Teja Y D

LinkedIn| teja.yerramsetti2@gmail.com | (+1) 812-671-5428

Summary

Data Science and AI professional with 8+ years of experience in machine learning, AI-driven risk assessment, and financial analytics across banking, taxation, and credit risk domains. Expertise in Generative AI, NLP, Graph Neural Networks (GNNs), and MLOps to build scalable AI-driven solutions. Proven ability to design fraud detection models, real-time credit risk assessment systems, and tax compliance automation leveraging Azure OpenAI, Spark Streaming, Kafka, and Hadoop. Strong background in big data processing, predictive modeling, and AI-driven automation to enhance operational efficiency and decision-making in finance and regulatory sectors.

Technical Skills

Languages: Python, R, SQL, Scala, Javascript, HTML
Machine Learning: Supervised (Linear regression, Decision Trees, SVM, KNN, Naive Bayes, Ensemble methods, XGBoost, Random Forest), Unsupervised (K-Means, PCA, SVD, LDA)
Deep learning and NLP: Pytorch, TensorFlow, Tokenization, Encoding, Named Entity Recognition, Sentiment Analysis, Topic Modeling, Recurrent Neural Networks (RNNs), LSTMs, Convolution Neural Network(CNNs)
Generative AI Frameworks : Langchain, LLMs(GPT, BERT, LLama, MistralAI, Falcon etc.,), Prompt Engineering, RAG(Retrieval-Augmented Generation), Fine Tuning(LoRA, QLoRA)
Distributed Computing: PySpark, Hadoop, Databricks, Hive, ETL/ELT, Snowflake
Cloud: Azure (OpenAI, AI studio, AI search, ADF), AWS(Sagemaker, Bedrock, EC2, EMR, Lambda), GCP
Other Skills: Git, MLFlow, DagsHub, MLOps, Jenkins, Docker, Kubernetes, Agile methodology, Jira, Tableau, powerBI

Professional Experience

OSU Wexner Medical Center | Senior Data Scientist — *Gen AI & MLOps | Azure OpenAI, LangChain, Azure Cognitive Search, Spark, Azure ML*
Aug 2023 – Present

- Developed an AI-powered RAG (Retrieval-Augmented Generation) clinical assistant to help physicians interactively navigate extensive surgical protocol documents. Leveraged Azure OpenAI and LangChain to create a chat-based interface that extracts precise SOPs and workflows from a vectorized knowledge base, reducing manual search time by 60% and improving compliance with procedural accuracy.
- Engineered document ingestion and indexing pipelines to transform unstructured clinical guidelines into vector embeddings using Azure Cognitive Search and custom chunking strategies, enabling fast semantic search and real-time answer generation in natural language.
- Built and deployed an agentic AI interface for HIPAA compliance enforcement, monitoring employee training status using automated data pipelines. Integrated Azure Data Factory, Spark, and SQL-based validation logic to dynamically track completion of mandatory HIPAA courses.
- Automated access control workflows using event-driven triggers and AI agents that grant or restrict access to internal healthcare systems based on real-time compliance status, ensuring adherence to federal guidelines and reducing manual oversight.
- Implemented robust MLOps practices using Azure ML and Kubernetes, automating model lifecycle, prompt template updates, and chatbot versioning, resulting in 50% faster iteration cycles and consistent performance in clinical environments.

Income Tax Department, India | **Data Scientist/Analyst** |*Machine learning, EDA, Tableau, MLflow, Tableau* April 2017 - August 2022

- Developed an AI-powered taxpayer risk assessment system leveraging Graph Neural Networks (Neo4j) and XGBoost, identifying high-risk entities and leading to the detection of ₹800 Cr ($100M) in tax discrepancies.
- Built an anomaly detection model using ElasticSearch and Spark, linking tax filings, bank transactions, and GST records to flag fraudulent entities, improving fraud identification accuracy by 35%.
- Engineered an NLP-driven document processing pipeline with Tesseract OCR and spaCy, automating extraction and validation of tax filings, reducing manual review efforts by 60%.
- Implemented a large-scale data pipeline using Hadoop and Spark, optimizing ETL workflows for ingesting millions of financial transactions daily, reducing processing time by 50%.
- Designed interactive compliance dashboards in Tableau, providing real-time tax collection insights and aiding

audit teams, improving policy enforcement efficiency by 40%.
- Deployed an entity-resolution system using fuzzy matching algorithms, linking shell companies and undisclosed income sources, strengthening tax compliance monitoring and reducing investigation time by 30%.

Wells Fargo, India | **Data Analyst** |*Machine learning, EDA, Tableau, MLflow, Tableau* July 2015 - March 2017

- Developed SQL-based financial risk models to analyze customer credit behavior, improving early fraud detection and reducing delinquency rates by 12%.
- Built interactive dashboards in Tableau to track key banking metrics, enabling leadership to make data-driven decisions and improving operational efficiency by 20%.
- Designed automated ETL workflows using SSIS and Hadoop, streamlining data ingestion from multiple banking systems and reducing processing time by 40%.
- Conducted advanced data analysis using SAS to identify anomalies in loan approval trends, enhancing fraud detection capabilities for mortgage applications.
- Optimized financial reporting processes in Excel and SQL, automating data reconciliation and reducing manual workload for analysts by 35%.

Projects

IntelliQuery: AI-Powered Research Assistant | *Claude, Azure OpenAI, LangChain, FAISS, Bedrock pipelines.*

- Developed a custom conversational model leveraging RAG frameworks to summarize scientific articles and extract web-based information, enhancing answer relevance by 30% through advanced prompt engineering.
- Deployed scalable CI/CD pipelines with AWS Bedrock, reducing researchers' time-to-insight by 50% and enabling faster decision-making.

Knowledge Extraction and Q&A System | *Hugging Face Transformers, LangChain, Pinecone, FAISS, AWS SageMaker.*

- Fine-tuned advanced LLMs like Claude and LLaMA2, integrated with RAG using FAISS, to extract insights from technical documentation web sites, achieving 90% accuracy in real-time question-answering.
- Implemented and scaled the solution on AWS SageMaker, achieving a 35% improvement in response times and a 40% reduction in manual data review, significantly enhancing research efficiency and workflow optimization.

Stance Detection | *Python, HTML, API, NLP, Machine learning, web scraping*

- Implemented advanced techniques, including contextualized pre-training and various machine-learning algorithms.
- Achieved optimal performance with 94.77% training accuracy and an 86.23% testing accuracy through the ensemble method for text classification.

Route Planner | *Python, statistics, Git, Machine learning, Artificial Intelligence*

- Developed an algorithm to find optimal routes across the extensive U.S. Road network, similar to Google Maps
- Utilized A* search and structured data into a graph for efficient routing, even considering scenarios like package delivery constraints on high-speed roads

Home Credit Default Risk (HCDR) | *Python, Hypothesis testing, A/B testing, EDA, data visualization, Docker, ML*

- Designed and implemented advanced machine learning models, achieving a 91.95% accuracy in predictive loan risk assessments.
- Successfully applied various machine learning models, with Logistic Regression leading in performance

Education

Indiana University-Bloomington Aug 2022 –May 2024
Master of Science in Data Sciences *GPA: 3.95/4.0*

IIIT– Nuzvid July 2011-June 2015
Bachelor of Technology *GPA: 9.01/10.0*