

MADHUSUDAN KUMAR

Machine Learning Engineer

✉ max12525k@gmail.com

☎ +91 9969072292

🔗 max12525k

🌐 max12525k



Machine Learning Engineer with 5.6 years of experience in applying machine learning, natural language understanding, Generative AI and software engineering to solving problems in smart customer interaction, threat mapping, predictive maintenance and revenue forecasting. Major career highlights include:

- Designed data generation service for rapid dataset delivery model finetuning for given use case.
- Advanced research on multimodal conversational agent by extending previous research.
- Led cross functional team to deliver AI-powered application with 50+ backend APIs.
- Developed Python Package to reduce model integration time by 60%.

SKILLS

Python

PyTorch

LangChain

GenAI

RAG

Transformers

HuggingFace

Pandas

Scikit-learn

FastAPI

VectorDB

MongoDB

MYSQL/PostgreSQL

Git

CI/CD

AWS

Linux

Docker

PROFESSIONAL EXPERIENCE

Machine Learning Engineer

Meta | Unifytech

📅 Nov 2023 - Ongoing

📍 Mumbai

🔗 Project Management

- Managed 5+ member team of ML engineers and annotators, achieving 10% acceleration in project milestone delivery.

🔗 AI for Smart Glasses

- Spearheaded the design and development of scalable data generation service using LangChain, tailored to support diverse finetuning use case, leading to significant reduction in dataset (100K+ samples) delivery time.
- Designed and implemented robust evaluation framework, leveraging quantitative and qualitative metrics to assess large-scale data generation, resulting in enhanced dataset quality by 40% while reducing annotation time by 60%.
- Fine-tuned LLaMA 3 (8B) model on custom dataset, achieving 86% performance improvement in answering user queries in structured format and deployed containerized model for seamless company-wide access on Meta Infra.

🔗 Research on Conversational Agents

- Validated findings from research papers by conducting ablation study and identifying key gaps for investigation.
- Extended text-based agent architecture (COOPER) research to build multimodal conversational agents, showcasing performance gains in specialized domains from existing agents.

🔗 HRMS Chatbot

- Developed a modular, distributed document processing pipeline with FastAPI, MongoDB and FAISS, streamlining embedding generation and storage for efficient scalability.
- Boosted HR Policy QA performance through strategic RAG implementation and optimized COT prompts.

LLAMA3

LangChain

Transformers

PEFT

Cosine Similarity

LLM Evaluation

Docker

FastAPI

Senior Data Scientist

SecureThings.AI

📅 Sep 2022 - Sep 2023

📍 Mumbai, India

🔗 Knowledge Graph Generation with AI

- Built knowledge graph generation pipeline leveraging SpaCy, extracting 50+ cybersecurity labels and 20+ relationships from 2M+ semi-structured texts with 70%+ accuracy, reduced extraction time by using OpenAI and LangChain.

🔗 Product Development and deployment

- Built and deployed distributed backend application using FastAPI, AWS Cognito, Docker and MongoDB, supporting a multi-screen product (15+) with extensive API integration (50+), ensuring seamless user experience.

- Deployed AWS pipeline using IaC, reducing management time by 70%, costs by 50%, and deployment time by 80%.

NLP NER PyTorch SpaCy LangChain FastAPI MongoDB Docker IaC LLM OpenAI

Senior Data Scientist

Tata Consultancy Services

📅 Apr 2019 - Sep 2022

📍 Mumbai, India

🔗 AI Platform for Telecommunication

- Led the team of 5 developers for web-based ML exploration tool for 40% faster deployment of ML solution.
- Designed 10+ ML solutions for telecom fault and performance management leading to decrease in 15% reduction in events and enable close loop automation for 25% events across 5000+ sites.

Data Scientist

🔗 AI Platform for Telecommunication

- Proposed, designed & built Python package for exploration & integration of 24+ ML use cases across 4+ applications, reducing model training and integration time by 60%.
- Enabled one day ahead predictive maintenance, reduced ticket generation by 6% and optimized network performance using forecasting, clustering and anomaly detection techniques, reducing the operation cost by 18% (POC).

🔗 ML for Revenue Forecasting in Insurance

- Automated and optimized monthly forecasting of claims paid and premium collected, reducing average MAPE by 10%

Xgboost Scikit-learn SARIMAX Mlxtend Anomaly Detection Plotly Pyod Pytest AZURE VM

PATENT

Patent

- Awarded patent: "METHOD AND SYSTEM FOR ANOMALY DETECTION OF NETWORK KPI" Intellectual Property India Patent No.:202221056578

EDUCATION

Internship

Chiba University

👤 Prof. T. Sakurai

📅 May 2016 - Jun 2016

📍 Chiba, Japan

- Explored linear diffusive coupled neurons (FHN coupled system) with external forcing to study pattern formation
- Conducted nonlinear stability analysis to obtain significant parameters & visualize their behavior near stability points to understand contribution to pattern formation

Visualization MATLAB Linear algebra Overseas Experience

M.Tech in Mechanical Engineering

IIT-B

📅 Jul 2017 - Nov 2018

- Proposed and implemented OWL-ontology based template to digitally store properties of machine component from different vendors, optimizing the digital processing of components information.
- Designed and built process to predict the failure time of new machine on the network based on other machine failure data by modelled estimation of reliability of mechanical components from loading condition and similarity of components, minimizing the operation downtime.

B.Tech in Mechanical Engineering

IIT-B

📅 Jul 2013 - Jun 2017

- Attended courses related to statistics, data analysis, computer programming and mechanical engineering.
- Conducted research in non-linear dynamics, brain entrainment using mathematical modeling and simulations.
- Headed official video channel of IIT-BBC, leading a team of 6 conveners, creating 60+ videos and covering 35+ events, with nearly 2 million views.