Kartik Gupta

+1 413 468 0353 kartikgt23@gmail.com kartikgt.github.io kartikgupta23

Education

- **University of Massachusetts Amherst** *2022 to 2024*
  MS in Computer Science
- **Indian Institute of Technology Kharagpur** *2013 to 2018*
  BS + MS in Electrical and Computer Engineering

Experience

- **Mathematica** *WA, USA*
  Data Scientist Intern *June 2023 to Aug. 2023*
  - Full stack data analytics and visualization tool for community health outcome prediction using R Shiny and AWS
  - Data scraping, wrangling, and transformation into pandas data frames for Sparse PCA and feature store curation
  - Developed ML models (Random Forest, Lasso Regression) for prediction of health scores and county similarities
  - Automated database creation in Redshift and RDS (SQL, Python, and Boto) orchestrated with AWS Step Functions
- **Adobe** *MA, USA*
  Student Researcher *Jan. 2023 to May 2023*
  - Developed neural image compression algorithms for optimal latency and storage in computer vision ML pipelines
  - Benchmarked texture recognition task performance for the baseline model and our custom neural architecture
  - Demonstrated improvements in top-5 accuracy for 1 bit (43% to 90%) and 8 bits (89% to 92%) per pixel quantization
  - Exhibited the generalization capabilities on satellite image classification on RSCNN7 and RESISC45 datasets
- **Uber** *May 2021 to Aug. 2022*
  Software Engineer 2
  - Developed and deployed a Spark and Presto query optimization tool with automated JIRA tracking, saving $1.5M
  - Performance tuned and optimized 50+ Apache Spark pipelines correcting for OOM, skew, and small file issues
  - Developed 3 real time event streaming pipelines using Apache Flink and Kafka (Java, Scala) for live reporting
  - Modeled central 10PB+ table employing SQL structs, exploding views, self joins, CTEs, and window functions
  - Managed and curated feature stores for deployed ML models, in collaboration with a team of applied scientists
  - Architected 15+ Hive based batch analytics pipelines with incremental updates for slowly changing dimensions
- **Envestnet Yodlee** *Oct. 2020 to Apr. 2021*
  Member Technical Staff
  - Lead the design and development of a Neo4j, a NoSQL graph database, based Master Data Management System
  - End-to-end application development on AWS stack using EC2 hosting, S3 storage, and Lambda event triggers
- **SAP Labs** *June 2018 to Oct. 2020*
  Software Engineer
  - Developed a Java and selenium based automation test suite for periodic application monitoring and logging
  - Modeled ARIMA based time series prediction used for planning resource assignment for application traffic
  - Visualized multi-variate distributions for health metrics with dashboards in Grafana, Splunk, and Kibana
  - Built an organization wide data lake on Hadoop HDFS as an SSOT for application and network data storage
  - Architected and deployed Spark and Hive based data pipelines for ETL of 10TB+ daily data volume into HDFS
  - Engineered a Java application for interactive API based querying of the data warehouse in Elastic Search
  - Wrote Ansible and Terraform scripts for migration of on-prem services to Google Cloud Platform (GCP)

Projects

- **Simulating interactions of LLM agents and evaluating in-context learning** *Aug. 2023 to present*
  - Utilized Langchain to create agents from LLama2 7b and simulated inter-agent debate and collaboration scenarios
  - Generated synthetic data from debate responses and analyzed conversational drift, topic modeling, and perplexity
  - Quantified incontext learning and influence via measuring delta across iterations with aspect based sentiment analysis
  - Inspired by Socratic AI, introduced a third instruction model based agent to act as an adjudicator of the debate
  - Few-shot prompting on common debate topics, arguments, and summarization from web and Reddit to the adjudicator
  - Experimented with prompting techniques (Chain-of-Thought, Tree-of-Thought) and quantized models (GPTQ, AWQ)
  - Used Pytorch, HuggingFace, Langchain, Transformers, Accelerate
- **Transfer Learning and Few Shot improvements for medical deep learning** *Sept. 2022 to Dec. 2022*
  - Utilized pretrained RESNET18 and VGG16 as backbones and partially finetuned on Covid-19 classification dataset
  - Simulated real-world medical cases by introducing class imbalances leading to increase in AuC (0.64 to 0.86)
  - Increased size of training corpus via data augentation using image transformations (affine, gaussian, perspective)
  - Implemented 4-way 5-shot learning schemes on trained model to further enhance accuracy on novel test datasets
  - Used Pytorch, Pytorch Lightning, Torchvision, Pytorch Hub
- **Deploying Sentiment Analysis ML model on Microsoft Azure** *May 2020 to June 2020*
  - Developed and registered a sentiment analysis model for classifying IMDB reviews on Azure Machine Learning Studio
  - Implemented endpoints (online and batch) for inferencing via API and deployent and autoscaling with Kubernetes
- **Translating images to natural language sentences for captioning and comment** *Mar. 2017 to Apr. 2018*
  - Researched image to text translation using self attention embedded LSTM modules on CNN feature embeddings
  - Improved performance on BLEU-3 and BLEU-4 compared to baseline BRNN model (30.4 to 34.4, and 20.3 to 24.3)

Skills

- **Languages**: Java, Python, Scala, Golang, R, SQL, C++, Shell scripting
- **Tools and Platforms**: Linux SUSE, Kubernetes, Docker, GitHub, Jenkins, SQL Workbench, DBeaver, JIRA, ServiceNow
- **Cloud Platforms**: Amazon Web Services (AWS), Microsoft Azure, SAP Cloud Platform
- **Data**: Apache Spark, Apache Flink, Kafka, Hive, Presto, AWS EMR, Pandas, Tidyverse, Numpy, Sklearn
- **Databases**: MySQL, Redshift, RDS, Elastic Search, HDFS, Amazon S3, Apache Pinot, Amazon Athena
- **Deep Learning Frameworks**: Pytorch, Pytorch Lightning, Keras, Tensorflow 2.0, Langchain, LlamaIndex

Certificates

- **Microsoft Certified Azure Data Scientist Associate**: Credential ID - H446-0997
- **Neo4j Certified Professional**: Credential ID - 17127043

Honors

- Selected as a beta tester for Manning publication of 'Real-time Stream Processing with Kafka and Spark"
- Graduate Teaching Assistant for Advanced Machine Learning (Fall 2023) and Computer Vision (Spring 2024)
- Taught Data Science and Data Engineering to experienced professionals as an instructor at Scaler Academy