# VOONA SRIRAJ

Github | Linkedin | srirajvoona2004@gmail.com | +91 637041597

## Professional Experience

**Machine Learning Research Intern** **August 2024 - December 2024**
SRM University AP Guntur, Andhra Pradesh

- Performed sentiment analysis on Twitter dataset analyzing public sentiment toward COVID-19 vaccinations using Python and machine learning algorithms, achieving high accuracy in sentiment classification
- Implemented comprehensive data preprocessing pipeline including text cleaning, tokenization, and feature extraction for social media data analysis
- Developed Logistic Regression model with cross-validation and performance optimization, creating actionable insights for public health communication strategies

## Education

**SRM University AP** **December 2022 - Present**
Bachelor of Technology in Computer Science and Engineering CGPA: 8.4
**Gandhi Public School** **2020 - 2022**
Intermediate CBSE Gunupur, Odisha

## Technical Skills

- **Programming Languages:** C++, Python, SQL, Java, HTML, CSS
- **Data Analysis & Visualization:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, SQL, Excel, Power BI, Tableau
- **Machine Learning & Deep Learning:** Regression, Classification, XGBoost, Random Forest, ANN, CNN, RNN, Transformers; end-to-end model building and evaluation.
- **Generative AI & NLP:** LangChain, OpenAI API, Hugging Face, FAISS, Retrieval-Augmented Generation (RAG), AI Agent Design, Text Preprocessing, Tokenization, Prompt Engineering
- **MLOps & Deployment:** Streamlit, Flask, GitHub Actions, Docker
- **Cloud Platforms:** AWS (S3, EC2 basics)
- **Databases:** MySQL, MongoDB

## Technical Projects

**WebAssist RAG-Powered AI Web Assistant** | *Streamlit, LangChain, FAISS, HuggingFace, LLaMA*

- Built Streamlit-based chatbot leveraging LangChain, FAISS vector database, and HuggingFace embeddings for semantic search over C++ documentation with context-aware responses
- Integrated ChatGroq LLaMA-3 model using Retrieval-Augmented Generation architecture delivering real-time LLM querying with source document visibility
- Engineered end-to-end ML pipeline including document loading, text chunking, vector storage, and semantic search capabilities for enhanced question-answering performance

**Medical AI Chatbot** | *Flask, Groq (Llama 3.1 8B), Pinecone, LangChain, Sentence Transformers, PyPDF, Docker,AWS*

- Engineered a Retrieval-Augmented Generation (RAG) architecture integrating Groq's Llama 3.1 8B model with Pinecone vector database to achieve contextually accurate medical responses through semantic similarity search.
- Automated processing of medical PDF documents using PyPDF and LangChain for text extraction, chunking, and embedding generation with Sentence Transformers for semantic search.
- Developed a clean, responsive chat interface with real-time messaging using Flask and Bootstrap, supporting concise and relevant answers sourced from the medical knowledge base.
- Deployed scalable cloud infrastructure using Docker containerization, automated CI/CD pipeline through GitHub Actions, and AWS services (ECR, EC2) with environment-based configuration management for production deployment.

**Visual Image Caption Generator using Deep Learning** | *Python, TensorFlow, Keras, OpenCV, COCO Dataset*

- Developed a CNN-LSTM model that generates natural language captions for images by extracting features with a convolutional neural network and generating text with LSTM, using the COCO 2017 dataset.
- Implemented an efficient data preprocessing pipeline including grayscale image conversion, caption tokenization, sequence padding, and dataset augmentation for improved model training.
- Applied deep learning techniques with Adam optimizer, categorical crossentropy loss, and ReLU activation to enable accurate image feature extraction and meaningful caption generation for diverse image inputs.

## Certifications and Training

- **Machine Learning and Deep Learning** - IBM SkillBuild
- **Natural Language Processing and Computer Vision** - IBM SkillBuild
- **A Joy of Learning Python** - NPTEL