Ansh Kumar Dev

520-599-5589 | anshkumardev@gmail.com | linkedin.com/in/ansh-kumar-dev/ | github.com/Anshkumardev

Education

University of Arizona — Tucson, AZ
Master of Science in Data Science — *Aug. 2023 – May 2025*

Gautam Buddha University — Greater Noida, India
B.Tech. in Artificial Intelligence — *Aug. 2019 – May 2023*

Experience

AI Engineer Intern — Jun 2025 – Present
Right Skale — *Pleasanton, CA*

- Built a **multi-modal AI** pipeline for SEC 10-Q and balance-sheet ingestion using containerized microservices (LlamaParse/Extract, vector-DB), processing **100+ filings/hour**.
- Developed **GenAI-powered assistants** with Bedrock Llama 3 and prompt chaining to summarize filings and extract business-critical insights for stakeholders.
- Designed fair and explainable **predictive models** for entity normalization and format classification, enabling consistent cross-firm financial comparison.
- Orchestrated scalable **machine learning pipelines** using AWS AI Agents and Step Functions, achieving **90% automation**; ensured CI/CD, observability, and SOX-compliant deployment via CloudWatch, X-Ray, and IAM.
- Partnered with product and engineering teams to align GenAI model outputs with end-user workflows, refining tools through iterative **feedback loops** and internal experimentation.

Data Scientist (Graduate Research Assistant) — Aug 2024 – Jan 2025
University of Arizona — *Tucson, AZ*

- Developed an end-to-end **Python (Pandas, SQL)** ETL pipeline to ingest and preprocess satellite imagery & sensor data; engineered time-series features for a CV–RNN model, reducing aquifer forecast RMSE by **18%**.
- Conducted exploratory analysis and **data wrangling** in **Jupyter** and SQL; mentored **15+ students** on **statistical modeling** and ML experimentation best practices.
- Validated forecasting models using **Bayesian hyperparameter tuning**, cross-validation, and **significance testing** to ensure reproducible and **bias-mitigated predictions** for water resource planning.
- Collaborated with academic stakeholders to define outcome-driven modeling goals aligned with long-term environmental policy decisions.

Research Data Scientist — Jun 2022 – Nov 2022
Gautam Buddha University — *Greater Noida, India*

- Applied **statistical time-series modeling** to analyze IoT traffic, detecting anomalies and improving throughput by **20%** with latency reduction of **25%**.
- Designed and benchmarked **QoS-aware streaming algorithms** with rigorous statistical evaluation, published in e-Prime (Elsevier).
- Implemented real-time anomaly detection protocols in Python and C++; elevated model accuracy by **40%** through **experiment-driven development** and root-cause analysis (LICTIC 2023).

AI Engineer Intern — Apr 2022 – Jul 2022
Merkletree Technologies — *Delhi, India*

- Enhanced face-auth model accuracy by **25%** via TensorFlow Faster R-CNN fine-tuning and targeted data augmentation.
- Productionized model using Docker & Kubernetes, reducing inference latency **35%** and cloud costs **20%**.
- Delivered frictionless biometric login for **50k+ users** via on-device inference integration and Flask REST APIs.
- Streamlined deployments with MLflow-tracked CI/CD pipelines to ensure reproducibility and traceability.

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.

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