RISHITA PRIYADARSHINI SARAF

rishitasarafp@gmail.com | P: +91-9490463841 | www.linkedin.com/in/rishita-priyadarshini-saraf | github.com/Rishita-P-Saraf

EDUCATION

Vellore Institute of Technology, Bhopal 2022- 2026
BTech In Computer Science and engineering | CGPA: 8.76 (expected)

TECHNICAL SKILLS

- **Programming Languages:** Python, C++
- **Databases:** MySQL
- **Technical Skills:** LLMs, RAG, Generative AI, Data Analysis, Microsoft Excel, Power BI, SQL, Deep Learning, Machine Learning Algorithms, NLP, Image Processing
- **Tools & Frameworks:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, TensorFlow, Keras, OpenCV, LangChain, Flask, Streamlit, FastAPI, Plotly, Power BI, Microsoft Excel

UNIVERSITY PROJECTS

Hindi Word Embeddings with FastText August, 2025

- Trained CBOW and Skip-gram models on large-scale Hindi corpus using FastText, capturing semantic and syntactic word relationships.
- Designed a text preprocessing pipeline (cleaning Unicode, punctuation, normalization) to prepare raw Hindi articles for embedding training.
- Evaluated embeddings via nearest neighbor queries, demonstrating meaningful clustering of related terms in Hindi NLP.
- Tech Stack: Python, FastText, Pandas, NumPy, Regex

MRI Brain Tumor Detection using Deep Learning June, 2025

- Developed a custom Convolutional Neural Network (CNN) in PyTorch to classify MRI brain scans as *Tumor* or *Healthy*, achieving 100% accuracy on the validation dataset.
- Designed a scalable preprocessing pipeline with a custom PyTorch Dataset class for image loading, resizing, and augmentation, enabling efficient GPU-accelerated training on large MRI datasets.
- Conducted extensive model evaluation using confusion matrices, loss/accuracy plots, and convolutional feature map visualizations to ensure interpretability and reduce risks of overfitting.
- Tech stack: Python, PyTorch, NumPy, OpenCV, scikit-learn, Matplotlib, Seaborn

Driver Drowsiness Detection using YOLOv5 July, 2025

- Implemented a real-time driver monitoring system using the YOLOv5 object detection model to classify driver states (*awake vs drowsy*) from webcam feeds.
- Collected and annotated a custom dataset with LabelImg, designed dataset configuration files, and trained YOLOv5 models with optimized hyperparameters for reliable detection.
- Integrated the trained model into a live inference pipeline with OpenCV, enabling continuous monitoring and laying the foundation for future safety alerts on drowsiness events.
- Tech stack: Python, PyTorch, YOLOv5, OpenCV, NumPy, LabelImg

WORK EXPERIENCE

Data Analyst Intern - NullClass (Remote) Jan 2025 – Feb 2025

- Analyzed Google Play Store app reviews and created an interactive HTML dashboard using Plotly.
- Performed data preprocessing and extracted sentiment trends to guide app improvements.
- Tech Stack: Python, Pandas, Plotly, Matplotlib, Seaborn, Jupyter Notebook

Data Science Intern - Cognifyz Technologies (Remote) Dec 2024 – Jan 2025

- Developed a linear regression model to predict restaurant ratings using customer data
- Conducted data cleaning, feature selection, and visualized insights using Seaborn and Matplotlib.
- Technologies: Python, Pandas, Matplotlib, Seaborn, Scikit-learn

ADDITIONAL INFORMATION

Certifications:

- IBM GEN AI Using IBM Watsonx certificate
- Oracle Data Science Professional Certificate

Achievements: Ranked 78th in the Zelestra Hackathon on HackerEarth for building a solar panel efficiency prediction model using XGBoost, achieving 89.88% accuracy.