Varsha Hemakumar

+1 716-533-7266 | shavarsha02@gmail.com| linkedin.com/in/varsha-hemakumar| github.com/VarshaHemakumar

Education

State University of NewYork at Buffalo Buffalo, NY
Master of Science in Computer Science - CGPA 3.8/4 Aug. 2024 – Present
Relevant Coursework - Data Intensive computing, Intro to Machine Learning, Algorithms and Analysis, Computer Security, Operating Sys, Deep Learning, Data Models and Query Language, Computer Vision

Sri Sivasubramaniya Nadar College of Engineering Chennai, India
Bachelor of Technology in Information Technology - CGPA 9.2/10 Aug. 2020 – May 2024
Relevant Coursework - Advanced Data Structures, Database Management Systems, Artificial Intelligence, Network Security

Indian Institute of Technology Madras - Diploma in DataScience Chennai, India
Relevant Coursework - Machine learning Foundation, Statistics I and II, Mathematics I and II, MLP

Technical Experience and Projects

Springer Capital May 2025 – July 2025
AI and Data Science Intern Remote, USA

· Performed in-depth EDA on 50+ real-world email and chat datasets to assess structure and sentiment labeling.
· Created a consolidated report evaluating dataset usefulness for multi-class sentiment classification based on format, label quality, and conversation threading.
· Defined classification schemas and led dataset selection for downstream model development, aligning with project standards and documentation workflows.

Ohm Clouds Mar. 2024 – May 2024
Machine Learning Intern Chennai, India

· Conducted a 2-month research internship at Ohm Clouds, analyzing and evaluating recommendation algorithms (e.g., collaborative filtering, content-based) to enhance personalized systems. Compiled a detailed report on algorithm performance, identifying optimal use cases for implementation.

Vayusastra Indian Institute of Technology Madras Apr. 2023 – Jul 2023
Internet of Things Intern Chennai, India

· Completed a three-month training and internship focused on gaining comprehensive insights into the Internet of Things (IoT), during which I executed a Smart Bin project using IoT technologies like AdaFruit, MQTT, IFTTT, and Blynk, successfully implementing waste segregation through voice commands.

Brief Me the Case: AI- Powered Legal and News Summarization System Mar 2025 – May 2025

· Built a full stack summarization system combining extractive, abstractive and rewriter models for high accuracy summarization of legal and news documents.
· Trained a LoRA optimized DistilBERT regression model for extractive sentence scoring with ROUGE supervision, achieving ROUGE-1: 0.25, ROUGE-L: 0.16.
· Fine-tuned facebook/bart-large-cnn for abstractive summarization with a custom PyTorch training loop, achieving BERTScore: 0.87, ROUGE-1:0.45 and GPT-2 Perplexity:19.19.
· Developed a T5 and Bart based rewriter pipeline to humanize summaries, improving fluency and readability.
· Deployed the full pipeline using FastAPI, enabling real-time summarization and humanization via a professional web interface.

CourseBase: AI-Powered Course Management and Recommendation System Jan 2025 – Apr 2025

· Developed a full-stack web platform using Flask and PostgreSQL for dynamic course enrollment, attendance tracking, and AI-based course recommendations.
· Designed a relational schema supporting 10,000+ students, 50+ courses, and 15,000+ attendance records; enforced BCNF through decomposition and implemented complex SQL constraints.
· Built and deployed a hybrid recommender system combining collaborative filtering (cosine similarity) and content-based filtering (course difficulty, credits) to suggest optimal course paths.
· Integrated a Tableau dashboard for visual analytics showing top-rated courses, attendance trends, and recommendation sources.

Comparative Analysis of DenseNet- 161 and CDCN++ For Face Anti-Spoofing Nov 2023 – Feb 2024

· Achieved a 12% improvement in spoof detection accuracy on a dataset of 50,000+ images by optimizing DenseNet-161 and CDCN++ architectures in Python.
· Reduced false positives by 18% using advanced depth map generation and co-occurrence feature extraction techniques.
· Applied CNN-based facial recognition and model evaluation techniques to enhance biometric security in computer vision systems.

Skills and Certificates and Publications

Languages: JavaScript, Python, SQL, C++, C, Java

Skills: Machine Learning (PyTorch, TensorFlow), Deep Learning, Docker, HTML, CSS, JavaScript, AngularJS, Angular, NodeJS, Django, Power BI, Tableau, MATLAB, SQL, Scikit-learn, XGBoost, NumPy, Pandas, Hadoop, Neural Networks, CNN, RNN.

Certificates: Foundation in Data Science and Programming (IIT Madras), Programming in Java, Introduction to Programming through C++, Design & Implementation of Human-Computer Interfaces, Programming in Modern C++

Publications: Presented the paper **Stress Detection in Videos Using Machine Learning** at ICIPCN 2024, Kathmandu University, published by IEEE Conference.Developed a non-intrusive stress detection system using CNN and ResNet on FER-2013, analyzing real-time video feeds for stress classification.