

Ansh Kumar Dev

520-599-5589 | anshkumardev@gmail.com | 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

AI Engineer Intern <i>Right Skale</i>	Jun 2025 – Present Pleasanton, CA
<ul style="list-style-type: none">Collaborated with cross-functional teams to deliver customized AI-powered analytics pipelines, ensuring accurate onboarding and preprocessing of SEC 10-Q filings and balance sheets for downstream analytics.Developed a scalable, containerized ETL pipeline (LlamaParse, LlamaExtract) integrated with AWS Step Functions and tracked via JIRA, automating 90% of ingestion-to-retrieval workflows.Built and deployed a RAG-based reporting system using OpenSearch and Bedrock Llama 3, enabling real-time insights and semantic document search across 1M+ data chunks.Partnered with implementation teams to resolve data quality issues and validate transformed data outputs; maintained data pipelines using SQL, Python, and GitHub Actions.Created modular analytics components enabling reuse across clients; improved data reporting accuracy by integrating stakeholder feedback into prompt design and validation workflows.	
Data Scientist (Graduate Research Assistant) <i>University of Arizona</i>	Aug 2024 – Jan 2025 Tucson, AZ
<ul style="list-style-type: none">Designed an end-to-end ETL pipeline in Python/SQL for satellite and sensor data used in environmental forecasting; improved forecast accuracy by 18% through robust feature engineering.Supported onboarding of environmental data for research use, ensuring data validation, cleaning, and transformation for reproducibility and analytic workflows.Created visual dashboards in Jupyter and Power BI for reporting hydrologic patterns and insights to academic stakeholders.Led reporting template creation and data visualization training for students and researchers, increasing efficiency in report generation and presentation.	
Research Data Scientist <i>Gautam Buddha University</i>	Jun 2022 – Nov 2022 Greater Noida, India
<ul style="list-style-type: none">Analyzed network and sensor data using statistical modeling and time-series analysis to uncover anomalies and optimize QoS parameters.Authored reporting protocols and dashboards to visualize data quality metrics, supporting compliance with data transmission standards.Published research on data streaming reliability and built tools to benchmark transmission performance under different control models (e-Prime, Elsevier).Collaborated with implementation and faculty teams to onboard and transform new IoT datasets from lab systems into structured analytic pipelines.	

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