Abhishek Teja Goli

+19374437879 | g.abhishekteja@gmail.com | Dayton, OH, USA | Linkedin

WORK EXPERIENCE

Topcon Healthcare   Remote
Data Science Intern   *June 2025 – Present*

- Automated Excel report generation by developing Python scripts that parsed system log files, reducing manual data entry time by 80%.
- Fine-tuned the Automorph deep learning pipeline on Azure Databricks to improve retinal image segmentation accuracy for NW400 images by 30%.
- Built scalable data workflows in Azure Databricks, leveraging distributed computing to accelerate image model training by 2x.
- Co-developed an intern social app to streamline daily check-ins and event tracking, enhancing intern engagement across 20+ team members.

Research Assistant, KASTLE lab

Wright State University   *March 2025 – Present*

- Conducting applied research under Dr. Cogan Shimizu, focusing on knowledge graphs, LLMs, and semantic technologies.

DataBeat   Telangana, India
Data Analyst Intern   *Feb 2024 – July 2024*

- Engineered deep learning models using ResNet, LSTM, and RNN to improve crop disease classification accuracy to 94%.
- Redesigned Python-Django data ingestion pipelines, cutting processing time by 40%. and implementing best practices for data warehouse management.
- Streamlined ETL workflows using Azure Data Factory and Python, resulting in 50% faster data processing and enabling real-time analytics for client dashboards.
- Built a predictive analytics model for ad forecasting (Nexstar Media Group), reducing manual reporting by 70% and improving forecast accuracy by 15%.

PROJECTS

Organization-Specific LLM Application | Link

- Architected an enterprise-grade Retrieval-Augmented Generation (RAG) system enabling creation of custom LLMs using proprietary data, increasing information retrieval accuracy by 85%.
- Engineered a scalable vector database with FAISS and SQLite3, integrating IndexIVFFlat indexing for sub-500ms query responses on 10GB+ datasets. Implemented Llama 3.2 (3B parameters) with plans to upgrade to Llama 3.3 (70B) for enhanced performance.
- Developed a Python-based RESTful API processing multiple files formats and transforms, optimizing embedding efficiency by 40%.

Smart Movie Recommendation System | Link

- Engineered an emotion-based recommendation engine using Natural Language Processing and deep learning (LSTM, RNN) analyzing user sentiment to deliver personalized movie suggestions with 95% accuracy.
- Integrated hybrid filtering algorithms (collaborative and content-based), improving recommendation relevance by 35%, scaling to 100+ concurrent users using load balancing and database sharing.
- Enhanced precision-recall metrics through A/B testing, improving model accuracy by 20%.

Mental Health Severity Classification System | Link

- Developed an ML model for mental health severity classification, achieving 92% accuracy.
- Evaluated Decision Trees, Naive Bayes, and KNN, optimizing F1-score by 15%.

EDUCATION

- **Wright State University**   Dayton, OH
  Master of Science in Data Science | GPA: 3.6/4.0   *Aug 2024 – Present*
- **Mahatma Gandhi Institute of Technology**   Telangana, India
  B. Tech in Computer Science and Engineering (Data Science) | GPA: 3.06/4.0   *Aug 2020 – June 2024*

TECHNICAL SKILLS

Programming & Database	: Python, SQL, Postgres
Libraries & Frameworks	: Pandas, numpy, Flask, Streamlit, Tensorflow, Pytorch
Data Science and Analytics	: ETL Pipelines, Data Cleaning, Data Wrangling, Statistical Modeling, A/B Testing, Hypothesis Testing
Data Visualization	: Matplotlib, Seaborn, Plotly, Tableau, Power BI
ML Frameworks:	: Regression, Classification, Clustering, Predictive Modeling, RAG, LLMs, NLP, Gen AI
Cloud and Tools	: Cloud platforms (AWS, GCP), Azure Databricks, Git, Jupyter Notebooks, VS Code, Excel

CERTIFICATIONS

- Certified in **Data Science** by IBM.
- **Google Data Analytics** from Google
- Certified by NPTEL in **Natural Language Processing** (NLP).
- Earned **Python** and **Machine Learning** certification from **Kaggle**.