Subhash Shivram Mokase

Contact No.: +91-9021214347

E-mail: subhashnmokase@gmail.com

Professional Summary

Experienced **Machine Learning Project Engineer** with **3.1+ years** in **Data Science, Machine Learning, and Generative AI**. Proficient in developing and deploying **ML/DL models**, implementing **NLP solutions**, and utilizing **Generative AI** to drive business insights. Skilled in end-to-end **data processing, model training, deployment, and optimization**.

Work Experience

CalSoft Pvt. Ltd., Pune (February 2022 – Present)

Designation: Machine Learning Project Engineer

Key Responsibilities:

- Developed and deployed ML and deep learning models for predictive analytics and business solutions.
- Implemented classification algorithms to build high-performance predictive models.
- Performed Exploratory Data Analysis (EDA), feature engineering, scaling, binning, and encoding techniques.
- Worked on fine-tuning Generative AI models like GPT, BERT, and LLMs for real-world applications.
- Built end-to-end machine learning pipelines, from data ingestion to model deployment.
- Integrated AI/ML solutions with cloud platforms like AWS.

Technical Skills

Machine Learning & AI:

- **Algorithms**: Regression, Classification, Clustering, Ensemble Methods (Bagging & Boosting), Supervised & Unsupervised algorithm, Reinforcement Learning.
- **Techniques**: EDA, Feature Engineering, PCA, Feature Scaling, Encoding, Data Normalization, Data visualization, Predictive modelling.
- **ML Pipelines**: Model Development, Deployment, Monitoring, Hyperparameter Tuning

Natural Language Processing (NLP):

- **Text Processing**: Tokenization, Lemmatization, Vectorization, Nlp Pipeline.
- **Libraries**: NLTK, spaCy, TextBlob, Hugging Face, Transformers.

Deep Learning:

- **Neural Networks**: Perceptron, ANN, CNN, RNN, LSTM, GRU.
- **Frameworks**: TensorFlow, Keras, PyTorch, OpenCV.

Generative AI:

- **Generative AI**: Hugging Face, OpenAI, GPT, BERT, LLM, LangChain, LangGraph, RAG, AI Agents, LoRA, QLoRA, Prompt Engineering, Transformer Architecture.

Databases & Cloud Technologies:

- **Databases**: MySQL, MongoDB, ChromaDB, Pinecone.
- **Cloud**: AWS (EC2, S3, SageMaker, Deployment).

Development Tools & Soft Skills:

- **Programming**: Python, SQL, Flask, FastAPI, Git, GitHub, Jira.
- **Data Tools**: NumPy, Pandas, SciPy, Scikit-learn, Seaborn, Matplotlib, Excel, OCR (Tesseract), Docker, PyMuPDF, Unstructured Library.
- **Operating Systems**: Linux, Windows, Ubuntu, Fedora.
- **MATHS AND STATS SKILLS:** Hypothesis Testing, Gradient Descent, SAS, Statistical modeling, Statistical tests.
- **IDE**: Jupyter Notebook, VS Code**,** PyCharm, Spyder, Google Colab, Power BI.

- **Soft Skills**: Problem-Solving, Teamwork, Communication, Adaptability, Honest, Collaboration, Interpretation.

Certifications:

- Python & Data Science, Red Hat (RHCSA Linux/RHCE) & Generative AI.

Education

- **M.Tech (CSE)** | **CGPA**: 9.0 (Pursuing)
- **B.E (CSE)** | **Dr. BAMU University** | **73%** | First Class with Distinction
- **Diploma (CSE)** | **MSBTE** | **75%** | First Class with Distinction

---

Project Experience

1. Diabetes Prediction (Predicting whether a patient has diabetes or not using ML models)

Roles & Responsibilities:

- Performed EDA, data cleaning, feature engineering, and model training.
- Implemented classification algorithms with high prediction accuracy.
- Optimized models using hyperparameter tuning and performance metrics.

2. Medical Insurance Price Prediction (Predicting charges for medical insurance plans)

Roles & Responsibilities:

- Developed ML models to estimate medical insurance charges.
- Built supervised learning models for charge prediction.
- Applied scaling, binning, and encoding techniques for improved model performance.

3. Automated Complaint Routing Using NLP & BERT Fine-Tuning (Classifying and routing customer complaints efficiently in financial institutions)

Roles & Responsibilities:

- Preprocessed large-scale text data, handling missing values and text cleaning.
- Applied BERT fine-tuning and LSTM models for classification.
- Implemented tokenization, word embeddings, and attention mechanisms.
- Integrated the trained model into a complaint routing system.
- Evaluated model performance using F1-score, precision, recall, and confusion matrix.

4. Generative AI Project: Q&A on PDFs using RAG (Retrieval-Augmented Generation)

This project builds a Q&A system for PDFs using Retrieval-Augmented Generation (RAG)**.** It finds relevant information from unstructured documents and provides accurate answers using LLMs (GPT, BERT, LLaMA)**.** Vector databases (ChromaDB, Pinecone, FAISS) improve search and retrieval for better responses.

Roles & Responsibilities:

- Develop and fine-tune LLM**s** for document understanding and Q&A generation.
- Implement RAG pipeline for efficient retrieval and answer synthesis.
- Preprocess PDF documents (OCR, text extraction, chunking, embedding generation)**.**
- Implement vectorization techniques for improved document retrieval.

5. Generative AI Project : Minutes of Meeting (MoM) Build an AI system to convert audio recordings into structured text using speech-to-text models. Process and analyze transcripts using LLMs & NLP techniques.

Roles & Responsibilities:

- Develop and fine-tune speech-to-text models (Whisper, Wav2Vec2) for accurate transcription & Preprocess and clean audio data for better transcription quality.
- Integrate Unstructured Library to handle various document and audio formats.
- Optimize LLMs (GPT, BERT, LLaMA**)** for text summarization and contextual understanding.
- Implement RAG (Retrieval-Augmented Generation) to enhance answer accuracy.
- Optimize vector search using ChromaDB, Pinecone, FAISS for efficient retrieval.