Navya Gupta

+12179042122 | navya.kgupta2@gmail.com | https://www.linkedin.com/in/navyag20/

Work Experience

Kalderos Inc. Chicago, IL, U.S.

Research Data Scientist *June 2025 - Present*
- Improved customer and merchant address matching accuracy from 6% to 30% by fine-tuning a transformer encoder on labeled pairs and added LLM-based pairwise preference learning to sharpen match decisions.
- Delivered a 46% lift in recovered revenue by shipping an anomaly-detection pipeline that combines Random Forests on tabular signals with text embeddings from event descriptions.
- Drove research-to-production transfer with PM, data, and platform teams; documented on system architecture, preference optimization, and distillation practices used to launch and operate models at scale.

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

Machine Learning Intern *May 2024 - May 2025*
- Deployed Apache Spark/Dask data pipeline for 10TB+ UAV imagery processing, enabling large-scale agricultural analysis.
- Leveraged Vision Transformer (ViT) models for crop maturity prediction, boosting forecasting accuracy by 32%.
- Developed Spearman correlation model to quantify moisture-yield relationships generating $20M in operational cost savings.
- Established CI/CD pipeline using Jenkins and Docker for 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 efficient probability distribution predictions in large gene datasets, implementing parallel processing and advanced data handling strategies for big data, resulting in a 45x computational speedup.
- Developed and deployed NLP pipelines using fine-tuned 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 maintaining regulatory compliance.
- Optimized client-side printing algorithms, resulting in a 20% reduction in delays; recognized as Q1 2023's best feature.
- Applied A/B testing frameworks to assess feature performance across 1M+ user base, leading to iterative product improvements.

Citrix Research & Development, Bangalore, India

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) , GPA: 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
ML Frameworks: JAX, PyTorch, TensorFlow, scikit-learn, pandas, NumPy, SciPy, LangChain, Hugging Face, OpenAI API
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, A/B Testing
DevOps & Deployment: Git, Docker, Jenkins

Projects (View all at: https://github.com/kalderos-ngupta)

Personalized Investment Reports with RAG AI (Code: https://github.com/kalderos-ngupta/rags-pir) *UIUC*
- Engineered a RAG pipeline using TF-IDF, combined with LangChain's memory augmented vector retrieval to extract and rank.
- Designed a report generation system with NER for company-specific insights and BERT-based analysis for risk assessment, using OpenAI's GPT models to improve investment insight accuracy by 22% in backtesting against real market events.

Fraud Detection using Multi-Agent Systems (Code: https://github.com/kalderos-ngupta/fraud-detection) *UIUC*
- Developed a multi-agent fraud detection system utilizing Graph-based Transaction Analysis with spectral clustering and Chain of Thought (CoT) to detect anomalous transaction patterns, reducing false positives by 28% in simulated banking datasets.

TextQuraton (Code: https://github.com/kalderos-ngupta/TextQuraton) *Bajaj Finance HackRx - Hackathon*
- Developed an app leveraging DistilBERT with a context and attention module to parse medical bills, achieving a 93.67% accuracy.
- Implemented spatial transformers and RCNN with bidirectional LSTM to process handwritten bills, enabling fraud detection.