SHREYAS TEMBHARE

sst9704@g.rit.edu — +1 (585)568-9782— linkedin.com/in/shreyas-tembhare/ — github.com/ShreyasTembhare

EDUCATION

Rochester Institute of Technology, NY, USA

Expected: May 2026

Masters of Science in Data Science

GPA: 4.00

Savitribai Phule Pune University, Pune, India

July 2024

Bachelors of Engineering in Artificial Intelligence and Machine Learning (AI & ML)

GPA: 3.6

SKILLS

- **Programming Languages and Databases:** Python, Java, SQL
- **Libraries and Tools:** Scikit-learn, TensorFlow, PyTorch, Pandas, NumPy, Matplotlib, Seaborn, Plotly, PdfPlumber, PyPDF2, LightKurve, Streamlit, AWS SageMaker, Tableau, Power BI, Apache Spark, Git
- **Machine Learning:** Deep Learning (CNN, RNN, LSTM), NLP (TF-IDF, Cosine Similarity, Transformers), Time Series Forecasting, Statistical Modeling (Regression, Classification, Clustering), Hypothesis Testing, Model Optimization (Beam Search, Word Embeddings), Model Tuning, Cross-Validation
- **Data Manipulation & Analysis:** Data Wrangling, Feature Engineering, Missing Data Handling, Outlier Detection

EXPERIENCE

Rochester Institute of Technology

Rochester, NY

Research Assistant

October 2024 - Current

- Conducting research on multimodal machine learning models, focusing on token length variations and retrieval accuracy optimization.
- Evaluating state-of-the-art models (CLIP, OpenCLIP, Long-CLIP, ALIGN, BLIP-2) on diverse datasets (ShareGPT4V, ROCO, Fakeddit, Urban1k) to analyze model robustness and retrieval efficiency.
- Investigating the impact of dataset variations on retrieval thresholds, identifying computational bottlenecks affecting ML performance.

TechR

Remote

Machine Learning & Software Engineering Intern

Apr 2023 - Jul 2023

- Designed and optimized an ML-powered Resume Analyzer using Python, TF-IDF, Cosine Similarity, improving text processing speed and classification accuracy by 25%.
- Developed and fine-tuned a BERT-based text summarization model, reducing computation time by 30% through model optimization techniques.
- Deployed an AWS cloud-based NLP model, streamlining the resume-job matching process and increasing user adoption by 25%.

YBI Foundation

Remote

Machine Learning Intern

Feb 2023 - Mar 2023

- Developed an ML-driven predictive model using Python, Scikit-learn, and hyperparameter tuning, demonstrating strong foundations in optimization mathematics and resource scheduling.
- Processed large-scale data pipelines using NumPy, Pandas, and Matplotlib, focusing on scalable ML solutions.
- Implemented feature selection techniques to reduce model complexity while maintaining high predictive accuracy.
- Conducted statistical validation (ANOVA, confidence intervals) to ensure model reliability and optimize decision-making.

PROJECTS

Exoplanet Detection Tool:

- Optimized compute graph processing by implementing parallelization and caching, reducing computation time by 40%.
- Developed custom outlier detection algorithms and noise reduction techniques, increasing model accuracy by 25% and improving large-scale dataset efficiency.
- Created advanced visualizations using Matplotlib and Seaborn, incorporating statistical confidence intervals for improved data interpretation and anomaly detection.

Resume Analyzer:

- Developed and deployed an NLP-powered resume analysis system using Streamlit and AWS SageMaker, reducing processing time by 40% and improving classification accuracy by 15%.
- Engineered an ML model using TF-IDF, Cosine Similarity, and Scikit-learn, achieving high-precision text classification for resume-job matching, aligning with AWS's focus on NLP optimization.
- Fine-tuned the NLP pipeline with domain-specific embeddings, enhancing contextual understanding and model generalization, demonstrating expertise in data-driven ML refinement.

Ecoharvestors:

- Designed an AI-powered crop yield prediction model using Scikit-learn and XGBoost, improving predictive accuracy by 30%.
- Implemented a deep learning model in TensorFlow to analyze agricultural factors, achieving a 25% accuracy increase over traditional models.
- Developed a high-efficiency data preprocessing pipeline using Pandas and NumPy, reducing feature engineering time by 40%.
- Deployed the model on AWS with Docker, improving scalability and ensuring reliable performance for large datasets.