resource utilization by 26%.

- Launched a human-in-the-loop tool (Python backend + Excel/VBA front end), enabling ops staff to simulate constraints and push one-click schedule updates.

Dynamic Interval Prediction of Waiting Times in a Cancer Care Facility Jul 2023 – Nov 2024

- Framed wait-time prediction as a dynamic probabilistic-interval problem and designed practice-driven metrics (interval width vs. coverage) to evaluate usefulness for patient communication.
- Invented an adaptive quantile-selection procedure on top of Random Survival Forest to optimize the width–coverage trade-off, producing intervals 5 min narrower than symmetric 15/85 bands at matched coverage.
- Customized a neural network model with pinball-loss–based objectives to model the full waiting-time distribution.
- Proposed a conditional, real-time updater of the wait-time distribution, improving accuracy by 5% while reducing interval width by 8%.
- Added explainability (SHAP-based visualizations) and a clinician UI to interpret patient/provider timelines and surface actionable causes of delay.

Predictive Scheduling in Service Systems with Simulation and Optimization Mar 2023 – Present

- Architected a Python-based simulation platform to model a single-server queueing system with uncertain arrivals, no-shows, and walk-ins for multiple appointment scheduling policies (overbooking, wave scheduling).
- Optimized a weighted multi-objective that aligns schedule-induced starts/finishes to simulated flows via MAE and regularizes workload, idle time, and makespan, yielding 5–28% higher accuracy vs. heuristic baselines.
- Enforced feasibility by predicting a full Gantt chart under single-server, non-overlap, duration, and ordering constraints, contrasting with unconstrained per-customer ML timelines.
- Integrated Dynamic Time Warping (DTW) via Dynamic Programming (DP) to quantify the similarity between sequences of varying lengths.
- Selected a representative schedule by choosing the simulated realization with the minimal objective as a feasible, sub-optimal prototype and warm start for the optimizer.

Invited Talks

Oct 2025 "Dynamic Interval Prediction for Patient Wait Times In a Cancer-Care Facility," INFORMS Annual Meeting, Atlanta, GA.

Jul 2024 "Dynamic Interval Prediction for Patient Wait Times In a Cancer-Care Facility," INFORMS MSOM Conference, Minneapolis, MN.

Aug 2023 "Physician Rostering Problem with Downstream Capacity Constraints," INFORMS Healthcare Conference, Toronto, CA.

Skills

- Python, R, MATLAB, SQL, VBA, Gurobi; PySpark, Databricks, Snowflake; Scikit-learn, TensorFlow, Keras, NLTK.
- Passed CFA Level 1 Exam, CFA Institute, 2018