Ankit Kumar

14th May 2025 Graduation | Open to relocation | OPT EAD US Work Authorization (Until June 2028)
akumar37@gmu.edu | in/027kumarankit | github.com/ankitkr8540 | ankitkr8540.github.io/portfolio/

Technical Skills

Machine Learning & AI: PyTorch, TensorFlow, Scikit-learn, PySpark, LLMs(BERT, LLaMa), Hugging Face, RAG Systems(Sentence Transformers,FAISS), Recommendation System(ALS), Topic Modeling(LDA), Data Preprocessing(TF-IDF, NLTK)
Software & Cloud Engineering: AWS (Lambda, S3, CloudWatch, Glue, SageMaker, Bedrock, EMR, S3, OpenSearch Service, QuickSight), Python, SQL, ETL Pipelines, Web Scraping(Scrapy), Matplotlib, CI/CD, Version Control

Experience

UniConnect: AI-Powered University Resource Platform December 2024 - Present

Founder and Developer

- Designed and deployed an AWS-based pipeline to store university website data in S3, automating monthly web scraping with Scrapy, AWS Lambda, and CloudWatch, while processing data using AWS Glue for structured storage.
- Currently implementing a Retrieval-Augmented Generation (RAG) and RLHF model on Amazon SageMaker, leveraging SentenceTransformers, FAISS, and Hugging Face's 'trlX' to enhance retrieval accuracy based on user feedback.

Accenture Pvt. Ltd February 2021 – July 2023

Packaged App Development Analyst

- Designed and implemented a hybrid financial forecasting system by combining LSTM for long-term market trends, XGBoost for capturing key financial indicators like P/E ratios, moving averages, and volatility metrics, and SARIMA for handling seasonal fluctuations in NYSE-listed stocks, leading to more accurate and explainable predictions.
- Built a sentiment-driven prediction pipeline by fine-tuning BERT and RoBERTa on financial data, extracting insights from earnings calls, market news, and analyst reports, and integrating these sentiment scores into forecasting models, improving predictive accuracy and risk assessment for institutional clients.
- Developed custom Chart.js plugins to extend core visualization capabilities, including interactive drill-downs and cross-chart filtering, reducing individual dashboard implementation time from 2 weeks to 4 days, while being adopted by 3 different project teams serving Fortune 500 clients.

Projects and Academic Experience

Pylinguist | *Python, Cross-Lingual NLP, under review at ACL BEA workshop* November 2024

- Built a multi-stage code translation pipeline processing 550K+ Python samples by creating a custom dataset for keyword mapping, and using LLMs (GPT-4, Claude, DeepseekAI), achieving 85% BLEU score and 98% sematic similarity across 7 widely-spoken languages through bidirectional translation validation and back-translation testing.

Pattern-Exploiting Training: Reproducibility Study | *PyTorch, RoBERTa, Few-shot Learning* October 2024

- Extended PET framework with dynamic pattern contextualization and adversarial perturbation testing, analyzing model robustness across linguistic variations and achieving consistent 88%+ accuracy on AG News across different prompting patterns with only 100 training examples.

Hybrid Recommender System with PySpark | *PySpark, ALS(Alternating Least Squares), Random Forest* October 2024

- Engineered a three-component recommendation system using PySpark for 25M ratings across 62K movies, integrating ALS matrix factorization, item-item collaborative filtering, and Random Forest regression, achieving RMSE of 0.8069.
- Implemented performance optimizations with 100-partition data distribution, strategic caching, and broadcast variables, creating an efficient weighted ensemble that balanced prediction accuracy and recommendation ranking quality.

Predictive Analytics for Job Fraud Detection | *PySpark, ML Classification, NLP* September 2024

- Implemented a fraud detection pipeline using PySpark ML with TF-IDF vectorization (n-gram=1-3) and optimized MLP Classifier (84% accuracy, 0.81 F1-score) on 18K job postings, enabling automatic flagging of suspicious listings.
- Engineered a dual-vector NLP system processing job titles (15-dim) and descriptions (5000-dim) with NLTK, applying PCA and SHAP analysis to identify significant fraud indicators and create a risk-scoring framework for content review.

Graduate Research Assistant, Teaching Assistant, Developer | *George Mason University* July 2024 - Present

- Developed a novel alignment-based distance function using multiplicative similarity to identify high-probability topic overlaps in LDA models. Applied it to analyze global and regional policy alignments in national AI strategies over 50+ documents.
- Hybridized LDA with fine-tuned ChatGPT 3.5 to improve topic interpretation and summarization. Designed optimized prompts for generating concise, context-aware summaries from top topic words.
- Designed and deployed an end-to-end RPA solution for GMU Fiscal Services that automated deposit transactions, saving 504 hours annually while handling the full development lifecycle from requirements gathering to UAT implementation.

Certifications

AWS Certified Developer Associate (DVA-C02) | AWS Certified Machine Learning Engineer - Associate (MLA-C01)

Education

George Mason University August 2023 - May 2025

Masters in Computer Science, Machine Learning Concentration, ***GPA:*** *3.93/4.0* *Fairfax, VA*

Courses: Mining Massive Datasets with MapReduce, Advanced NLP, Machine Learning, Data Mining, Artificial Intelligence

Extra-Curriculars

Google Developer Group on Campus at George Mason University August 2024 - Present

Co-Lead

- Co-founded GMU's first GDG chapter, growing membership to 100+ students through strategic outreach and campus partnerships.
- Organized bi-weekly ML/AI workshops covering RAG systems, prompt engineering, and ML frameworks, developing leadership skills while creating a collaborative tech community.