

Vikram Pande

vikramspande7@gmail.com | Portfolio | GitHub | LinkedIn | LeetCode | (984)-286-7150 | Charlotte, NC

Education

North Carolina State University

Master of Science in Electrical Engineering; **GPA: 3.97/4.00**

Aug 2022 – May 2024

Raleigh, NC

Courses: Random Processes, Image Processing, Neural Networks, Advanced Machine Learning, Object Oriented Development, Pattern Recognition, Cloud Computing, Natural Language Processing, Computer Vision, Probabilistic Graphical Models

Savitribai Phule Pune University

Bachelor of Engineering Electronics and Telecommunication; **GPA: 7.93/10.00**

Aug 2016 – May 2020

Pune, India

Courses: Machine Learning, Image Processing, Data Structures & Algorithms, Computer Networks, Linear Algebra, Vector Calculus

Experience

Dentsply Sirona

Machine Learning Engineer

June 2024 – Present

Charlotte, NC

- **Led development and productionization** of ML systems and infrastructure, owning end-to-end codebases; **adopted by multiple** internal sales teams.
- Built an LLM-RAG-based agent with a vector database for semantic search across 500+ product documents, handling 1,500+ monthly product queries with 85% accuracy, fine-tuned via RLHF with actual user feedback.
- Improved recommendation precision by **25%**, achieving **0.83 Precision@K** and **0.64 NDCG**, by designing a **two-tower neural recommender** system in **PyTorch**.
- Automated email summarization and classification with **LLMs**, cutting manual effort by **~60%**, and streamlining reporting.
- Boosted churn prediction accuracy by **18%** on **100K+ records** using probabilistic models (BG/NBD) and RFM features; improved product sales forecasting accuracy by **20%** using statistical and ML models (ARIMA/Prophet) in **PySpark**.
- Reduced model deployment time by **40%** by implementing scalable **MLOps pipelines** for legacy and new models.

Sozzani Lab, NCSU

Research Assistant

Sep 2023 – May 2024

Raleigh, NC

- Achieved **96% F1-score** in **protein sequence classification** by training a custom **CNN-Attention-LSTM** model on Arabidopsis data; published in [Nature](#).
- Reduced model training time by **60%** by parallelizing **neural network training** using **High Performance Computing**.
- Developed an **Autoencoder & KMeans pipeline** to identify and cluster **plant subtypes** for improved phenotype mapping.

Syngenta

Data Science Intern

June 2023 – Dec 2023

Durham, NC

- Applied **HDBSCAN** clustering and **t-SNE** dimensionality reduction to **25k×25k** genomic datasets for **subgroup discovery**.
- Optimized data processing by **55%** by automating **ETL pipelines** for genomic data using **Python** and **SQL**.
- Built **interactive Tableau dashboards** to display **real-time KPIs** for stakeholders across **4 countries**.

Active Robotic Sensing Lab, NCSU

Research Assistant

Jan 2023 - Sep 2023

Raleigh, NC

- Conducted research to segment damaged and healthy leaves of the **St. Augustine turfgrass** under drought conditions.
- Implemented **HSV space** thresholding and mean clustering for segmentation and achieved an average **SSIM of 72%**.
- Utilized YOLOv8 for object detection and achieved **mAP of 89.8%**, studied data augmentation techniques and transfer learning with **ViT-16** vision transformer for semantic segmentation.

Accenture

Machine Learning Engineer

May 2021 – June 2022

Pune, India

- Improved overall performance by **15%** of an **AI-based document processing platform** serving **25+ enterprise clients**.
- Achieved **82% F1-score** in multi-class document classification by developing an **OCR & BERT-based NLP pipeline**.
- Improved email sentiment analysis model accuracy by **15%** by implementing **TF-IDF, WordNet, and Naive Bayes**.
- Developed an **NER Module** using **LayoutLMv2** to extract entities, boosting parsing accuracy by **25%**, enabling automation.
- Built an OCR-engine **recommender system** using **RandomForests**, reducing client-side processing time by **20%**.

Projects

[NLP] **LOLgorithm: Humor Classification** (Python, PyTorch, TensorFlow, SciKit-Learn) [[ArXiv](#)]

Dec 2023

- Leveraged **ColBERT** dataset to examine the humor content in a sentence and verify the linguistic theory of humor.
- Created hand-crafted **syntactic and semantic** features modifying the embeddings from **NRClex, Word2Vec, and WordNet**
- Utilized **contextual BERT embeddings** and improved model accuracy by **14%** with all features using **Colbert model**.

[Computer Vision] **Explainable AI for DeepFake Detection** (Python, PyTorch) [[GitHub](#)]

Nov 2023

- Achieved an F1 score of **98%** with **XceptionNet** for deepfake detection on **FaceForensics++** and **Celeb-DF** datasets.
- Applied **Explainable AI (XAI)** methods such as **GradCAM, LIME, and LRP** to highlight the relevance of input to the prediction and improved transparency and interpretability.

Tehchnical Skills

Programming Languages : Python, C++, C#, R, MATLAB, SQL

Libraries : Scikit-learn, Matplotlib, Seaborn, Langchain, LlamaIndex, NLTK, SpaCy, Hugging Face Transformers, MLFlow, OpenCV, Pillow, MMCV, OpenVINO, Detectron2, AutoML

Tools & Frameworks: PyTorch, TensorFlow, Keras, Databricks, PySpark, Azure, AWS, Tableau, Git, Docker, Kubernetes, HPC