Siddharamayya Mathapati

AI/ML Engineer

msidrm455@gmail.com | +91 9740671620 | Belagavi, KA

Portfolio | GitHub | Linkedin

PROFESSIONAL SUMMARY

Innovative AI Engineer with 4+ years of experience in designing, developing, and deploying cutting-edge AI models and systems. Proficient in Large Language Models(LLMs), Natural Language Processing (NLP), Deep Learning, and MLOps, with a proven track record of building scalable AI solutions that improve prediction accuracy by 25%, reduce latency by 15%, and cut cloud costs by 25%. Passionate about leveraging AI to solve complex business problems and deliver impactful, data-driven solutions.

EDUCATION

ACHARYA INSTITUTE OF TECHNOLOGY   2018 - 2020
Master's Computer Applications   Bangalore
CGPA: 7.2

BASAVPRABHU KORE COLLEGE   2015 - 2018
Bachelor Computer Application   Chikodi
CGPA: 6.8

SKILLS

AI/ML	Machine Learning, Deep Learning, Natural Language Processing (NLP), Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), Computer Vision, Reinforcement Learning
Frameworks	TensorFlow, PyTorch, Hugging Face, LangChain, OpenAI GPT, Scikit-learn
Cloud Platforms	AWS SageMaker, Google Cloud Platform (GCP), Microsoft Azure, Firebase
Tools	Docker, Kubernetes, Jenkins, Git, MLflow, Airflow, RabbitMQ, Postman
Databases	MySQL, PostgreSQL, MongoDB, NoSQL, VectorDB, ChromaDB, FAISS.
Soft Skills	Team Collaboration, Leadership, Problem-Solving, Communication, Agile Methodologies, Cross-functional Coordination

Technical Tools

AI/ML Tools	TensorFlow, PyTorch, Hugging Face, OpenAI GPT, LangChain
Cloud Platforms	AWS SageMaker, Google Cloud Platform (GCP), Microsoft Azure
DevOps Tools	Docker, Kubernetes, Jenkins, Git, MLflow, Airflow
Databases	MySQL, PostgreSQL, MongoDB, NoSQL, VectorDB, ChromaDB

Languages

English - Advance	Kannada – Native	Telugu – Advance	Malayalam - Beginner
Hindi – Advance	Tamil - Intermediate	Marathi - Advance	

EXPERIENCE

CAPGEMINI TECHNOLOGY SERVICE LIMITED INDIA | sENIOR SOFTWARE ENGINEER Navi Mumbai | May 2021 – Present

- Developed and deployed machine learning models for predictive analytics in a cloud environment using AWS SageMaker, improving prediction accuracy by 25% and reducing processing time by 40%.
- Engineered advanced conversational AI solutions using LLMs, improving response accuracy and contextual relevance for industries like banking, resulting in a 35% increase in customer satisfaction and a 20% boost in engagement metrics.
- Fine-tuned pre-trained models using Hugging Face, enhancing NLP capabilities in custom applications, leading to a 30% increase in model performance for text classification tasks.
- Led the design and optimization of AI-powered features like recommendation systems, reducing latency by 15% and improving user engagement by 20%.
- Implemented Retrieval-Augmented Generation (RAG) to enhance the reliability of responses in NLP systems, reducing hallucinations by 50% and increasing the overall efficiency of information retrieval by 40%.
- Optimized cloud infrastructure with AWS services for scalable and cost- efficient AI model deployment, achieving a 25% reduction in cloud costs and 30% improvement in processing speed.
- Orchestrated DevOps workflows using Docker and Kubernetes, ensuring seamless and efficient deployment of AI-driven applications, cutting deployment times by 45%.
- Automated end-to-end testing using Python-based frameworks, improving the reliability and stability of AI models, with a 20%reduction in model errors during deployment.
- Automated **end-to-end testing** for AI models using Python-based frameworks, reducing model errors during deployment by **20%**.
- Created **AI-powered recommendation systems** for e-commerce platforms, increasing user engagement by **20%** and reducing latency by **15%**.

X-Ciencia Technology|AI RESEARCH INTERN | March 2020 – April 2021
- Researched and implemented **deep learning models** for computer vision tasks, achieving **90% accuracy** in image classification.
- Developed **reinforcement learning algorithms** for autonomous systems, improving decision-making efficiency by **25%**.
- Published research papers on **AI ethics and bias mitigation** in machine learning models, contributing to the development of fairer AI systems.

PROJECTS / OPEN-SOURCE

RECRUITMENT_HELPER_AI_WITH_LLAMA | LINK   Jupyter Notebook, Python
- AutoRecruiterAI is an AI-powered solution designed to automate the process of extracting job information from career pages and generating personalized cold emails for potential clients. It saves time, enhances efficiency, and eliminates manual effort in collecting job data and creating outreach emails.

STOCK-PRICE-PREDICTION-USING-LLMS | LINK   Python, PowerShell, Roff, Shell, Batchfile, JavaScript
- This project is an AI-powered finance agent that leverages large language models (LLMs) to analyze stock market trends, predict future prices, and provide insights into stocks. It integrates data from various sources, including Yahoo Finance for stock data and NewsAPI for stock news articles. The agent uses this data to make predictions and assist in financial decision-making.

CREDIT-RISK-ANALYSIS-WITH-CATBOOSTCLASSIFIER | LINK   Python
- This project focuses on predicting the likelihood of credit card defaults using machine learning. With

financial institutions heavily reliant on accurate predictions to minimize risk and improve profitability, this project provides an automated solution to assess creditworthiness effectively.

LINKEDIN-COMMENT-POSTER-WITH-AI | L<small>INK</small>   Python

- The LinkedIn Activity Scraper is a Python-based automation tool that extracts user activity data, including posts, likes, shares, and comments, using Selenium and BeautifulSoup. It ensures dynamic content handling, error management, and ethical scraping practices for insightful analysis.

MEDICAL-ASSISTANT-AI | L<small>INK</small>   Python

- This project is a web application that utilizes Google Generative AI to assist medical professionals by analyzing medical images and providing diagnostic insights. The application allows users to upload images and receive analytical feedback through a chat interface.

STUDENT_STUDY_HELPER_AI_WITH_GOOGLE_GEMINI | L<small>INK</small>   Python

- This Project provides a chatbot built using LangChain, Gemini LLM, and PDF processing to answer questions based on uploaded PDF documents. The chatbot is designed to help students by answering their queries from the context of the uploaded files.

SIMPLE_SQL_WITH_GOOGLE_GEMINI_AI | L<small>INK</small>   Jupyter Notebook, Python

- This project demonstrates how to integrate Google Gemini LLM with a MySQL database using LangChain for natural language queries. It uses semantic similarity-based few-shot learning to generate accurate SQL queries and fetch results from the database.

CERTIFICATIONS

- LLM Engineering: Master AI, Large Language Models & Agents - U<small>DEMY</small>
- MLOps Bootcamp: Mastering AI Operations for Success Udemy - U<small>DEMY</small>
- Python for Data Analysis & Visualization - U<small>DEMY</small>
- Deep Learning Masterclass with TensorFlow 2 - U<small>DEMY</small>