Yash Khaitan

Ashoka University, Haryana, India
yash.khaitan_ug2023@ashoka.edu.in — +91 74340 88328 — LinkedIn

EDUCATION

Ashoka University, Haryana, India — Aug 2023 — Aug 2027
BA. Economics with Research (Minor in Mathematics) — Major & Minor GPA: 4.00/4.00
Dean's list in all semesters — Cumulative GPA: 3.97/4.00

EXPERIENCE

Centre for Effective Governance of Indian States — *Data Science Intern* — New Delhi, India | May 2025 — July 2025

- Led a **predictive modeling** study to forecast child growth outcomes using contemporary ML methods, integrated anthropometric and **geospatial data**, engineered relevant features, and **authored a whitepaper** on results and policy implications for MoWCD.
- Applied econometric and spatial analysis techniques on **NFHS-5 and PLFS microdata** to quantify regional disparities in **socio-economic indicators across regions**, using Stata, Python, and QGIS for data preprocessing, modeling, and visualization.
- Conducted a systematic review and process mapping of the **Support for Statistical Strengthening** scheme (MoSPI), developing **consultation tools** and state-specific briefs to recommend structural and institutional improvements for enhanced implementation.

Centre for Data Science and Analytics — *Data Engineering Intern* — Haryana, India | Dec 2024 — Jan 2025

- Designed and deployed end-to-end **data pipeline**s for Ashoka University's **DataLake** project, enabling seamless integration of structured (Excel) and unstructured (image) data into a **PostgreSQL database** with dynamic schema handling.
- Developed a **Django-based frontend** interface to facilitate the upload and management of Excel and image data, streamlining researcher workflows and improving data submission efficiency.
- Implemented data validation, preprocessing, and visualization workflows to ensure metadata consistency, enhance analytical readiness, and improve data quality, accessibility, and usability.

Nature Conservation Foundation — *Spatial Data Science Intern* — Arunachal Pradesh | May 2025 — Present

- Worked with PhD researchers and Professors to analyse Landsat and Sentinel imagery via **Google Earth Engine** to map long-term deforestation and land-use transitions, creating baseline environmental insights for the previously understudied Northeast region.
- Produced high-resolution geospatial datasets by training supervised **ML classifiers** on satellite imagery and ground-truth points to enable quantitative analysis of forest loss patterns, contributing to the core dataset for a series of forthcoming papers.
- Using **econometric techniques** to identify socio-environmental deforestation drivers and evaluate forest management regimes.

Krea University, Prof Rohan Gudibande — *NLP Analyst* — Chennai, India | May 2025 — Present

- Developed a pipeline to scrape details of 1,000+ land conflict events and 5,000+ related news articles via LLM-generated web search queries, ingesting the data into a **SQL database** to support analysis of the **economic and financial impacts** of land disputes.
- Built and deployed a custom **NLP pipeline** to classify conflict events into predefined categories based on textual features, **automating the classification of 1,000+ reports** with over 95% accuracy using Python libraries including spaCy and scikit-learn.
- Conducted preliminary analysis and created **visualizations** using Python to explore and summarize patterns in the collected data.

Ashoka University, Prof Parush Arora — *Econometrics Research Assistant* — Sonipat, India | March 2025 — May 2025

- Conducted simulation experiments on MATLAB to evaluate Bayesian opinion pooling approaches for **inflation density forecasting**, contributing to the working paper "Regularized Opinion Pools for Density Forecasts Under a Bayesian-Inspired Framework."
- Implemented **Markov Chain Monte Carlo methods** to estimate model parameters and compare multiple pooling specifications.
- Tested over 100 model configurations by systematically varying regularization techniques and prior distributions on datasets from the Survey of Professional Forecasters **(Federal Reserve Bank of Philadelphia)** to identify optimal ensemble forecasting strategies.

POSITIONS

Director of Research & Data Insights — Ashoka Data Society — September 2024 — Present

- Lead a 20-member team on three interdisciplinary data projects in financial analytics, digital health, and remote sensing.
- Conducted workshops on SQL, financial modelling, and spatial analytics to enhance the team's data-driven research skills.

Director of Finance — 180 Degrees Consulting Ashoka — October 2023 — December 2024

- Led outreach for finance and managed ongoing relationships with partners, **successfully securing funding** and support for events.
- Led a project with the Government of Punjab to explore labor market patterns and **design vocational skill-training pathways** for high school graduates, combining labor analysis, field research, and sector mapping to align training with regional needs.

PROJECTS

A Deep Learning and Econometrics Approach to Forecasting NIFTY Movements
Github Repository
Compiled financial and economic data from the **World Bank and Yahoo Finance** to train and compare various deep learning models (RNN, LSTM, GRU), rigorously benchmarking them to identify the most robust architecture for volatile time-series data. **Forecasted quarterly NIFTY movements** while employing econometric models to uncover key financial drivers and analyze macroeconomic shocks.

Portfolio Optimization Dashboard
Github Repository
Built a Portfolio Optimization Dashboard integrating the Yahoo Finance API to extract and process **real-time financial data** for portfolio construction. Applied quantitative methods and Monte Carlo simulations (10,000 portfolios) to **optimize asset allocation** using Sharpe, Sortino, CVaR, and other risk-adjusted metrics, aimed at enhancing investment decision-making and portfolio risk-management.

SKILLS & OTHER DETAILS

Programming & Languages: Python, R, Stata, Google Earth Engine, SQL, LaTeX, MATLAB, HTML, CSS, JavaScript
Technical Skills: Machine Learning, Deep Learning, Financial Modelling, Time Series Forecasting, NLP, GIS, Remote Sensing
Hobbies & Interests: Guitar, Piano, Table-Tennis, Badminton, Cycling
Languages: English, Hindi, Gujarati (Elementary), Spanish (Elementary)