

Ansh Kumar Dev

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

EDUCATION

| | |
|--|--|
| University of Arizona <i>Master of Science in Data Science</i> | Tucson, AZ Aug. 2023 – May 2025 |
| Gautam Buddha University <i>B.Tech. in Artificial Intelligence</i> | Greater Noida, India Aug. 2019 – May 2023 |

EXPERIENCE

| | |
|--|---|
| Jr. Data Analyst Intern <i>Right Skale</i> | Jun 2025 – Present Pleasanton, CA |
| <ul style="list-style-type: none">Download, organize, rename, and version-control SEC 10-Q and balance-sheet filings, managing 1M+ documents across a confidential data processing pipeline.Execute data entry and document management workflows using custom microservices and Python scripts; maintain data accuracy and integrity.Create and reconcile transaction-level financial data in Excel and SQL for ingestion into cloud-based vector databases and reporting tools.Circulate and track authorization documents and escalation notices for error handling and system-level issue triage.Collaborate with client-facing teams to resolve data issues, ensuring timely communication and escalation protocols. | |
| Data Scientist (Graduate Research Assistant) <i>University of Arizona</i> | Aug 2024 – Jan 2025 Tucson, AZ |
| <ul style="list-style-type: none">Built automated ETL pipelines in Python and SQL to process satellite and sensor data; delivered datasets with validated accuracy and timely updates.Performed data wrangling, exploratory analysis, and documentation for data onboarding into ML modeling tools.Maintained strict data confidentiality while validating datasets for time-series forecasting and publishing reproducible results. | |
| Research Data Scientist <i>Gautam Buddha University</i> | Jun 2022 – Nov 2022 Greater Noida, India |
| <ul style="list-style-type: none">Conducted high-accuracy data processing, cleaning, and organization of IoT logs for anomaly detection and streaming analysis.Documented protocols for data quality assurance and reconciliation of statistical anomalies across time-series logs.Protected confidential information during research and publication processes for academic and industrial data. | |
| AI Engineer Intern <i>Merkletree Technologies</i> | Apr 2022 – Jul 2022 Delhi, India |
| <ul style="list-style-type: none">Prepared and documented model deployment data in Excel and JSON, ensuring consistent versioning and reproducibility of results.Participated in data onboarding and reconciliation of facial recognition datasets, improving pipeline accuracy and reducing errors.Collaborated with team leads to validate inputs, update records, and maintain data handling standards in production. | |

PROJECTS

| | |
|---|----------|
| Healthcare Accessibility & Risk Mapping Platform <i>Project Link</i> | Apr 2025 |
| <ul style="list-style-type: none">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 <i>Project Link</i> | Mar 2025 |
| <ul style="list-style-type: none">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 |