

# Sneha Priyadarshini S S

Bengaluru, India • snehapriyadarshini99@gmail.com • 9087047407

## Technical Skills

- **Machine Learning & AI:** Supervised & Unsupervised Machine Learning, Statistical Modeling, Probabilistic Models (HMM), Data Preprocessing, Feature Engineering, Model Evaluation
- **Deep Learning:** ANN, CNN, RNN, GANs, Diffusion Models
- **Computer Vision:** Image Segmentation (U-Net), CNN Architectures (ResNet, Inception)
- **NLP:** Text Classification, BERT-based Embeddings
- **Programming:** Python (NumPy, Pandas, TensorFlow/Keras, Matplotlib), SQL
- **Enterprise Tools:** Informatica, Autosys, Ansible Tower, JIRA, Databricks

## Education

### Master of Technology in Data Science

Current CGPA: 9.09

Amrita Vishwa Vidyapeetham, Coimbatore

July 2024 - Present

### Bachelor of Engineering in Computer Science

CGPA: 7.45

Coimbatore Institute of Technology, Coimbatore

July 2016 – Sept 2020

## Professional Experience

### Intern

SAP Labs, Bengaluru

August 2025 – Present

(4 months)

- Working on **Intelligent Scenario Lifecycle Management (ISLM)**, SAP's framework that supports the **end-to-end lifecycle of ML models** including training, deployment, activation, inference, and monitoring.
- Assisted with onboarding and configuration of ISLM-supported scenarios such as **Generative AI, Predictive Analytics (PAL/APL), Business Entity Recognition, Document Information Extraction, Data Attribute Recommendation, and Personalized Recommendation**.
- Worked with **Prompt Registry** within SAP ISLM to manage, version, and reuse prompts for **Generative AI scenarios**.

### Test Analyst | Test Engineer

Infosys Limited, Bengaluru

March 2021 – July 2024

(39 months)

- Led a **Quality Assurance team** of 3 testers, collaborating with developers and stakeholders to enhance **data pipeline architecture** and ensuring smooth project planning and releases.
- Enhanced and optimized **complex SQL queries** for ETL workflows and increased the speed **by 20%**, laying a strong foundation for data preprocessing.
- Created **defect management strategies** and identified anomalies, improving data quality to **87%** making it reliable for decision-making.

## Projects

### Satellite Image Dehazing Using Haze Conditioned Latent Diffusion Models (Ongoing)

- Segmented hazy satellite images to identify thin, moderate and thick haze regions using U-Net Model to implement haze conditioned dehazing achieving an average IOU of 75.6.
- Encoded segmented inputs into a **latent vector space** for efficient sampling and reconstruction through U-Net based encoder-decoder
- Trained a **Latent Diffusion Model** to learn haze-to-clear image generation within the latent space.
- Decoded latent outputs back to image space and evaluated performance using **PSNR, SSIM, and MSE** metrics.

## Customer Purchase Prediction Using Hybrid HMM and Machine Learning

- Built a hybrid model combining **Hidden Markov Models (HMM)**, a Probabilistic Graphical Model, for sequential user behaviour and machine learning classifiers (Logistic Regression, XGBoost, LightGBM) for purchase intent prediction.
- Modelled event sequences to extract HMM log-likelihood features and session-level characteristics.
- Achieved highest accuracy of 94.91% with XGBoost and 94.81% with Voting Ensemble, confirming the strength of sequence-aware hybrid modelling for e-commerce analytics.

## Drug–Protein Interaction Prediction Using Machine Learning & Deep Learning

- Developed an end-to-end **computational drug discovery pipeline** using the **BindingDB dataset** to predict drug–protein binding interactions, reducing dependency on costly wet-lab experiments.
- Built **classical ML models** (Random Forest, XGBoost, LightGBM, Naïve Bayes) using **Morgan fingerprints** for drug molecules and **3-mer frequency vectors** for protein sequences, leveraging **statistical modelling** principles.
- Designed a **deep learning pipeline** combining **ChemBERTa** (for SMILES embeddings) and **ProtBERT** (for protein sequence embeddings) fused via MLP layers for interaction prediction.
- Achieved strong predictive performance, demonstrating that BERT-based bio–chem embedding fusion outperforms traditional handcrafted feature approaches for Drug–Target Interaction (DTI) prediction.

## Complex Word Identification Using Machine Learning and Deep Learning

- Built a Complex Word Identification (CWI) system to classify words as simple or complex using both Machine Learning and Deep Learning models for text simplification applications.
- Engineered linguistic, morphological, and contextual features (e.g., word length, syllable count, POS tags, readability scores) and integrated them with **BERT** embeddings for richer text representation.
- Implemented and compared classical ML models (**Logistic Regression, Random Forest, Gradient Boosting, XGBoost**) and DL models (**CNN, BiLSTM, FCNN**) using the English CWI dataset (~30K samples).
- Achieved the best performance with a Fully Connected Neural Network (**F1 = 0.81**), showing that context-aware deep models combining BERT embeddings and handcrafted features outperform traditional approaches.

## Certifications

- SAP Generative AI Developer – SAP (In Progress)
- 30 Days of Learning Databricks – Databricks Academy  
Focus: Apache Spark basics, data processing, and analytics workflows
- Infosys Certified Python Associate
- Generative AI Landscape certification - Infosys
- ETL Using Informatica (Infosys Certification)
- Business Communication Excellence Program Level 6