Sundar Velmurugan

246 , Arugeri , Cuddalore -606111, TN| velsundar490@gmail.com | +91-9360784554 | LinkedIn | GitHub

About

- Dedicated Data Science professional with a robust foundation in mathematics and proficiency in Python, Generative AI with Large Language Models, Machine Learning, and Deep Learning.
- Skilled in building predictive models, developing AI-driven applications, and conducting real-time data analysis.
- Eager to leverage analytical expertise and technical capabilities in a dynamic, innovative environment.

Experience

VDart – AI/ML Intern **Jan 2025 –May 2025**

- Gained hands-on experience in machine learning, deep learning, and AI model development.
- Worked with data preprocessing, model training, and performance optimization.
- Explored computer vision and NLP techniques for real-world AI applications.
- Collaborated on AI-driven solutions, enhancing problem-solving and analytical skills.

Projects

Smart AI Virtual Interview System for Recruitment and Skill Assessment

- I built the Smart AI Interview System, a web platform that automates aptitude, Q&A (based on job descriptions), and HR interviews using AI to make hiring faster and fairer. Candidates log in, check their camera and microphone, and take their interviews. Admins can schedule interviews and view results on a dashboard with charts and performance statistics. The system uses TF-IDF, Sentence Transformers, and LLaMA 2 for answer evaluation, along with YOLOv5 and Faster R-CNN for monitoring—automatically ending interviews after 3 violations. Speech is transcribed using the Google Web Speech API. The platform is built with FastAPI and Bootstrap.

Multimodal PDF Q&A Chatbot using OCR and LLM

- I built a PDF Q&A Chatbot that lets users upload PDFs and ask questions. It extracts text (including from images using OCR), generates embeddings with Hugging Face and FAISS, and uses Llama 2 via Ollama to answer questions based on document content. Built with Streamlit and LangChain.

Real-Time Seatbelt Detection with Fine-Tuned Vision Transformer

- Built a seatbelt detection model using a pretrained Vision Transformer combined with fully connected layers for binary image classification. Applied data augmentation, fine-tuning, and training techniques like early stopping. Achieved high accuracy on real-world images using PyTorch.

Developed a multimodal AI-powered traffic accident prevention system

- Integrated Whisper for audio, LLaVA for image-to-text processing, and deep learning models (Transformers, ARIMA, Prophet, LSTM, TCN) for accident forecasting. Analyzed the FARS 2020 dataset, achieving high prediction accuracy, and built a real-time interactive system with Gradio to enhance autonomous driving safety.

Disease Prediction System using XGBoost and Streamlit

- Preprocessed symptom data with one-hot encoding, handled class imbalance with SMOTE, and optimized features using SelectKBest. Tuned hyperparameters with RandomizedSearchCV and evaluated the model using F1 scores. Deployed a user-friendly Streamlit app for real-time disease prediction.

Technical Skills

Programming Languages: Python, SQL, R Language
Data Analysis Tools: PowerBI (Beginner), Excel (Beginner).
Tools and software: VS code, GitHub
Operating System: Windows

Education

SASTRA University | *Master of Science* **2023 - 2025**
M.Sc. Data Science *– CGPA:8.25* *Thanjavur, Tamil Nadu*

A.V.V.M Sri Pushpam College | *Bachelor of Science* **2020 - 2023**
B.Sc. Mathematics *– CGPA: 8.32* *Thanjavur, Tamil Nadu*

Certifications

- OSINT Using Python
- Workshop in Clinical Data Analytics (2024)
- Mathematical techniques