Vinayaka M Hegde

vinayakah@tamu.edu | (979) 721-2288 | linkedin.com/in/vinayaka2000 | https://github.com/Vinayaka2k

EDUCATION

Texas A&M University — College Station, TX
Masters in Computer Science (MCS) — *May 2026*
Relevant Courses: Software Engineering, Algorithms, Artificial Intelligence

PES University — Bangalore, India
Bachelor of Technology in Computer Science and Engineering — *May 2021*
GPA: 9.01/10.00. Relevant Courses: Operating Systems, Big Data, Cloud Computing, Web Technologies
Awards: Prof. CNR Rao Scholarship (top 10%)

SKILLS

Programming Languages: Python, Javascript, Java, Ruby, C++, HTML, CSS, SQL
Frameworks & Libraries: Flask, React.js, PyTorch, NLTK, pandas, numpy, scikit-learn
Cloud & Tools: Apache Hadoop, Git, Docker, AWS, Postman, Heroku, Github Actions

EXPERIENCE

PES Labs — Bangalore, India
Software Engineer - ML Team | Project: *Developing and deploying ML workloads on Nvidia Jetsons* — *Jan 2023 – Jul 2024*

- Achieved ***up to 3x*** reduction in inference latency by leveraging the **TensorRT** framework and offloading selective model layers to the Deep Learning Accelerator for optimized execution
- Reduced ***model complexity by 12%*** by leveraging PyTorch's qnnpack engine to lower the precision of weights via **INT8 quantization**
- Developed a tool using PyTorch's runtime dispatcher to compute the Floating Point Operations (FLOPs) of models, resulting in a ***47% reduction*** in the FLOPs after applying the above optimization methods
- Implemented object detection models such as Mobilenetv3, Resnet50, YOLOv8, and containerized them using **Docker**
- Identified excessive system calls as the primary source of overhead in containerization, leading to significant insights that culminated in multiple research papers

Leadsquared — Bangalore, India
Software Development Engineer 1 | *Integrations Team* — *Jul 2021 – Dec 2022*

- Migrated backend ***RESTful APIs*** from on-premises to **AWS EC2**, reducing infrastructure costs by ***Rs. 92K***, and ***improving performance by about ~ 30%***
- Optimized long-running and bulk-update tasks by transitioning to **Python**-based batch jobs, ***reducing execution time by 23%***
- Orchestrated ***Adobe E-Sign API integration*** for Mercer, creating a customized workflow that automated document signing

Hewlett Packard Enterprise — Bangalore, India
Software Developer Intern | Project: *Abstractive text summarizer* — *Jan 2020 – Jul 2020*

- Implemented a ***scalable,*** fault-tolerant Database-as-a-Service (DBaaS) for the backend of a cloud-based application
- Zookeeper ensured ***high availability and fault tolerance,*** and RabbitMQ streamlined communication across nodes
- Utilized nginx as reverse-proxy, and configured an application load-balancer, achieving ***15% increase in throughput***

SELECTED PROJECTS

SQL Engine for vast datasets | Big Data Analytics — *(MapReduce, Hadoop Distributed FIle System, Apache)*

- Engineered a map-reduce based SQL engine, similar to Hive, capable of query parsing, selection, and aggregation
- The mapper filters rows based on the WHERE clause, and reducer aggregates the results using an aggregate function

Enhanced Neural Machine Translation using Attention | *Deep Learning* — *(Autoencoders, LSTM, RNNs)*

- Optimized a translation model, achieving ***97% accuracy*** with attention, improving upon the baseline of 94%
- Employed LSTM-based sequence-to-sequence architecture with dropout regularization

MyBlogApp | *Deep Learning*

- Developed a Node.js/Express backend for a blogging app, integrating JWT-based authentication middleware, and managed data models with MongoDB, achieving 99% uptime on Render via automated Github Actions
- Architected the frontend using React components, and designed custom hooks that resulted in 28% reduction in page load times

SELECTED PUBLICATIONS

[1] Towards Efficient Scheduling of Concurrent DNN Training and Inferencing on Accelerated Edges, ***CCGrid 2023***

[2] Performance Characterization of Containerized DNN Training and Inference on Edge Accelerators, ***HiPC 2023***