

SUDARSHAN VEMARAPU

Hyderabad, India | +91 6305434790 | sudarshanvemarapu@gmail.com
Portfolio: bheemisme.in | GitHub: github.com/bheemisme | LinkedIn: linkedin.com/in/bheem-is-me

PROFESSIONAL SUMMARY

Adaptable and results-driven AI Engineer with hands-on experience spanning Machine Learning, Generative AI, Data Science, and Full Stack Development. Experienced in finetuning models, AI agent development, RAG pipeline development etc. Strong foundational expertise in Python, machine learning frameworks, and database management.

TECHNICAL SKILLS

Languages:	Python, JavaScript, TypeScript, SQL
Machine Learning & GenAI:	PyTorch, TensorFlow, Scikit-learn, OpenCV, NLTK, LangChain, Transformers, Fine-Tuning, LoRA, LLMs
Cloud & DevOps:	GCP (Vertex AI, Cloud Run, App Engine), AWS (EC2, Lambda), Heroku, Git, Docker, REST APIs
Data Analysis & Vis:	Pandas, NumPy, Matplotlib, Seaborn, Plotly
Database Systems:	MySQL, MongoDB, DynamoDB, Firestore
Web Development:	React.js, Next.js, React Query, Tailwind CSS, Express.js, Nest.js, FastAPI

PROFESSIONAL EXPERIENCE

Supaboard

Machine Learning Engineer Intern

Bengaluru, India

Jan 2025 – March 2025

- Designed and developed an advanced multi-class sentiment analysis model to predict over 28 distinct emotion tokens, expanding beyond traditional polarity-based classifications.
- Engineered a large-scale, balanced dataset of 111,866 training samples, 11,198 validation samples, and 5,598 test samples by scraping and aggregating reviews (Amazon, Flipkart, ProductHunt) and social media feeds (Twitter).
- Utilized data augmentation techniques and AI-assisted labeling to eliminate class imbalances and improve training dataset quality.
- Conducted extensive model experimentation establishing baseline metrics with FastText and zero/few-shot LLM classification benchmarks.
- Fine-tuned BERT-base and RoBERTa-base transformer architectures via Low-Rank Adaptation (LoRA) and custom classification layers, achieving an overall evaluation accuracy exceeding 85%.
- Explored alternative modeling strategies including masked language modeling for emotion token inference to benchmark against direct classification.

Unified Mentor

Data Science Intern

Remote

June 2024 – July 2024

- Performed advanced descriptive analysis and derived actionable insights from multiple large, structured datasets.
- Developed interactive dashboards using Python's Plotly library to translate raw data trends into clear, visualization-driven business storytelling.
- Conducted a comprehensive bird strike analysis to uncover correlation patterns across flight altitudes, atmospheric factors, and aircraft phases of flight.
- Analyzed heart disease datasets to identify and quantify relationships between health-related patient attributes and clinical disease occurrence.

- Delivered an employee attrition analysis mapping correlations between job-related factors (salary, work-life balance, tenure) and workforce attrition rates.
- Deployed all production dashboards to Heroku, enabling remote, interactive access to data insights for decision-makers.

Fyipen Pvt. Ltd.

Full Stack Developer Intern

Remote

March 2023 – May 2023

- Contributed to full-stack engineering efforts across three production web applications: a tourism portfolio, an ed-tech platform, and an e-commerce platform.
- Leveraged a diverse technology stack including Next.js, React.js, Nest.js, FastAPI, and MongoDB.
- Resolved critical frontend hydration errors within a Next.js tourism platform and engineered robust API testing suites to guarantee system stability.
- Implemented frontend wireframes, built reusable React components, and integrated React Query for asynchronous client-side API state management.
- Developed highly optimized backend REST API routes using FastAPI and Nest.js, implementing custom server-side caching mechanisms to drastically minimize response times.
- Applied backend development best practices in application logging, performance optimization, and scalability to deliver production-ready code.

KEY PROJECTS

Contract Intelligence Platform (Generative AI & LLMs)

[GitHub](#)

- Engineered a Generative AI legal technology solution to manage, analyze, and validate legal contracts against specific regulatory standards.
- Built an autonomous AI agent using the LangChain framework to extract metadata and drive an interactive contract-analysis chat interface.
- Deployed the solution architecture as a high-performance REST API.

Book Recommendation System (Deep Learning)

[GitHub](#)

- An end-to-end recommendation system for Amazon Book Reviews, built entirely on Google Cloud Platform (GCP). The project covers the full machine learning lifecycle: automated data ingestion from Kaggle, data cleaning and feature engineering in BigQuery, scalable model training using Vertex AI, and an API/inference layer for serving predictions.

Unpaired Image-to-Image Translation (Computer Vision / GANs)

[GitHub](#)

- Developed a computer vision style-transfer architecture from scratch using PyTorch to perform unpaired image translations (Monet paintings to real photos and vice versa).
- Implemented Cyclic Generative Adversarial Networks (CycleGANs), training the architecture on a T4 GPU to achieve a Fréchet Inception Distance (FID) score of 63 at approximately 30 training epochs.

Neural Lossy Compression (Deep Learning / VAEs)

[GitHub](#)

- Designed and trained a lossy image compression model leveraging a Variational Autoencoder (VAE) optimized on the Fashion-MNIST dataset.
- Executed 20 hours of continuous model training on a T4 GPU, reaching a stable training loss of 3.1 and test loss of 3.3 while maintaining high visual similarity to original input images.

EDUCATION

M.Sc. in Computer Science

GPA: 7.925 / 10

University of Madras - Chennai, India

May 2025

B.Sc. in Computer Science

GPA: 9.6 / 10

MIT World Peace University - Pune, India

May 2023