KARTIK GUPTA

kgupta@umass.edu | kartikgupta23 | kartikgt | +1 (413) 468-0353

Education

University of Massachusetts, Amherst *Sep 2022 – Dec 2023*
Computer Science | 3.95/4.0 M.S.

Indian Institute of Technology (IIT) Kharagpur *2013 – 2018*
Electronics & Comm. | Honors Distinction B.S. + M.S.

Experience

Mathematica Policy Research *Chicago, IL*
Data Science Intern *Jun 2023 – Aug 2023*

- Designed an application which analyzed multi-domain features to estimate Social Determinants of Health for all 3200 US counties and facilitated knowledge transfer between highly correlated counties with differing health outcomes. Employed Sparse PCA, Lasso regression, and Random forests.
- Piloted a Langchain application to query an LLM in contextualizing digital medical records to predict readmissions. Evaluated various prompting techniques, which led to a 34% increase in accuracy.
- Created datasets capturing prices offered by health insurance providers for a range of medical procedures including medicaid reimbursements at varying granularity levels. Orchestrated AWS Lambda computations on RDS and Sagemaker using Step Functions written in Python botocore.

Adobe Research *Amherst, MA*
Graduate Researcher *Feb 2023 – May 2023*

- Researched enabling computer vision tasks of recognition and segmentation on the compressed representations of image via learning based compression codecs similar to the efforts led by JPEG AI team.
- Implemented an auto-encoder based compression model using Pytorch Lightning to get latent image representations. Lightning framework allowed modularizing code with appropriate abstraction levels.
- Compared performances of multiple compressed models differing in specified bitrate and quantization parameters. Evaluation criteria included latency, task accuracy, PSNR, and memory utilization. Measured and stored the model results in a Weights and Biases project to efficiently publish outcomes.

Uber *Bangalore, India*
Machine Learning Engineer II *May 2021 – Aug 2022*

- Developed a stream processing application offering real-time analytics to merchants. Wrote stateful transformations on Kafka streams using Apache Flink, leading to 13% increase in customer retention.
- Deployed prediction models on production while monitoring data distribution shifts, data freshness, and model performance alongside triggers for automated retraining, in line with MLOps best practices.
- Expanded canonical coverage of organization data resources from 19% to 36% by revamping the existing datasets and datamodels into a more efficient and standardized data model.
- Achieved savings of $1 million by developing a resource monitoring tool. Using pattern matching, it automatically detected resource-intensive queries by identifying inefficient implementations.
- Facilitated the integration of Cornershop and Drizly data into the Uber ecosystem. Designed a unified data model and wrote comprehensive PRDs to plan efficient transfer of data solving data silos problem.

Envestnet Yodlee *Bangalore, India*
Senior Data Scientist *Oct 2020 – Apr 2021*

- Implemented a Master Data Management (MDM) system built on Neo4j, a NoSQL graph database, to establish a centralized source of truth for financial metadata. Automated with AWS Data Pipelines.
- Created data pipelines triggered by the ingestion of materialized views in S3, coordinated via AWS Lambda to transform and load the data into the Neo4j database. Designed the graph schema using Cypher scripts.

SAP Labs *Bangalore, India*
Data Scientist *Jun 2018 – Sep 2020*

- Developed and configured dashboards for metrics visualizations on Splunk, Kibana, and Grafana.
- Reduced SLA-unmet costs by $1.4 million and saved 30 man-hours per week by implementing a Hadoop data lake and ElasticSearch warehouse solution that offered a unified view of the entire data ecosystem in accordance with CI CD practices using Jenkins, enabling teams to efficiently triage critical issues.
- Provisioned infrastructure and deployed applications on a new data center on Google Cloud Platform (GCP) for migrating legacy shared services to cloud using Terraform, Ansible, GIT, and Jenkins.

Academic Projects

Dealing with Scarce Data Domains

- Deep Learning applications in health are constrained by lack of trainable data, data scarcity problem.
- Transfer Learning via pretrained Resnet50, and VGG16 finetunned for Covid19 Classification task.
- Augmented the data by creating a larger training corpus via combining multiple transformations.
- Implemented a Few Shot model and performed comparative analysis with Transfer Learned models.

Captioning Images to Natural Languages sentences using Deep Neural Networks

- Identified objects in images generated as feature vectors after processing via a pre-trained CNN.
- Used a self attention model to identify features at each time step from the outputs generated.
- The output from the attention model was fed to the LSTM along with the previous hidden state.
- Generated step-by-step captions for each of the identified features using an LSTM based RNN model.

Computer Vision Projects

- Implemented photometric stereo to get surface depth and true reflectance from shaded images.
- Developed demosaicing methods on Bayer filtered inputs to generate colored images.
- Learned blur kernel and the sigma-map for non uniformly blurred images to deblur a blurred image.
- Autoregressive model based completion of corrupted and incomplete 3D point clouds and meshes.
- Comparative study of image segmentation methods (UNet and Mask-RCNN) for medical images.

Skills and Certifications

Microsoft Certified Azure Data Scientist Associate: ***Credential ID H446-0997***
Neo4j Certified Professional: ***Credential ID 17127043***
Statistical Analysis: Bayesian Modeling, Multivariate Analysis, Hypothesis testing, Statistical modeling
Languages and Packages: Python, Pandas, Botocore, Selenium, Scala, Java, Shell, R, R Shiny, SQL, Luigi
Machine Learning: PyTorch, Lightning, Langchain, LLMs, Torch Profilers, Hugging Face, OpenCV
Data Processing: Apache Flink, Spark, Hadoop, Neo4j, Elasticsearch, Kafka, Hudi, RDS, Redshift, MYSQL
Business Domains: US Healthcare, Delivery Platforms, Financial Markets, Human Capital Management

Extracurriculars

Grader for Advanced Machine Learning course (COMPSCI 689) at UMass Amherst *2023*

- Assisted Professor with crafting assignments and evaluating copies of 100+ graduate students.
- Resolved students' questions during office hours and reiterated concepts taught in the lecture.

Data Science mentor and faculty at Scaler *2022*

- Taught data science and engineering concepts, as used in the industry, to seasoned professionals.
- Designed the curriculum and researched relevant upto-date concepts in Machine Learning and NLP.

Indian military training *2013 - 2014*

- Cadet at National Cadet Corps Bengal EME Coy, trained under Air Force base Salua, West Bengal.
- Experienced in leading a platoon of 50+ cadets during marches and instilling military rigour.