

NIVEDITHA SRIKANTH

Chennai, India | +91 9363043890 | ns225@uw.edu | Portfolio | LinkedIn | GitHub

TECHNICAL SKILLS

- ML & AI:** Python, R, scikit-learn, PyTorch, TensorFlow, XGBoost, Random Forest, BIRCH, LOF, GMM (Mclust), KMeans, PCA, t-SNE, NLP (TF-IDF, BERT, spaCy), CNNs, Sequence Models, Feature Engineering, Anomaly Detection, Model Evaluation
- Data & Cloud Infrastructure:** SQL (Athena, PostgreSQL, MySQL), AWS (S3, Glue, SageMaker, Lambda, Redshift), ETL Pipelines, Docker, Git, Power BI, Tableau, Jupyter, MS Excel
- Competencies:** Research Communication, Cross-team Collaboration, Technical Writing, Problem-Solving, Independent Research

EXPERIENCE

The Cigna Group

Bengaluru, KA, India

Data Science Associate Analyst

Jan 2026 – Apr 2026

- Built anomaly detection and risk scoring models on healthcare claims data; improved detection coverage by **20%**, contributing to **\$400K-\$700K** estimated savings
- Applied BIRCH + LOF on **10M+** records; identified **18%** high-risk providers and improved lift by **11%** and precision by **13%**

Bioinformatics Lab, REC (Supervised by Dr. Sujata Roy)

Chennai, TN, India

Undergraduate Research Assistant – Computational Bioinformatics (ML)

Jan 2025 – Apr 2025

- Developed unsupervised Machine learning pipeline on **49k feature** gene expression data (GSE57329) using PCA + GMM (G=9); achieved silhouette score **0.81**
- Reduced feature space by **98%** and identified immune-related gene clusters via PPI and pathway enrichment (KEGG, Reactome)

EDUCATION

Indian Institute of Technology Guwahati (Daksh Gurukul)

Guwahati, AS, India

Micro-Degree Program, Computer Science

June 2024 – Oct 2025

- Coursework:** Data Structures & Algorithms, Discrete Mathematics, Database Systems, Cloud Architecture

Rajalakshmi Engineering College, Anna University

Chennai, TN, India

Bachelor of Technology, Biotechnology

Oct 2021 – May 2025

- Coursework:** Python for Machine Learning, Probability & Statistics, Data Structures & Algorithms, Bioinformatics

PROJECTS

Mclust-Based Gene Expression Clustering for Immune Checkpoint Analysis | R, Mclust, PCA, t-SNE, UMAP, ggplot2, KEGG

- Clustered **41K+ mouse aortic endothelial genes** (GEO: GSE57329) via PCA + Gaussian Mixture Models to map immune check points in T2D-atherosclerosis; identified hub genes **Cd4, Cxcl10, Fmo3** via STRING/CytoHubba MCC ranking, validated with silhouette avg. **0.81**, t-SNE, and KEGG/Reactome enrichment

Cerebrovascular Accident Prognosis using ML | Python, scikit-learn, XGBoost, Random Forest, Decision Tree, AUC-ROC

- Developed and evaluated supervised ML models on clinical data (age, BP, cholesterol, history), achieving **99.9% accuracy** with Random Forest and **97.9%** with XGBoost for predicting cerebrovascular disease risk (WCONF 2023 — Best Presentation Award)

Healthcare Claims Fraud & Anomaly Detection | Python, Scikit-learn, PostgreSQL, SQL, pandas, psycopg2, Matplotlib, Power BI

- Developed an ETL workflow on **1M+ Medicare claims**, detecting **3%** anomalies using rule-based SQL flagging and provider outlier scoring (BIRCH, LOF), and prioritizing the top **1%** high-risk claims (**128K**) through an interactive Power BI dashboard

Scientific Paper Clustering | Python, scikit-learn, TF-IDF, PCA, t-SNE, UMAP, K-Means, NLP, topic modelling, Jupyter Notebook

- Built a scientific paper clustering pipeline using NLP embeddings, applying PCA and K-Means on **10K+ arXiv research papers** with multi-language preprocessing and TF-IDF variants, improving literature retrieval efficiency by **40%**

PUBLICATIONS

Malware Classification using ML with Feature Selection (Cuckoo Sandbox) | Procedia Computer Science, 2024

- KNN + Boruta automated feature selection for malware family prediction; **93%** precision, **96%** recall, **92%** F1 on phishing detection in Cuckoo sandbox environment

Brain Tumor Segmentation using Multi-Task Learning (GlobalNet, FusionNet) | IEEE CSNT, 2023 (Best Presentation and Paper Award)

- 3-architecture deep learning pipeline (**GlobalNet, Multi-task Learning, FusionNet**) for brain tumor segmentation; improved accuracy and resilience across multi-modal medical imaging

Speech Signal Segmentation using Kernelized Deep Networks and Clustering | IEEE CSNT, 2023

- Kernelized deep networks with **MFCC/LPC + spectral clustering (graph Laplacian)**; superior segmentation accuracy with reduced complexity on TIMIT and BURN benchmark corpora

Neural Image Captioning with Visual Attention for Accessibility | IJSRSET, 2023

- CNN + RNN** captioning model with visual attention for accessibility; converts image content to audio descriptions for visually impaired users via **2-phase** text-to-speech pipeline