

# Harsh Gupta

Bloomington, IN

930-904-3651 – [hagupta@iu.edu](mailto:hagupta@iu.edu) – [linkedin.com/in/harshg1347](https://www.linkedin.com/in/harshg1347) – [github.com/Harsh1347](https://github.com/Harsh1347) – [Google Scholar](#) – [YouTube](#)

## EDUCATION

### Indiana University Bloomington

Aug 2024 – May 2026 (Expected)

*Master of Science in Data Science (Current GPA: 4/4)*

*Bloomington, IN*

*Courses: Fundamentals of LLM, Computer Vision, Applied ML, Data Mining, Statistics*

### Manipal University

Jul 2017 – May 2021

*Bachelor of Technology in Electronics and Communication, Minor in Data Science (GPA: 8.13/10)*

*Manipal, India*

## EXPERIENCE

### Indiana University Bloomington

Oct 2024 - Present

*Machine Learning Engineer* | *Azure, Python, SQL, Tableau, LLM,*

*Bloomington, IN*

- Built an **LLM-powered virtual teacher** using a RAG chatbot in Azure AI Studio to automate student query resolution and reduce educator workload.
- Developed **time series forecasting pipeline** in Azure, optimizing dining resource allocation from transactional and **reduced food wastage by 30%**.
- **Boosted accuracy by 21%** by building a data ingestion & feature engineering pipeline in Azure Data Factory.
- **Accelerated** prediction pipeline by **6×** using distributed Azure compute, **reducing residual error to 5%**.

### Deloitte - Strategy & Analytics

Sep 2021 - Jul 2024

*Data Engineer - Financial Sector* | *AWS, Spark, OracleDB, Postgres*

*Bengaluru, India*

- Designed **AWS-based ETL** infrastructure processing **15M+ records daily** across multiple business lines.
- Built an **end-to-end ETL pipeline** (Glue, Lambda, Aurora) integrating **500+ fields** from new source systems.
- Built a scalable data validation and logging framework to ensure ETL data integrity, **reducing debugging and issue resolution time by 33%**
- **Conducted training for 50+** new hires on Python and ETL concepts, accelerating onboarding time.

*Data Scientist - Customer Strategy & Pricing* | *Python, R, Machine Learning, AWS S3, Tableau, Excel*

- Implemented a **machine learning-based price recommendation strategy**, projecting a **\$1M–\$3M profit increase** across 17 regions and 5,000+ stores while maintaining guest count **loss under 1%**.
- Engineered a modular software architecture using Python and SQL, enabling dynamic pricing simulations and real-time analysis of price elasticity across different markets.
- **Streamlined** the Pricing-as-a-Service **pipeline, boosting efficiency by 30%** and supporting client acquisition through success story presentations.

### Samsung Research Institute


Jan 2021 - Jun 2021

*Research & Development Student Intern* | *Python, Numpy, OpenCV*

*Remote*

- Developed a computer vision algorithm to enhance mobile night sky photography, published a research paper , and received an Excellence Award for innovation.

## PROJECTS

**Premium Sneakers Discord Bot**  | *Python, MongoDB, Selenium, Heroku*

- Built a serverless real-time sneaker bot (Heroku + MongoDB) for a client, **generating \$300K** in sales across 1,000+ pairs.

**MockChain - AI Agent powered Interview Prep**  | *LLM, Langchain, Python, Ollama*

- Multi-agent AI-powered mock interview platform that simulates personalized interviews and delivers actionable feedback based on your resume and job description.

## TECHNICAL SKILLS

**Programming:** Python, R, PySpark, SQL

**Cloud & Data Platforms:** AWS, Azure, OracleDB, Postgres, Amazon Redshift, Amazon Aurora, Hadoop

**Machine Learning & MLOps:** Regression, Clustering, Computer Vision, Time Series Forecasting, Deep Learning, Recommendation Systems, Feature Engineering, Model Deployment, CI/CD Pipelines

**Data Visualization Tools:** Tableau, PowerBI, MS Excel, Matplotlib, Seaborn