

# Ansh Kumar Dev

520-599-5589 | [anshkumardev@gmail.com](mailto:anshkumardev@gmail.com) | [linkedin.com/in/ansh-kumar-dev/](https://linkedin.com/in/ansh-kumar-dev/) | [github.com/Anshkumardev](https://github.com/Anshkumardev)

## EDUCATION

### University of Arizona

Master of Science in Data Science

Tucson, AZ

Aug. 2023 – May 2025

### Gautam Buddha University

B.Tech. in Artificial Intelligence

Greater Noida, India

Aug. 2019 – May 2023

## EXPERIENCE

### Data Scientist (Graduate Research Assistant)

Aug 2024 – Jan 2025

University of Arizona

Tucson, AZ

- Analyzed large-scale satellite and sensor data using **Python (Pandas, NumPy)** and **SQL**; identified trends and outliers in time-series signals to improve groundwater forecasts.
- Performed **statistical modeling**, **regression analysis**, and hypothesis testing to validate environmental predictions; cut 12-month RMSE by **18%**.
- Created visualizations in **Jupyter** and presented findings to technical and non-technical audiences; supported decisions on water resource planning.
- Mentored 15+ students on **data wrangling**, **EDA**, and best practices in storytelling through data for research publications.

### Research Data Scientist

Jun 2022 – Nov 2022

Gautam Buddha University

Greater Noida, India

- Led analysis of large IoT datasets for trend and anomaly detection using **Python** and **statistical time-series modeling**; improved reliability by 20%.
- Conducted **root cause analysis** and used regression-based insights to optimize network throughput; reduced latency by 25%.
- Collaborated on algorithm development and published results in Elsevier; emphasized **data storytelling** for policy and technical audiences.

### AI Engineer Intern

Apr 2022 – Jul 2022

Merkletree Technologies

Delhi, India

- Performed A/B testing and impact analysis to optimize face-auth models; improved accuracy by **25%** through fine-tuned model experimentation.
- Used **Python** and **SQL** for data extraction, transformation, and visualization in model monitoring dashboards.
- Collaborated across data, engineering, and mobile teams to align ML outputs with user experience and product performance.

### MLOps Intern

May 2021 – Aug 2021

LinuxWorld Informatics Pvt Ltd

Jaipur, India

- Built and deployed predictive analytics pipelines to classify diabetes risk using historical health data; improved model F1 by **15%**.
- Conducted **Bayesian hyperparameter tuning**, performance comparisons, and regression diagnostics across experiments.
- Created **Power BI** dashboards and reports to present insights to stakeholders; simplified complex technical results into clear business actions.

## PROJECTS

### Healthcare Accessibility & Risk Mapping Platform | [Project Link](#)

Apr 2025

- Built a distributed **GeoPandas + OSMnx** pipeline unifying Geospatial, socioeconomic & clinical data across 3,000+ U.S. census tracts to identify underserved areas for policy action.
- Developed a **LangChain** multi-agent LLM workflow with **OpenAI API** and **FAISS RAG** and a **Streamlit** choropleth dashboard, enabling real-time Q&A and evidence-based insights.

### E-commerce Analytics Lakehouse | [Project Link](#)

Mar 2025

- Engineered an **Azure Data Factory to Databricks (PySpark)** ETL ingesting 1.2 TB/day of order & clickstream data to power near-real-time KPI dashboards.
- Modeled star-schema views in **Synapse SQL** and built live **Power BI** reports, Reduce manual reporting by 90%

## TECHNICAL SKILLS

**Languages/ML:** Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, OpenCV), R, SQL, C++

**Models:** RoBERTa, SVM, Decision Trees, K-Means, PCA, Forecasting & Statistical Analysis

**Cloud & MLOps:** AWS (EKS, EC2, S3), Azure (ADF, Synapse, Databricks), GCP, Docker, Kubernetes, Spark, CI/CD, MLflow

**Tools & APIs:** Flask, REST APIs, TensorBoard, Git, Linux, Power BI, Streamlit