MUTHU AJAY
Hyderabad | +91 8098060920 | muthuajay97@gmail.com | LinkedIn | GitHub

Highly accomplished and results-driven Data Scientist Consultant with 3+ years of experience at Ernst & Young, specializing in developing and deploying advanced AI/ML solutions. Proven ability to leverage cutting-edge technologies like CNNs, Vision Transformers, LLMs, and RAG models to solve complex business problems, improve efficiency, and drive data-driven decisions. Expertise in data engineering, model development, and cross-functional collaboration, consistently exceeding project goals and delivering impactful results.

WORK EXPERIENCE

DATA SCIENTIST - CONSULTANT | ERNST & YOUNG | SEPTEMBER 2023 - PRESENT

- **Enhanced Document Processing:** Developed and deployed CNN models in PyTorch for document classification, achieving **99% accuracy** and **reducing processing time by 60%**, significantly improving operational efficiency.
- **Improved PII Image Classification:** Implemented transfer learning with EfficientNet, achieving a **15% performance improvement** in PII image classification accuracy.
- **Contract Analysis Tool: Built using RAG, Chroma DB, and LLMs to automate legal contract review and improve team efficiency.**
- **Optimized SQL Query Performance with RAG:** Integrated LLM-powered Retrieval-Augmented Generation (RAG) models for SQL query optimization, resulting in a **30% enhancement in query performance**.
- **Data Curation for Visual Recognition:** Managed the collection and pre-processing of large, complex datasets of distorted alphanumeric characters from images, ensuring high data quality for model training.
- **Advanced Alphanumeric Character Classification:** Designed and implemented CNN models to classify complex, distorted alphanumeric characters, achieving **97% accuracy** on challenging visual recognition tasks.
- **State-of-the-Art Image Classification with Vision Transformers:** Designed and fine-tuned Vision Transformer (ViT) models for image classification, achieving state-of-the-art performance and improved robustness against adversarial attacks.
- **Multimodal AI System Development:** Implemented Vision Transformers within multimodal AI systems, integrating visual data with NLP for enhanced model understanding and decision-making capabilities.
- Developed and deployed a robust and efficient **XGBoost model** incorporating **dimensionality reduction** to manage **high-dimensional and high-cardinality features.** Achieved **94.9%** average cross-validation accuracy**,** ensuring robust performance, minimizing overfitting, and optimizing model efficiency with reduced computational costs.

DATA SCIENTIST - ASSOCIATE CONSULTANT | ERNST & YOUNG | JANUARY 2022 – AUGUST 2023

- **End-to-End Data Scanning Solutions:** Led the development of end-to-end data scanning ad hoc solutions for diverse databases (MySQL, Hive, Redis, MongoDB, Elasticsearch, custom databases) and unstructured data sources (HDFS, cloud storage).
- **Automated ETL for PII Identification:** Implemented an automated ETL pipeline based on microservice architecture using private cloud (Airflow, Docker, Kubernetes, DataBricks) for PII identification across multiple data sources, ensuring compliance with global data privacy regulations.

KEY PROJECTS

- **Contract Analysis Application with RAG:** Developed a sophisticated Contract Analysis application leveraging **Retrieval-Augmented Generation (RAG)**, **Chroma DB vector store**, and **Large Language Models (LLMs)**. This application automates the analysis of legal contracts, extracts key information for the legal team, **significantly reducing manual review time and improving legal team efficiency**.

- **CNN and VGG16 Architectures for Image Recognition:** Built CNN and VGG16 architectures, achieving high accuracy on benchmark image datasets, demonstrating a deep understanding of convolutional neural networks.
- **GPT-2 Language Model Pre-training:** Developed and pre-trained a GPT-2 architecture from scratch on a 10GB custom dataset, showcasing expertise in transformer-based language models and large-scale training.
- **Segment Anything Model (SAM) Implementation and Optimization:** Designed and implemented the Segment Anything Model (SAM) architecture, including custom dataset preparation, automated training pipelines, and extensive hyperparameter tuning to achieve optimized segmentation performance. Leveraged transfer learning and fine-tuning to adapt SAM to specific domains with limited data.

SKILLS

- **Programming:** Python
- **Machine Learning & Deep Learning:** PyTorch, Deep Learning, Transformers, LLMs, Gen AI, CNN, Vision Transformers, NLP, Reinforcement Learning, Classification, Regression, Statistical Modelling, AB Testing, Model Deployment
- **Data Analytics & Engineering:** Data Cleaning, Feature Engineering, EDA, Pandas, NumPy, Data Visualization, Time Series Analysis, ETL, Data Pipelines
- **Computer Vision:** Image Processing, Object Detection, Image Segmentation
- **Databases:** SQL, Hive/Hadoop, NoSQL
- **Visualization:** Power BI, Tableau, Matplotlib, Seaborn
- **Cloud Platforms:** Azure Services, AWS, Google Cloud Platform (GCP)
- **Version Control:** Git, GitHub

EDUCATION

- **PGDM – BUSINESS ANALYTICS | MAY 2021 | INTERNATIONAL SCHOOL OF MANAGEMENT EXCELLENCE, BENGALURU**
- **BACHELORS OF ENGINEERING – ELECTRONICS AND COMMUNICATION ENGINEERING | MAY 2018 | SAVEETHA ENGINEERING COLLEGE, CHENNAI**