

NAVYA SIVA RAJ

Data Scientist | Python, SQL, Machine Learning, NLP, Predictive Analytics

navyasraj2003@gmail.com (832) 724-9256 [linkedin.com/in/navya-s-raj/](https://www.linkedin.com/in/navya-s-raj/) github.com/navyasraj02 navyasraj02.github.io

EDUCATION

Master of Science, Computer Science · GPA: 3.8/4.0
University of Houston, Houston, TX

Aug 2024 – May 2026

Bachelor of Technology, Computer Science and Engineering (Honors) · GPA: 9.57/10.0
Kerala Technological University (KTU), Rajiv Gandhi Institute of Technology (RIT), Kerala, India

Nov 2020 – Jun 2024

TECHNICAL SKILLS

- **Languages:** Python, SQL
- **Data Analysis:** Pandas, NumPy, Exploratory Data Analysis, Statistical Analysis, Feature Engineering
- **Machine Learning:** Scikit-learn, PyTorch, Predictive Modeling, Classification, Regression, Cross-Validation, Hyperparameter Tuning, NLP
- **Visualization:** Matplotlib, Seaborn, Tableau
- **Data & Deployment:** MySQL, MongoDB, Jupyter, AWS (EC2, S3), Docker, Git, FastAPI, Flask

WORK EXPERIENCE

Application Developer | Hewlett Packard Enterprise Data Science Institute at UH

Jun 2025 – Present

- Automated semester report generation using Node.js, TypeScript, Vue.js, and MongoDB, reducing generation time by 30–44 minutes per report.
- Strengthened application reliability by implementing biweekly backups and global error handling for recovery and observability.
- Led initial development of a consent-based public-sharing feature that converts student records into shareable dashboards and profiles.
- Deployed a production full-stack application behind NGINX on Jelastic/Reclaim Cloud, supporting 1,100+ users.

Research Assistant | University of Houston, Sugar Land

Oct 2024 – Jul 2025

- Engineered NLP preprocessing pipelines to clean, standardize, and consolidate 10,000+ social media and forum posts collected from multiple sources over six months.
- Applied sentiment analysis and topic modeling to surface emerging trends in user behavior, discussion themes, and engagement patterns, delivering actionable insights to research faculty.

Machine Learning Intern | Indian Institute of Information Technology, Kottayam (IIITK)

Dec 2021 – Apr 2022

- Collaborated with six interns to build ML workflows using Python, Pandas, NumPy, and Matplotlib, including data cleaning, visualization, preprocessing, model training, and evaluation.

ACADEMIC PROJECT EXPERIENCE

Multi-Class Multi-Label Traffic Crash Cause Identification | University of Houston

Jan 2025 – May 2025

- Constructed an end-to-end ML pipeline on 794K+ Chicago crash records, mapping 40 contributory causes to five target classes and reducing 48 raw features to 18 predictors through feature selection.
- Evaluated multiple classification algorithms, including Decision Tree, Logistic Regression, kNN, Random Forest, and SVM, using accuracy, precision, recall, and F1-score.
- Optimized kNN hyperparameters to achieve 54.19% accuracy, a 15-point gain over the 39% baseline, making it the top-performing model across all evaluated classifiers.

Traditional ML vs. Self-Hosted LLM Benchmarking on AWS EC2 | University of Houston

Jan 2026 – Feb 2026

- Benchmarked TF-IDF/Logistic Regression and quantized Llama 3.2 3B inference across six AWS EC2 configurations, comparing latency, throughput, and cost under varying batch sizes and concurrency levels.
- Traditional ML achieved 140–160 ms average latency and up to 65 texts/second; c7g.xlarge delivered the best self-hosted LLM cost-performance at \$2.622 per 1M tokens.

Pseudostem Weevil Detector | RIT, Kottayam

Mar 2023 – Jun 2023

- Curated 6,000 field-recorded acoustic samples and transformed audio into Mel spectrograms using FFT-based signal processing for CNN classification, achieving 98% held-out test accuracy.
- Operationalized the CNN model through a web application, enabling farmers to perform real-time audio-based pest detection.

Agentic LMS Automation System | University of Houston

Jan 2026 – Present

- Developed an instructor-facing agentic LMS assistant for Moodle and Canvas that analyzes assignment context, identifies pending tasks, and drafts rubric-aligned feedback for instructor approval.
- Orchestrated multi-agent workflows using LangGraph, MAF, and GPT-4o tool calling, enabling agents to translate page context into adaptive browser actions.
- Built an asynchronous Playwright automation layer for authenticated LMS navigation and page extraction; containerized Moodle with Docker and traced agent execution in LangSmith.