Dinesh Vennapoosa

+1 (317) 720 7226 – Portfolio – dinesh.vennapoosa@gmail.com – LinkedIn – Github

SUMMARY

Experienced **Machine Learning Engineer** with a strong background in designing, developing and deploying advanced **ML models**. Proficient in implementing **MLOps** practices and leveraging cloud services like **AWS** for scalable and reliable machine learning solutions. Proven expertise in **generative AI**, **NLP**, and **computer vision**, with hands-on experience fine-tuning large language models and building robust **information retrieval systems**. Efficient at collaborating with cross-functional teams to integrate **AI-driven insights** into business and healthcare applications, driving efficiency and innovation. Passionate about utilizing **AI** to solve complex problems and enhance decision-making processes.

WORK EXPERIENCE

Generative AI Engineer

Dell Technologies, Texas, USA *February 2024 - Present*

- Designed, developed, and deployed **Generative AI** powered chatbots using **Retrieval-Augmented Generation (RAG)** with **Python**, enhancing conversational capabilities and improving information retrieval efficiency by 40%.
- Implemented an intent detection model that classified user utterances with 95% accuracy, streamlining interaction classification.
- Leveraged **vector databases** like **Faiss** to index fallout resolutions using Python, creating a comprehensive knowledge base from historical technical support and business user data.
- Applied **prompt engineering** techniques using **LangChain** and **Python** to connect **large language models (LLMs OpenAI GPT model** with data from Oracle, PostgreSQL, optimizing data integration.
- Evaluated retrieval efficiency and summarization quality at different stages of RAG pipeline and used OpenAI GPT model to generate accurate responses to the queries.
- Enhanced project management by integrating **Git** for version control and utilizing **CI/CD** tools like **Jenkins**, ensuring codebase integrity and streamlined development workflows and followed software development principles, design patterns, and best practices.
- Deployed the chatbot using **Flask** and **AWS** services with **Docker** and **Kubernetes**, ensuring scalable and robust implementation through efficient containerization.
- Collaborated with **cross-functional teams** including UX/UI designers and front-end developers to enhance the chatbot's user experience.

Machine Learning Engineer

Indiana University, Indiana, USA *August 2022 - December 2023*

- Led a research project on using **large language models (LLMs)** like GPT-3.5 in cultural studies to rank long-tail cultural concepts such as holidays across different countries.
- Developed advanced NLP techniques and generative AI models to create a novel few-shot question-answering task, utilizing GPT-3.5, Bloom 7b, and Llama 7b.
- Fine-tuned **Llama 7b LLM** using the **QLoRA** approach on a custom dataset, leveraging transfer learning and distributed training on NVIDIA A100 GPUs, enhancing cultural nuance recognition by 15%.
- Created a robust **RAG** application using LangChain for efficient Information Retrieval Systems, integrating GPT-4 Turbo OpenAI API and knowledge graphs like Neo4j to generate accurate clinical responses, reducing manual intervention by 60%.
- Collaborated with healthcare professionals to refine the RAG system, incorporating feedback to improve model performance, interpretability, and scalability.
- Optimized NLP tasks using Hugging Face Transformer models such as **BERT**, improving model performance through rigorous optimization techniques, hyperparameter tuning, and metric evaluations.

Machine Learning Engineer

Ugaan Labs Pvt Ltd, Karnataka, India *July 2020 - June 2022*

- Implemented **deep learning** and **computer vision** state-of-the-art object detection models using **YOLOv4 tiny, YOLOv7, YOLOv8** with TensorFlow, achieving a 10% increase in detection accuracy.
- Applied **OCR (Optical Character Recognition)** to extract text features from KYC documents, utilizing data annotation tools like Labelbox and data versioning with **DVC**, enhancing KYC automation and real-time processing efficiency.
- Enhanced system accuracy by 5% and facial recognition match accuracy by 8% through **model fine-tuning (transfer learning)** and automated **hyper-parameter tuning** frameworks, validated by **A/B testing**.

- Designed and executed complex SQL queries for **ETL processes** with tools like **Apache Airflow**, ensuring high-quality data for predictive modeling.
- Developed machine learning models with Python libraries such as **Pandas** and **scikit-learn**, including **SVM, Decision Tree, Random Forest, XGBoost**, optimized with **GridSearchCV**.
- Deployed solutions on **AWS**, utilizing **EC2, S3, SageMaker, Docker, and Kubernetes** for scalable implementation, with monitoring and logging via **AWS CloudWatch** and serverless functions using **AWS Lambda**, aligning with **MLOps** best practices using **MLflow**.

Data Analyst

Indian Institute of Technology Guwahati, Assam, India *July 2018 - May 2020*

- Developed **machine learning models** using **Decision Trees** and other algorithms, achieving a 95% accuracy rate in predicting equipment failures, reducing downtime and maintenance costs by 33%.
- Applied **signal processing** techniques combined with **machine learning** to analyze machinery vibration patterns, identifying potential faults and enabling timely maintenance actions using **Fourier Transform** for signal processing.
- Conducted **data cleaning** and **preprocessing** to ensure data quality and reliability, using techniques such as **normalization**, **outlier detection**, and **missing value imputation**.
- Utilized data visualization tools such as **Power BI** and **Matplotlib** to create insightful dashboards and reports, facilitating data-driven decision-making for maintenance and operations teams.
- Worked closely with **mechanical engineering** experts to integrate practical insights into **data-driven models**, ensuring technical soundness and real-world applicability.
- **Documented** model development processes, signal processing techniques, and maintenance recommendations to support knowledge transfer and scalability.

PUBLICATIONS

- By **artificial intelligence** algorithms and **machine learning models** to **diagnose cancer** Link
  - Conducted extensive research on the integration of **AI** in oncology diagnostics, focusing on imaging tests, endoscopic procedures, and biopsy and cytology tests to evaluate the effectiveness of **AI** in early cancer detection and prognosis.
  - Analyzed the proficiency of **AI systems** using advanced algorithms like **Convolutional Neural Networks (CNNs)**, demonstrating enhanced precision in diagnosing various cancer types and playing a crucial role in prognosis and treatment strategy determination.
  - Focused on algorithm-guided identification of oral cancer, utilizing **CNNs** to detect subtle variations in lesion heterogeneity, significantly improving early detection and treatment success rates.

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, Object-Oriented Programming (OOP), Data Structures, Problem Solving.
Machine Learning: Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Keras, TensorFlow, PyTorch, Deep Learning, NLP, Computer Vision.
Natural Language Processing: NLTK, SpaCy, Gensim, BERT, RoBERTa, Tokenization, Lemmatization, Stemming, Named Entity Recognition (NER), Part-of-Speech (POS) Tagging, Dependency Parsing, Sentiment Analysis, Text Classification, Language Modeling.
Generative AI: GPT-3.5, GPT-4 Turbo, Bloom 7b, Llama 7b, Large Language Models (LLMs), Prompt Engineering, Few-Shot Learning, Zero-Shot Learning, QLoRA, Retrieval-Augmented Generation (RAG), Vector Databases (Faiss), LangChain.
Cloud Computing & Big Data: AWS (EC2, S3, SageMaker, CloudWatch, Lambda), PostgreSQL, MongoDB, Spark, PySpark, Docker, Kubernetes.
Data Analysis: Numpy, Pandas, SciPy, Matplotlib, Seaborn, Plotly, Power BI, Statistical Analysis, Data Modeling, Feature Engineering, Data Cleaning, Data Wrangling.
Development and Deployment: Linux, Git, Continuous Integration/Continuous Deployment (CI/CD), Flask, Docker, Kubernetes, Model Deployment, ETL Pipelines, ML Pipelines, MLOps (MLflow, Kubeflow, TFServing).

EDUCATION

Indiana University Purdue University **Indianapolis, Indiana, USA**
Master's in Applied Data Science *Dec 2023*