

VANSHIKA NAMDEV

+1 (217) 721-6543 | vanshikanamdev9@gmail.com | [linkedin.com/in/vanshikanamdev](https://www.linkedin.com/in/vanshikanamdev) | github.com/vanshu25 | vanshu25.github.io

EDUCATION

University of Illinois Urbana–Champaign | *Master of Science, Statistics*

Aug 2025 - May 2027

KIET Group of Institutions | *Bachelor of Technology, Computer Science*

Jul 2019 - Jun 2023

SKILLS

- **Languages:** Python, SQL, R, Java, C, JavaScript
- **Machine Learning & AI:** PyTorch, TensorFlow, JAX, TF-Serving, Scikit-learn, Hugging Face Transformers, LLM tooling, Semantic Search, RAG, LLM Fine-tuning, NLP, Prompt Engineering, LLM Evaluation, Statistical Methods
- **Data Analytics & Systems:** Distributed Training, Large-Scale Datasets, ETL Pipelines, Apache Spark, Airflow, Embedding-based Retrieval, Experiment Tracking, MLflow, PostgreSQL, MongoDB, Data Analytics, Pandas, Microsoft Azure, Snowflake
- **Visualization & Deployment:** FastAPI, Flask, REST APIs, AWS (EC2, S3, Lambda), Git, Power BI, React, Streamlit

EXPERIENCE

SSA Marine (Carrix)

June 2026 – Aug 2026

Data Science Intern

Seattle, US

- Conducted Python-based geospatial and statistical analysis (hypothesis testing, confidence intervals) on 4M+ row container dwell-time datasets in Snowflake/Snowpark, informing pricing strategy and contributing to an estimated \$5–7M revenue increase
- Built a React-based AI research tool for Market & Acquisition analysis, combining an LLM-driven scoring engine with custom logic and formulas to support strategic decision-making for the Strategy team — improving research efficiency by 70%

Mercedes-Benz Research & Development India

Aug 2023 - Jul 2025

Data Scientist

Bengaluru, India

- Built an LLM-powered AI agent leveraging GPT-4 and a transformer-based RAG architecture to process 200+ compliance rulebooks, enabling automated report generation, high-risk flagging, and real-time query resolution — taken from POC to full production; awarded Gold Award
- Developed ensemble ML model (scikit-learn, XGBoost, SVM), automating 90% of manual classification tasks across 10+ carlines and reducing evaluation time from weeks to hours; applied pandas-based feature engineering and data analysis to drive model performance insights
- Co-authored SAE Mobilus publication on AI-driven logistics screening using BERT/transformer models achieving 83–85% accuracy on technical document classification

Nunam

Feb 2023 - Jul 2023

Data Engineer

Bengaluru, India

- Designed PostgreSQL and Apache Airflow pipelines automating daily ingestion of 10,000+ energy data points, reducing manual processing by 80% and enabling real-time battery performance monitoring.
- Improved ETL data quality by 60% via pandas/NumPy validation; deployed FastAPI microservices on AWS with TF-Serving for scalable model serving across 5+ downstream analytics teams.

University of Warwick

Jun 2022 - Aug 2022

Data Science for Social Good Fellow

Coventry, UK

- Selected as 1 of 15 fellows globally; built SCANR ensemble model (92.8% accuracy, 47.8% F2-score) for student dropout risk prediction for Buckinghamshire Council.
- Awarded Runner-up at National Data Science & AI Showcase among 50+ competing projects across UK universities; presented to 200+ stakeholders at Datafest (The Shard, London).

PROJECTS

Farmer's Friend – RAG-based Livestock Triage App

- Built end-to-end RAG pipeline in Python using Claude Sonnet, LangChain, and ChromaDB with Pandas-based preprocessing and optimized chunking for real-time veterinary query resolution over domain-specific livestock corpora
- Measured and optimized query latency, achieving sub-second retrieval times, and demonstrated live at UIUC hackathon with judges from agricultural and ML industry to win first place

SCANR – NEET Risk Prediction System | [GitHub](#)

- Developed ensemble model in Python combining Random Forest, XGBoost, and Logistic Regression to achieve 92.8% accuracy and 47.8% F2-score identifying at-risk students for Buckinghamshire Council
- Engineered 50+ features from 5 years of student records using Pandas on a large dataset; implemented a DVC-based reproducible ML pipeline with automated hyperparameter tuning and designed interactive dashboards in Power BI

PATENTS & PUBLICATIONS

- AI-Driven Logistics Impact Screening for Engineering Changes in the Automotive Supply Chain. *SAE Mobilus publication, 83–85% accuracy.*
Link: saemobilus.sae.org