Harini Mukta

Email: muktaharini@gmail.com| **Phone:** (913) 596-6706 | **Address:** South Lyon, Michigan 48178
LinkedIn: www.linkedin.com/in/harinimukta

Summary:

- Experienced **Generative AI** and **Data Science professional** with expertise in developing advanced, data-driven solutions that enhance client engagement and operational efficiency.
- Skilled in **predictive analytics** and **unsupervised learning**, translating complex data into actionable insights.
- Experienced in deploying scalable **AI solutions** on **cloud platforms** and using state-of-the-art frameworks, backed by a master's in computer science focused on Data Science.

Technical Skills:

- **Data Analytics and Visualization**: Power BI, Statistical Analysis
- **Data Science:** Artificial Intelligence, Machine Learning
- **Database:** MySQL, MongoDB
- **Programming Languages:** Python
- **Cloud:** AWS SageMaker, AWS EC2, AWS Bedrock
- **Web Technologies:** HTML, CSS, JavaScript, Streamlit
- **Tools:** ECS Delta, Trusted Link EDI, SSIS, SSRS, Scikit
- **IDEs:** VS Code, Jupyter notebook, Google Colab
- **Libraries & Frameworks:** PySpark, NumPy, Pandas, Matplotlib, Scikit Learn, Pytorch, Langchain, Keras
- **Generative AI:** Transformers, GPT, DALL-E, Langchain, Fine Tuning, Prompt Engineering, RAG, Whisper API Hybrid Search, Gemma, Llama
- **Vector Database:** Pinecone, ChromaDB, FAISS
- **AI Platforms:** OpenAI, Hugging Face, GROQ, Ollama

Work Experience:

Client: Elysium Analytics, Santa Clara, CA.

Role: Generative AI Engineer **[July 2024-Current]**

Project Description: To develop an AI powered chatbot capable of recognizing user emotions

Responsibilities:

- Spearheaded the development of an **emotion-detecting chatbot, leveraging Streamlit for seamless User Interaction**
- **Incorporated AI and Natural language processing** through **finetuning of LlaMA2 models for advanced emotional intelligence.**
- **Integrated OpenAI's Whisper** technology for accurate voice-to-text capabilities, enhancing real time user communication
- Focused on **User-Centric design for the chatbot interface**, ensuring adaptability and scalability to various user's inputs and scenarios

- **Implemented LangChain's Model I/O module** to efficiently interface with language models, enabling more effective **few shot prompt templates and token usage tracking**
- **Utilized LangChain's Retrieval module** to interface with application-specific data, enhancing the chatbot's ability to access relevant information
- **Applied LangChain's Callbacks feature** to log and stream the intermediate steps of any chain, improving transparency and traceability of the chatbot's decision-making process.

Client: Hallmark Cards| INFOSYS PVT LTD

Role: **Data Scientist** **[Sep 2021 – Dec 2022]**

Project Description: Focused on **sales forecasting** and enhancing **supply chain efficiency** through **advanced data integration techniques** and the implementation of **machine learning models** for predictive analytics.

Responsibilities:

- Deployed **machine learning models** using **Amazon SageMaker** for accurate sales predictions.
- Utilized **supervised learning algorithms** (**Decision Trees**, **Random Forest**, and **Gradient Boosting**) for forecasting sales and enhancing decision-making processes.
- Conducted **cross-validation** and **hyperparameter tuning** to improve model accuracy, reducing error rates.
- Optimized **SQL Server databases** to enhance data quality and ensure efficient data integration, while using **PySpark** for data cleaning, normalization, and data transformation, streamlining data
- Implemented **Trusted Link EDI** and **ECS Delta** for improved supply chain efficiency and sales analysis.
- Played a key role in **data cleaning** and **preprocessing** including **handling missing values** through imputation techniques, **removing duplicates**, and performing **outlier detection** using statistical methods. Employed **feature engineering** to create relevant variables and utilized **data normalization** and **scaling** techniques to ensure high-quality datasets for machine learning models.
- Contributed to sales decision-making by integrating **advanced reporting systems** and **data-driven insights** into business processes, utilizing **Power BI** for **data visualization** and **interactive dashboards**. Employed **DAX** for calculations and **ETL processes** for seamless data integration, enhancing **business intelligence analysis**.

Client: CMR INFOTECHS PVT.LTD

Role: Data Analyst Intern

Project Description: In my role as a **Data Analyst Intern**, I supported the company's IT consulting services, focusing on helping organizations optimize their use of technology.

Responsibilities:

- I gathered and **analyzed data** from various sources to identify trends that could guide strategic IT decisions for clients. I performed **statistical analysis**, including **regression**, to validate insights and help develop tailored technology solutions.
- Utilizing **Python** libraries like **Pandas** and **NumPy**, I cleaned and manipulated data to ensure accuracy, which was crucial for effective project outcomes. Created interactive dashboards and reports using **Power BI**, making it easy to present key metrics and insights to both stakeholders and clients.

Projects:

- **Automated Client Data Extraction and Email Generation for Companies** | *GenAI* Aug 2024
  This project develops a tool to automate personalized email pitches for service-based companies. By leveraging web scraping, **Large Language Models (LLMs)**, and **vector databases**, it enhances the efficiency of Business

Development Executives (BDEs) by reducing manual effort and speeding up project opportunity identification. Key results include over 50%-time savings, a 30% increase in potential project opportunities, and a 25% boost in project win rates.

- **RAG Document Q&A |** *GenAI* June 2024
  Created an interactive Q&A interface with **Streamlit** using **Retrieval-Augmented Generation (RAG)** to process research papers. It employs **PyPDFDirectoryLoader** for document ingestion, **RecursiveCharacterTextSplitter** for chunking text, and **OpenAIEmbeddings** with **FAISS** for efficient vector-based document retrieval. Users can query the system, receiving accurate responses generated by **ChatGroq** and **Llama3**, enhancing information retrieval and engagement.
- **Classifying Amazon Book Review Sentiments|** *Artificial Intelligence* Mar 2024
  Conducted sentiment analysis on Amazon book reviews to classify text into **positive and negative categories**. This project employed **Natural Language Processing (NLP)** techniques, utilizing **TF-IDF vectorization** for feature extraction and **Random Forest classification** to achieve an impressive **accuracy of 0.98**. Key performance metrics included **F1 score**, **precision**, and **recall**, demonstrating the model's robustness in accurately predicting sentiment.
- **Classifying Geographic Landscapes in the Intel Image Dataset|** *Artificial Intelligent* Dec 2023
  This project utilizes **Convolutional Neural Networks (CNNs)** for precise image classification across varied datasets. Key steps include **feature extraction**, **data preprocessing** (resizing, normalization, augmentation), and **model optimization** techniques like dropout and batch normalization to enhance accuracy and prevent overfitting. Model performance is assessed for detailed error analysis where delivered with **0.82 Test Accuracy**. This streamlined approach ensures high-performance image classification.
- **Tailoring Strategies through Client Attribute Analysis|** *Artificial Intelligence* Feb 2023
  Implemented **unsupervised learning techniques**, specifically **K-means clustering**, to segment clients based on a multitude of personal attributes. This approach facilitated the development and implementation of tailored strategies customized to individual preferences and characteristics. Utilized **evaluation metrics** such as the **Silhouette Score**, and **Within-Cluster Sum of Squares** to assess the quality and effectiveness of the clustering results
- **Insights from Diverse Datasets via Predictive Modeling and Classification** |*Machine Learning* Sept 2022
  Conducted a comprehensive analysis of insurance datasets utilizing various **classification techniques**, including **Random Forest, Logistic Regression, Decision Trees, and Gradient Boosting Classifier**. Among these, the **Random Forest model** demonstrated superior performance, achieving an **accuracy** of **0.9345**. This analysis involved the integration of key algorithms and advanced statistical models to enhance predictive accuracy and ensure robust classification outcomes.

Education:

Master of Science - Computer Science **-** Data Science **May 2024**

University of Missouri Kansas City **GPA:3.8/4**

Bachelor of Technology - Computer Science and Engineering **Jun2021**

Jawaharlal Nehru Technology University Hyderabad **GPA:8.5/10**

Certifications:

- **Academy Accreditation-Generative AI Fundamentals, Data Bricks**
- **Complete Generative AI Course With Langchain and Hugging Face, Udemy**
- **Foundational Generative AI, iNeuron**