KRISHNA CHAITANYA BHUPATHI RAJU

kbhupath@buffalo.edu | https://www.linkedin.com/in/krishna-chaitanya-9b28aa238/ | +1 (716) 715-0795| Buffalo, NY

EDUCATION

Master of Science in Engineering Science -Data Science
University at Buffalo, The State University of New York
- Cumulative GPA: 3.5/4

Bachelor of Engineering: Computation and Mathematics
Mahindra University

SKILLS

- **Programming Languages**: Python, C, SQL
- **Machine Learning**: PostgreSQL, Databricks, Python packages (e.g., scikit-learn, numpy, pandas, matplotlib, PyTorch, plotly), AWS, Azure.
- **Data Science Tool & Technologies**: Data science pipeline (cleaning, wrangling, interpretation), Statistics, Excel, Git, A/B testing, VSCode, HTML

PROFESSIONAL EXPERIENCE

Data Analytics Intern, Infobos Technologies Pvt Ltd Feb 2024-August 2024
- Worked on a subcontracted initiative related to Google's Project Nimbus, focusing on enhancing AI-driven response systems for enterprise clients.
- Conducted data preprocessing, feature extraction, and pattern analysis on large-scale user interaction datasets.
- Used Python(pandas,Numpy,sckit-learn) and SQL to build pipelines and prepare model inputs

Product/Data Analyst, Kofuku Idea Labs May 2023-August 2023
- Performed data cleaning, preprocessing, and EDA using Python (Pandas, NumPy, Seaborn)
- Analyzed user behavior and product metrics to derive actionable business insights.
- Supported A/B testing and marketing analytics to optimize product features and campaigns.
- Built visual dashboards to present findings to product and marketing teams.
- Collaborated cross-functionally to align data insights with business strategy.

PROJECTS

Flight Delay Prediction and Analytics System: Tech: Apache,Pyspark, ML,MCP Engineering
- Built a distributed Spark data pipeline integrating aviation + weather + turbulence datasets to engineer large-scale delay risk features and perform end-to-end ETL and analysis.
- Developed and optimized **Spark MLlib models (Logistic Regression, Random Forest, GBT)** using Pipelines, feature engineering, and cross-validation to predict flight delay probability.
- Designed an **AI-queryable analytics system** by combining GraphFrames (airport network analysis) with an MCP server, enabling natural-language access to real-time delay insights.

Healthcare Admissions Database System: Tech:PostgreSQL, Python, dbt, SQL, Docker, Streamlit
- Designed and implemented an end-to-end healthcare data pipeline transforming raw patient admissions CSV data into validated, analytics-ready OLAP models using PostgreSQL, Python ETL, and dbt.
- Built a **production-grade data architecture** with a normalized 3NF OLTP schema and dbt-powered star schema (fact & dimension tables), enabling scalable clinical and operational analytics.
- Delivered **advanced SQL insights and visualization tools**, using CTEs and window functions to analyze appointment no-shows, and deployed an interactive Streamlit dashboard within a fully containerized Docker environment.

Cybersecurity Data Lakehouse: Tech: Databricks,pyspark, Deltalake
- Built a Medallion Architecture (Bronze/Silver) on Databricks to process 40k+ cybersecurity JSON records.
- Normalized nested arrays using Spark explode operations for granular vendor risk analysis.