Chicago, IL
jyontika@u.northwestern.edu
215-479-8708

Jyontika Kapoor

Statistics PhD Student

https://jyontika.com
https://linkedin.com/in/jyontika
https://github.com/jyontika

Statistics PhD student focused on applied data science and behavioral analytics, with experience using real-world health, consumer, and operational data to build predictive models and decision-ready insights.

EDUCATION

Doctor of Philosophy (PhD) in Statistics and Data Science | *Northwestern University* Expected June 2029

Bachelor of Arts (BA) in Data Science | *Wellesley College* | GPA 3.82 May 2024

TECHNICAL SKILLS

Languages & Tools: Python, R, Git/GitHub, LaTeX, Tableau

Libraries: pandas, NumPy, scikit-learn, PyTorch, TensorFlow, tidyverse, mgcv

Methods: regression, machine learning, causal inference, missing data, survival analysis, model interpretation

SELECTED PROJECTS

Flight Delay Forecasting and Explainable ML | Evanston, IL September 2025

- Built predictive models on ~150,000 departures from Chicago airports to forecast arrival delays and flag high-impact delays, supporting scheduling efficiency, resource allocation, and passenger experience.
- Used model explainability to identify factors driving arrival-delay risk, including runway/taxi congestion and upstream departure disruptions, supporting decisions around scheduling, staffing, and customer communication.
- Created a random forest meta-classifier to choose between the model predictions for each flight, using model uncertainty features to improve reliability for operational planning and delay-risk forecasting.

Foreclosure Housing Price Prediction | Evanston, IL May 2025

- Built a predictive modeling pipeline for 8,000+ Illinois foreclosure sales, reducing prediction error from $134K to $91.5K using regression, tree-based methods, and neural networks to support real estate valuation and risk assessment.
- Mapped county-level prices and prediction errors to identify geographic pricing disparities and less reliable prediction regions, revealing location-based risk in foreclosure valuations.
- Used transfer learning to adapt Illinois-trained models to a smaller Wisconsin dataset, reducing prediction error from $177K to $99K and demonstrating how larger market datasets can improve small-data predictions.

Evanston Restaurant Recommender System | Evanston, IL April 2025

- Developed a restaurant recommendation system using 1,500+ reviews across 68 local restaurants to help users find relevant dining options faster, improving personalized discovery and engagement on recommendation-based platforms.
- Built a hybrid restaurant recommendation pipeline combining restaurant similarity, user-rating patterns, and review-text signals to replace generic popularity rankings with more personalized recommendations.
- Used Bayesian shrinkage and sentence-transformer embeddings to improve ranking reliability and reduce prediction error by 42% compared to demographic-only models, improving recommendation quality and user trust.

RESEARCH EXPERIENCE

Tetel Laboratory at Wellesley College | Lead Statistical Research Assistant | Wellesley, MA March 2023 – October 2024

- Led statistical analysis for a Mayo Clinic collaborative women's health study, integrating multimodal longitudinal health data from participant surveys, wearable activity data, food logs, and microbiome outcomes.
- Built analysis-ready datasets, assessed missingness, ran exploratory statistical tests, and developed a new causal inference framework to account for dependent observations in longitudinal health data.

Machine Learning Laboratory at Tufts University | Research Intern | Medford, MA June 2023 – August 2023

- Built an end-to-end Cook County, IL geospatial validation pipeline in Python, extending a Massachusetts-based opioid overdose prediction project to support external validation and public health planning.
- Adapted an outdated TensorFlow neural network architecture by reconfiguring model parameters for spatiotemporal opioid overdose prediction, enabling reuse of existing ML infrastructure on new geospatial data.