RAJU GUGULOTH | CS24M036

Indian Institute of Technology Madras

in LinkedIn   GitHub

Professional Experience

RINEX

Artificial Intelligence Intern   *Jul 2022 – Sep 2022*

1. Sentiment Analysis on Social Media Data

○ Built supervised models (Logistic Regression, Naive Bayes, SVM, Random Forest) for sentiment classification.

○ Preprocessed 150K+ Twitter/Reddit posts; did cleaning, tokenization, TF-IDF features.

○ Achieved **88.2%** classification accuracy with Random Forest; performed hyperparameter tuning and metric evaluation.

2. Rain Prediction Using Weather Data

○ Developed model to forecast "RainTomorrow" using a structured weather dataset (∼140K samples, 23 features).

○ Executed pipeline: ingestion, imputation, encoding, scaling, feature selection.

○ Random Forest achieved **88.7%** accuracy and **AUC-ROC 0.91**.

Tools: PyCharm, Google Colab, Kaggle.
Skills: Python, Pandas, NumPy, Scikit-learn, model evaluation, feature engineering.

Projects

RAG-Enhanced Information Retrieval System with LSA & LLM Integration *Independent Project 2025*

○ Dataset: Cranfield dataset (1,400 abstracts, 225 queries, 1,837 relevance judgments).

○ Implemented FAISS + LSA hybrid retrieval with GenAI (Perplexity API, HuggingFace LLMs).

○ Developed a Gradio-based chat interface with rapid contextual response and document citation.

○ **Tools:** Python, FAISS, Scikit-learn, HuggingFace, Perplexity API, Gradio.

Feedforward NN and Optimizers — Fashion-MNIST   **Feb 2025 - Mar 2025**

Instructor: Prof. Mitesh Khapra   *DA6401: Deep Learning*

○ Built a customizable feedforward neural network from scratch for Fashion-MNIST, achieving **84.14% test accuracy**.

○ Implemented six optimizers — **SGD, Momentum, Nesterov, RMSProp, Adam, Nadam** — and ran W&B sweeps, boosting accuracy by ∼3%.

○ Discovered an optimal configuration achieving ∼**98.4% test accuracy on MNIST** through loss function and optimizer performance analysis.

CNN and Transfer Learning — iNaturalist Classification   **Apr 2025**

Instructor: Prof. Mitesh Khapra   *DA6401: Deep Learning*

○ Designed and trained a flexible CNN architecture on the iNaturalist dataset, achieving **45.55% test accuracy** when trained from scratch through W&B-driven hyperparameter sweeps.

○ Implemented and optimized multiple configurations including filters, activation functions, dropout, and batch normalization using **Adam** and **Nadam** optimizers, leveraging Bayesian optimization for efficient search..

○ Fine-tuned a pretrained **ResNet50** on iNaturalist, boosting validation accuracy from **45.52% (scratch)** to **76.33%**, demonstrating the effectiveness of transfer learning.

Seq2Seq Transliteration — Dakshina Dataset   **May 2025**

Instructor: Prof. Mitesh Khapra   *DA6401: Deep Learning*

○ Designed and implemented an RNN-based Seq2Seq transliteration system on the Dakshina dataset, exploring Vanilla RNN, GRU, and LSTM architectures for character-level language mapping.

○ Executed Bayesian hyperparameter sweeps (93 trials) to fine-tune embedding size, hidden layers, dropout, learning rate, and training epochs, achieving **55.01% validation** and **54.74% test accuracy**.

○ Performed in-depth benchmarking and error analysis across architectures, uncovering GRU's superior generalization and optimal dropout settings for robust, overfitting-resistant models.

Semantic-Aware Information Retrieval System (Cranfield Dataset)

Instructor: Prof. Sutanu Chakraborti   *Natural Language Processing CS6370*

- **Designed** an advanced IR system by integrating **Latent Semantic Analysis (LSA)**, **K-Means clustering**, and **FastText query expansion** to overcome VSM limitations (high dimensionality, lack of semantics, vocabulary mismatch)and significantly improve retrieval accuracy, efficiency, scalability, and semantic relevance across documents.
- **Improved retrieval efficiency** by clustering semantically transformed document vectors, enabling faster searches without compromising accuracy, scalability, or overall system performance.
- **Enhanced recall and ranking quality** via query expansion with FastText embeddings, effectively handling synonymy and polysemy thereby improving query relevance, retrieval diversity, user satisfaction, and semantic coverage.
- **Evaluated** on the **Cranfield dataset**, achieving consistent gains in **Precision, Recall, MAP, F-score, and nDCG** over the baseline VSM, demonstrating robust performance, scalability and strong generalization across retrieval tasks.

Performance Evaluation of Container Orchestration

Instructor: Prof. Janakiram D   *Cloud Computing CS6847*

- Engineered large-scale load tests ($10 \rightarrow$ **10,000 RPS**) on a containerized Flask microservice deployed on **Microsoft Azure, ensuring scalable, resilient, high-performance cloud operations.**.
- Identified **critical degradation at 1,000+ RPS** in single-instance deployments, validating system bottlenecks via queueing theory.
- Implemented scaling (Docker Swarm: **5 replicas**, Kubernetes HPA: **1–10 pods**) achieving **13× latency reduction** and **2,857% performance improvement**.
- Demonstrated Kubernetes' **dynamic auto-scaling superiority**, providing actionable insights for production-grade scalability optimization.

Education

Program	Institution	Year
M.Tech. (Computer Science and Engineering)	Indian Institute of Technology, Madras	2026
B.Tech. (Computer Science and Engineering)	Kakatiya University, Warangal	2024
Intermediate (Class XII)	RD Junior College, Nayeem Nagar	2019
Secondary (Class X)	TSWREIS (Boys), Wardhanapet	2017

Course Work

- **Programming:** Advanced Data Structures and Algorithms, Advanced Programming Lab.
- **Machine Learning:** Natural Language Processing, Pattern Recognition and Machine Learning, Deep Learning, Linear Algebra & Random Process.
- **Systems:** Operating Systems, Cloud Computing, Applied Cryptography, Digital Systems Testing and Testable Design.
- **Networks:** Computer Networks.
- **Databases:** Database Management System, SQL.

Technical Skills

- **Programming Languages:** C, C++, Python, Java.
- **Tools:** Visual Studio Code, Jupyter, Git.
- **Databases:** MySQL ,PostgreSQL,MongoDB.
- **Operating Systems:** Windows, Linux.
- **Frameworks:** TensorFlow, PyTorch, AWS, Pandas, NumPy, Scikit-learn.

Positions of Responsibility

CAD for VLSI – Prof. Gopalakrishnan Srinivasan

Teaching Assistant   *Jul 2024 – Nov 2024*

- Guided students in CAD algorithms for VLSI, assisted lab sessions, and evaluated project submissions.

Design and Analysis of Algorithms – Prof. Jayalal Sarma

Teaching Assistant   *Jan 2025 – May 2025*

- Supported students in problem-solving sessions and contributed to assignment preparation and grading.