SAI DHRUV YELLANKI HANMANTHRAO

College Park, MD| +1 (240)-960-1300 | dhruvsai@umd.edu |LinkedIn | Github

PROFESSIONAL SUMMARY

Data Science professional with experience in building real-time ETL pipelines and KPI dashboards using Azure Databricks, SQL, and Power BI for financial analytics. Developed machine learning models with PyTorch, XGBoost, and Prophet for time series forecasting and churn prediction. Implemented Retrieval-Augmented Generation (RAG) systems using Transformers and LangChain for multi-document NLP tasks such as sentiment analysis and document summarization.

EDUCATION

UNIVERSITY OF MARYLAND, COLLEGE PARK — College Park, MD, USA
Master of Science in Data Science; Cumulative GPA: 3.89/4.0 — Aug 2024 - May 2026

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY — Chennai, TN, India
Master of Science in Data Science; Cumulative GPA: 3.74/4.0 — Jun 2019- May 2023

- **Publication**: Y H Sai Dhruv, (2023) Speech Emotion Recognition Using LSTM Model (A Deep Learning Approach). International Conference on Recent Trends in Data Science and its Applications (ICRTDA 2023) | DOI | Github Link

WORK EXPERIENCE

WATZMANN CONSULTING PVT LTD — Hyderabad, TG, India
Data Analyst — Jun 2023 – Jun 2024

- Built automated ETL pipelines using Azure Data Lake for scalable storage and Azure Databricks for high-speed distributed processing of procurement and finance data, reducing processing delays and improving data readiness by 16% for real-time reporting.
- Applied SQL-based transformations in Databricks to clean, standardize, and integrate raw datasets (invoices, purchase orders, payment logs), enhancing data reliability and achieving a 15% increase in data accuracy and 20% reduction in manual errors.
- Designed 12+ interactive Power BI dashboards to visualize KPIs such as inventory turnover, delivery performance, and supplier efficiency, improving operational visibility and reducing report generation time by 30%.
- Streamlined reporting workflows by integrating auto-refresh schedules into Power BI and delivering features through Agile sprints, reducing manual dashboard updates by 18% and increasing stakeholder decision-making speed by 25%.

RAMPHAL TECHNOLOGIES PVT LTD — Hyderabad, TG, India
Data Science Intern — Dec 2022 – May 2023

- Developed and fine-tuned flight delay prediction models using PyTorch, XGBoost, and clustering techniques, leading to a 12% improvement in simulation accuracy and a 7.5-minute reduction in projected delays.
- Architected a real-time pipeline using RabbitMQ and PostGIS to stream and process 10,000+ flight events per second, reducing false positives in congestion alerts and improving system throughput.
- Created Tableau dashboards to track flight metrics and congestion patterns, reducing manual reporting time by 30% and increasing data accessibility by 25%.

PROJECTS

LLM-BASED MULTI-PDF RETRIEVAL-AUGMENTED GENERATION (RAG) SYSTEM | Github Link — Mar 2025

- Engineered a real-time question-answering system that enables querying across up to 10 PDF documents. Implemented semantic search using Hugging Face embeddings and FAISS, and generated responses via LangChain integrated with Groq's LLaMA3 model. Deployed the solution with Streamlit, reducing document review time by over 40%.
  Tech Stack : Python, LangChain, Hugging Face, FAISS, Groq API, LLaMA3, PyPDF2, Streamlit.

BITCOIN PRICE TIME SERIES FORECASTING WITH FACEBOOK PROPHET | Github Link — Jan 2025

- Implemented and containerized a forecasting pipeline using Facebook Prophet on 13 years of Bitcoin data. Optimized changepoint and seasonality parameters, improving predictive accuracy by 8.3% (RMSE). Deployed an interactive Streamlit dashboard to deliver 365-day forecasts with 95% confidence intervals.
  Tech Stack: Python, Facebook Prophet, Pandas, Matplotlib, Scikit Learn, Plotly, Streamlit, Docker, Git, Web Scraping

CUSTOMER CHURN PREDICTION & RISK MODELING | Github Link — Nov 2024

- Developed and tuned Random Forest, SVM, and Logistic Regression models to classify customer churn using financial and engagement data. Engineered indicators like zero-balance flags and inactivity drop-offs, improving feature relevance and increasing model precision by 11%. Final model achieved 86.5% accuracy and 0.86 AUC, supporting retention efforts projected to reduce churn by up to 20%.
  Tech Stack: Python, Scikit-learn, Pandas, Matplotlib , Predictive Modelling, Cross-Validation, Hyperparameter Tuning.

SKILLS

Programming Languages: Python(Pandas, NumPy, Scikit-Learn, Matplotlib, Seaborn, Plotly, Dash, PySpark, SpaCy, NLTK, TensorFlow), SQL , R.
Databases: MySQL, PostgreSQL(PostGIS),Oracle, NoSQL(MongoDB).
Cloud Services: Microsoft Azure (Databricks, Data Lake), AWS.
Data Science & Methodologies: ETL Development, Data Cleaning and Transformation, Feature Engineering, Predictive Modeling, Time Series Forecasting, Deep Learning (LSTM, Transformers), NLP (NER, Sentiment Analysis, Embeddings), Model Evaluation (Accuracy, ROC-AUC), Agile Methodology
Tools & Technologies: Power BI, Tableau, Kibana, Apache Kafka, Apache Flink, RabbitMQ, Docker, Streamlit, Jupyter Notebook, Git, Visual Studio Code, JIRA