Shubham Salokhe

MACHINE LEARNING ENGINEER

852/1 'B', Subhash Road, Raviwar Peth, Kolhapur 416012

8552974448 | salokheshubham00@gmail.com | shubhamsalokhe.github.io | ShubhamSalokhe | shubhamsalokhe | shubhamsalokhe

Summary

Experienced ML engineer with 4+ years of practical experience in machine learning, deep learning, and NLP. Seeking a challenging role to create advanced AI-ML solutions, enhance customer experience, and drive business success through collaboration within a dynamic team.

Work Experience

Xoriant Software Technology
Pune, Maharashtra

AI/ML Engineer
April. 2024 - Present

- Designed and implemented a Human-in-the-Loop (HITL) financial query agent using LlamaIndex workflows and the FastAPI WebSocket architecture. The system intelligently parses natural language queries to extract financial parameters, dynamically requests missing information through interactive clarification steps, and retrieves precise financial data from structured databases. This agent enhances user experience by guiding users through complex multi-parameter queries while maintaining real-time communication and ensuring data accuracy through human validation checkpoints.
- Implemented an SQL agent to automate database queries and streamline data retrieval processes. This agent is designed to handle dynamic query generation, allowing for efficient interaction with complex databases. By automating these interactions, the SQL agent provides timely insights and facilitates better data management, enhancing overall operational efficiency.
- Implemented advanced document summarization techniques to extract key insights from large texts. This system leverages artificial intelligence algorithms to condense long documents into concise summaries, enhancing the accessibility and usability of information. By automating the summerization process, users can quickly grasp the main points of complex documents without the need for extensive reading.
- Developed a question-answering system capable of extracting accurate responses from CSV datasets. This system allows users to query large CSV files and receive precise answers based on the data contained within. By enabling efficient information retrieval, the solution helps users quickly find and use the data they need, improving decision-making processes.

Codiant - A YASH Technologies Company
Indore, Madhya Pradesh

Machine Learning Engineer
Jan. 2022 - April 2024

- Designed and integrated a conversation retrieval chat bot using OpenAI GPT-3.5 Turbo, LangChain, and Faiss vector database. Developed a personalized chat bot tailored to both customer and service provider needs.
- Leveraged pre-trained models to extract feature embedding from text, images, and audio, enabling the development of a highly personalized recommendation system.
- Designed, integrated, and deployed deep learning models to power recommendation systems, including Session-based, Knowledge-based, Collaborative Filtering, and Content-Based approaches. Additionally, implemented an Incremental learning system for continuous improvement with Kubeflow.
- Utilized prompt engineering techniques, including Chain of Thought, Zero Shot, and Few Shot prompting, to improve AI interpretation of datasets.
- Leveraging GPT-4 for profound conversational insights and improved team communication, while also utilizing GPT-2's text generation capabilities to create tailored job descriptions for enhanced hiring processes and attracting top-tier talent. Developed Document based Question Answering using GPT-3.5/RoBERTa to fetch aircraft details from bunch of documents.
- Developed and optimized AI-generated text prompts for various applications, focusing on accuracy and relevance.Collaborated with product development as a Generative AI subject matter expert, contributing to the creation of scalable, resilient, and ethical AI solutions.
- Applied prompt engineering best practices to reduce hallucinations and enhance response accuracy.

TechHelper Technologies
Bhopal, Madhya Pradesh,

Jr. Machine Learning Engineer
Nov. 2020 - Jan. 2022

- These projects offer a rich array of experiences. From creating a COVID-19 chatbot with Rasa for vital information dissemination to conducting sentiment analysis on airline data through a Streamlit app, they hone conversational AI and data analysis skills. Building a Flask-SQLite web app for URL shortening provides hands-on application development expertise. Crafting the K-Nearest Neighbor algorithm from scratch enhances machine learning proficiency. Lastly, participating in a FIFA 19 dataset hackathon for EDA and applying clustering techniques like K-Means and PCA deepens insights into unsupervised learning, making for a well-rounded skill set.
- have extensive experience in machine learning and data analysis, having implemented 10 different classification techniques for leaf classification, conducted heart attack analysis and prediction using SVM, Random Forest, and logistic regression, and built a simple linear model from scratch with gradient descent, and successfully predicted house prices using advanced regression techniques

Certificates

Neural Networks and Deep Learning
deeplearning.ai

Jun 2023 – March 2023

- Deep Learning Specialization

Improving Deep Neural Networks: Hyper-parameter Tuning, Regularization and Optimization   *deeplearning.ai*

Jun 2023 – Dec 2023
- Deep Learning Specialization

Data Science Professional Certificate   *IBM, Coursera*

Jun 2020 – Feb 2021
- IBM Data Science Professional Certificate-Coursera

Machine Learning   *Stanford University, Coursera*

Jun 2020 – Sept 2020
- Stanford University-Coursera -Andrew NG

Recommendation Systems on Google Cloud   *Google Cloud, Coursera*

Mar 2023 – Oct 2023
- Advanced Machine Learning on Google Cloud Specialization

Skills

Machine Learning	Transformers, LLM's, RAG (Retrieval augmented generation), Predictive Modeling, AWS, Microsoft Azure, GCP, Keras, TensorFlow, PyTorch, lightning.ai, RASA, XGBoost, CatBoost, LightGBM, Numpy, scikit-learn, Pandas, Pyplot, seaborn, Prompt-Engineering, Llama-index, Workflow, Langchain, LangGraph, Faiss, Chroma, Pinecone, Vector Database, Milvus, Generative-AI, Huggingface
MlOps	MLflow, Docker, Weights & Biases, Kubeflow, DVC
Databases	PostgreSQL, MySQL, MongoDB, Oracle Database, Redis, SQLite, Elasticsearch, Faiss, Pinecone, Weaviate, Milvus, Chroma
Back-end	Flask, FastAPI, Django, Sanic, Streamlit
Front-end	React, HTML5, LESS, SASS
Programming	Python, C++, SQL, LaTeX
Languages	English, Marathi, Hindi

Education

Shivaji University   *Kolhapur, Maharashtra*
Bachelor of Mechanical Engineering   *Mar. 2020*
- GPA: 2.9 / 4.0