SHASHWAT KUMAR

New York, USA | +1 (917) 295-7422 | shashwat.kumar@columbia.edu |linkedin.com/in/shashwatkumar53/

EDUCATION

Columbia University in the City of New York — New York, United States
Master of Science in Data Science (GPA: 3.75/4) — Aug 2024 - Dec 2025

Vellore Institute of Technology — Vellore, India
Bachelor of Technology - Computer Science and Engineering (GPA: 9.06/10) — Aug 2020 - Jun 2024

WORK EXPERIENCE

NOKIA — New York, NY
AI & Data Intelligence Co-op — Sep 2025 - Present

- Building an unsupervised RCA engine for T-Mobile alarms using temporal clustering, graph-based topology metrics, and rule-based fault segregation, with BERTopic for alarm classification and topic extraction.
- Engineering an agentic AI anomaly detector that integrates LLMs with RAG (FAISS + vector embeddings) to detect KPI anomalies, apply contextual reasoning over telecom data, and generate explainable root-cause narratives.

INTUIT: *AI & Data Science Intern* — Mountain View, CA
Spend Alert System — May 2025 - Aug 2025

- Deployed a Databricks ETL pipeline unifying S3 tables from 14 ad channels; computed weekly spend deltas vs rolling/YoY benchmarks and generated JSON payloads for LLM ingestion to track ~$300M media spend.
- Productionized an LLM spend-alert agent via Superglue; benchmarked GPT-4.0 vs O3, engineered prompts for multi-dimensional insights (publisher, bucket, product), and automated Slack alerts with 97% accuracy vs. SQL baseline.

Advertisement Creative Analysis

- Built a reproducible creative analysis pipeline with GCP/Drive asset retrieval and context-aware fuzzy matching (RapidFuzz + format/size checks), achieving 95%+ mapping accuracy across folders.
- Generated multimodal embeddings with SigLip (image), V-JEPA (video), MSClap (audio), and Jina AI (text); applied PCA and K-Means for clustering, followed by KeyBERT to analyze creative messaging themes at scale.

IBM x Vodafone Idea Limited — Pune, India
Artificial Intelligence Engineer Intern — Dec 2023 - Jun 2024

- Developed a Retrieval Augmented Generation chatbot with Mistral 7B LLM and Redis for efficient information retrieval and high-quality embeddings, reducing query response time by 75%.
- Integrated fine-tuned CogVLM, Tesseract OCR and PDF miners for data extraction with over 94% accuracy, further utilized Sentence Transformers and NLTK for tokenization and generating vector embeddings.

SKILLS

Coursework: Applied Machine Learning, Deep Learning-NLP, Statistical Inference & Modeling, Forecasting.
Programming languages: Python, SQL, C++, Java, MATLAB, R, Scala, JavaScript, HTML/CSS.
Libraries and Frameworks: PyTorch, TensorFlow, Transformers, Scikit-learn, Pandas, NumPy, OpenCV, Matplotlib.
Software and Tools: AWS(S3), GCP, Databricks, Superglue, PowerBI, Tableau, Git, Apache Spark (PySpark), MLflow.

ACADEMIC PROJECTS

KPMG x Columbia DSI – Intelligent Policy Analysis System — Aug 2025 – Dec 2025

- Implemented an LLM-RAG pipeline using BGE-M3 embeddings to automate Q&A across 3K+ Medicaid policy documents, achieving a 20% improvement in answer accuracy over baseline retrieval methods.
- Architected a hybrid retrieval framework with FAISS + Neo4j KG, deploying Llama-3.1-8B for reasoning-intensive queries and Qwen-2.5-7B for low-latency inference to deliver real-time, explainable policy insights.

Online Retail and Electricity Demand Forecasting — Jan 2025 - May 2025

- Designed product-based clustering (UMAP, MiniBatchKMeans) and forecasting pipeline on over a million retail transactions with models (Prophet, BiLSTM, GLM, SARIMA) to forecast daily and weekly sales at 7.65% MAPE.
- Optimized an electricity-usage forecasting pipeline with Savitzky–Golay smoothing, corrected train/test splits, and benchmarked BiLSTM, DeepAR, Prophet and Chronos to achieve 2.4% MAPE.