Navya Gupta

+1(217)904-2122 | navya.kgupta2@gmail.com | https://www.linkedin.com/in/navyag20/ | https://navyagpt.github.io/

Summary: Data Scientist (3+ YOE) focused on multimodal ML and product measurement. Shipped transformer and LLM systems and scalable pipelines on large datasets; partnered with PM/engineering to drive product-led growth decisions via experimentation and quantitative insights.

Work Experience

Kalderos Inc. Chicago, IL, U.S.

Data Scientist   *June 2025 - Present*
- Improved entity resolution quality by 5x by fine-tuning a transformer encoder on labeled pairs and optimizing ranking via pairwise preference learning.
- Built and deployed a scalable anomaly detection and scoring pipeline (Random Forest baseline + evaluation harness) for large-scale healthcare claims data, driving a 46% lift in product revenue vs. rule-based analytical baselines.
- Owned end-to-end measurement, constructed reproducible SQL analysis datasets, defined success metrics/guardrailsran cross-validation (F1/ROC-AUC, calibration, ablations), and tuned thresholds to align with product risk and cost tradeoffs.
- Monitored models with PM/data/platform: implemented data quality checks (schema/volume, null/outliers, leakage), built telemetry for drift/perf decay (distribution, calibration drift), and authored docs/reports that influenced roadmap prioritization.

Bayer Research & Development, Champaign, IL, U.S.

Data Science Co-op/Intern   *May 2024 - May 2025*
- Built distributed Spark/Dask data pipeline for 10TB+ UAV imagery processing, enabling large-scale multimodal sensing analysis.
- Built a spatiotemporal Vision Transformer for image time series to predict crop maturity, reducing RMSE by 32% vs. ConvLSTM.
- Developed Spearman correlation statistical model to quantify moisture-yield relationships saving $20M in operational cost.
- Established CI/CD using Jenkins and Docker for end-to-end model deployment, reducing deployment time from 48h to 3h.

University of Illinois Urbana-Champaign, Champaign, IL, U.S.

Graduate Research Assistant- Machine Learning   *Jan 2024 - May 2024*
- Developed KDE++, a tree-based data structure for on-device learning of probability distributions in large-scale gene datasets; implemented parallel processing and data handling strategies for big data, resulting in a 45x computational speedup.
- Developed NLP pipelines by fine-tuning LLMs, improving named entity recognition (NER) accuracy by 34%.
- Engineered point-cloud architecture for spatial data analysis, achieving 29% higher precision, bridging domain expertise with ML.

Citrix Research & Development, Bangalore, India

Software Development Engineer   *Jul 2022 - Jul 2023*
- Enhanced Citrix Workspace App(CWA) for HTML5 and ChromeOS through targeted development, utilizing user data analysis and cloud automation, driving client base growth to 1M monthly active users while ensuring regulatory compliancy.
- Optimized client-side printing algorithms, resulting in a 20% reduction in delays; recognized as Q1 2023's best feature.
- Applied A/B testing framework to assess log pipelines to improve feature performance across 1M+ users.

Software Development Intern   *Jan 2022 - Jun 2022*
- Developed visualization and analytics optimization with Google Analytics 4 (GA4) adopted by 10+ CWA teams.

Education

University of Illinois Urbana-Champaign (UIUC)

Master of Science, Statistics, GPA: 4.0/4.0   *Aug 2023 - May 2025*
- Achievements: Awarded Excellent Teaching Assistant (2024)

Manipal Institute of Technology

Bachelor of Technology, Computer Science & Engineering, (Minor: Big Data) , CGPA: 9.4/10.0   *Jul 2018 - Jul 2022*
- Achievements: Dean's List : 2019, 2021, 2022, Department Highest Achiever for the year 2021-2022.

Skills

Programming Languages: Python, Java, C/C++, SQL, R, JavaScript, MATLAB
Product Analytics & Experimentation: A/B testing, metric design, calibration, offline/online evaluation, hypothesis testing
ML Frameworks: JAX, PyTorch, TensorFlow, TinyML, pandas, NumPy, MLX, SciPy, LangChain, Hugging Face, MicroPython
RAG & NLP: BM25, TF-IDF, BERT, NER, Transformers, Retrieval Pipelines, Chain of Thought, Document Chunking
Cloud & Tools: AWS, SageMaker, MLflow, Power BI, Tableau, Azure, REST APIs, Postman
DevOps & Deployment: Git, Docker, Jenkins

Projects (View all at: https://github.com/navyagpt)

RAG-based Investment Insight Quality Evaluation (Code: https://github.com/navyagpt/rags-pir)   *UIUC*
- Designed an automated evaluation framework for retrieval and summarization quality (NER-based entity coverage, risk scoring, backtests vs. historical events) to quantify quality lift across variants. (Skills: RAGs, LangChain, TF-IDF, Gemini API)

MediVision (Code: https://github.com/navyagpt/MediVision)   *Bajaj Finance HackRx - Hackathon*
- Built an NLP extraction and classification pipeline for medical bills using DistilBERT achieving 93.67% accuracy; created error analysis and calibration checks to reduce high-cost misclassifications. (Skills: Reasoning, Visual Question Answering)