Akash Mishra

New York | akash.mishra13@gmail.com | +19294064445 | http://akashmishra.me

Computer Science Engineer with experience in Software Engineering and Machine Learning

GitHub: - https://github.com/AkashSky1994 LinkedIn: - https://www.linkedin.com/in/akash-mishra1994

Education

New York University, New York — *M.S. (Computer Engineering)* **September 2021 - May 2023**

Data Science | Machine Learning | Deep Learning | Computer Vision | High-Performance Deep Learning | ML in Cybersecurity | Internet Protocol

APJ Abdul Kalam University, India — *B.Tech. (Computer Science and Engineering)* **April 2013 - May 2017**

Skills

- ***Language*** **-** Nodejs, Golang, Python, JavaScript **|** ***Database*** **–** NoSQL (MongoDB), SQL (MySQL, PostgreSQL)
- **Other Tools -** PyTorch, PyTorch-Geometric, Elasticsearch, AWS (EC2, Kinesis, SQS, S3, Cloudwatch), GCP, Redis, Git, Docker, Snowflake, Hadoop, Airflow, Spark, scikit-learn, NumPy, Pandas, Huggingface, Dask, dbt, Terraform, Langchain, CI/CD, CircleCI

Experience

New York University, NY — ***Research Assistant*** **January 2023 - April 2023**

- Conceptualized and implemented a self-supervised Graph neural network which gave comparable ROC-AUC score of 0.65 to the SOTA model by leveraging ResNet50 and DistilBERT for image & text encoder along with masked autoencoders using PyG.
- Reduced the training time by 40% by deploying on Multi-GPU High Performance Server (Nvidia RTX8000 and Tesla V100 GPUs)

Thrasio, NY — ***Data Engineer Intern*** **June 2022 - August 2022**

- Implemented ETL pipeline using Amazon Ads API by leveraging Python, Celery for job scheduling, Redis and AWS for cloud infra.
- Prototyped data pipeline that handles around 20 billion rows, using tools like Snowflake for data warehousing, S3 for object storage, SQS for queue building and dbt for data transformation.
- Proposed an automated service that would reduce the developer resource allocated up to 40% by identifying multiple data sources, processing these raw data, and building data pipelines using Snowflake, SQS and Datadog Monitoring.

LBB (Little Black Book), India — *Software Engineer* **September 2019 - July 2021**

- Developed in-house search engine with Elasticsearch as an underlying technology that improved organic user engagement by 60% leveraging Golang, Python, AWS infrastructure, and NLP techniques like Topic modeling, Lemmatization, and Keyword extraction.
- Spearheaded a DEV team to build a containerized microservice to optimize pickup and delivery routes leading to reduction in logistics cost by 25% and average delivery time by 1 day using NodeJS, Redis, CI/CD tools and AWS Cloud for deployment.
- Collaborated with finance team to design a highly scalable distributed service (using Golang, AWS Kinesis, Autoscaling EC2, Redis) which tracks microtransactions within ecommerce ecosystem, automates payout and optimizes refunds, reducing the cost by 30%.

Bookchor, India — *Software Engineer* **July 2017 - August 2019**

- Created a recommendation system that employs Matrix factorization primarily SVD to provide personalized product suggestions and K-Nearest Neighbors to get similar product suggestions using prior reading preferences for users.
- Identified key problems and bottlenecks within existing architecture, took the initiative to improve the performance and scalability of services by implementing a distributed cache and a distributed sharded NoSQL database improving API response time by 100ms.

Projects

Depth Estimation by Self-Supervised Learning for Monocular Images [Link]

- Researched and trained self-supervised learning model for depth estimation on KITTI dataset by experimenting with various iterations of UNet Architecture involving Encoders (ResNet, ConvNext) and Up-sampling decoders like ESPCN achieving loss of 3.7.
- Enhanced our model which reduced the artifact problem in the disparity map and consequently reduced the overall RMS loss to 3.485 by superimposing a semantic map using Mask R-CNN pre-trained on COCO Dataset (built using PyTorch).

Hateful Meme Detection using Multimodal neural network [Link]

- Engineered various iterations of multimodal neural network to identify hateful intent from the large corpus of memes.
- Generated metrics (ROC-AUC: 0.59) comparable to models published during NeurIPS 2020 Challenge (by Facebook AI).

Collaborative-filter based Recommender System Using Apache Spark [Link]

- Modeled a large-scale recommender system using ALS factorization algorithm for implicit feedbacks and Apache Spark.
- Improved the performance by partitioning the data using parquet files and established a distributed parallelizable ALS module.

Stealthy Syntactical Backdoor Attack on BERT [Link]

- Innovated a syntactical backdoor attack on Language models using T5 Model along with SCPN to generate poisoned paraphrased sentences which is not perceivable by the standard audit process.
- Increased Attack success rate to 96.73% for stealthy syntactical attacks keeping the change in original accuracy negligible.

Job Posting Retrieval Webservice With Golang and Terraform [Link]

- Architected a Job posting Scrapping service in Golang which fetches jobs from multiple sources into a common database.
- Programmed a Github CI/CD pipeline for automated deployments and used Terraform for Infrastructure orchestration onto Cloud.