# Kaushik Parvathaneni

📞 +91 9949057441  ✉ kaushikparvathaneni26@gmail.com  🔗 Linkedin  ⬡ Github  👤 Portfolio

## Work Experience

**Senior Data Engineer**                                                    **February 2025 – November 2025**
*CMR Solutions LLC*                                                                       *Richmond, VA*
- Designed and automated scalable ETL workflows using Apache Airflow with Python, integrating data from APIs and Kafka streams, and internal systems, which reduced manual processing time.
- Developed and built modular DBT models on Snowflake to standardize business logic and support fast and testable reporting across teams which improved data trust and transparency.
- Deployed a unified analytics stack on AWS(S3, Redshift, Lambda) consolidating siloed data into a single platform which increased performance of the dashboard and reporting by 35%.

**Senior Data Engineer (Senior Consultant)**                               **April 2024 – February 2025**
*Hoosier Community Network*                                                              *Bloomington, IN*
- Led data infrastructure projects for small businesses using Python, MySQL, and complex SQL queries to clean transform, and analyze large data sets of over 500,000 rows, providing actionable insights while aligning with industry standards to support business strategy and decision-making. Also Conducted workshops on data visualization, analytics, and emerging technologies, enhancing the data literacy of the students.
- Designed Tableau and PowerBI reports, creating executive-level dashboards with DAX and Power Query to track KPIs like engagement, revenue, and operational efficiency; reducing the time to interpret data by 40%.

**Data Engineer**                                                          **August 2023 – February 2024**
*CMR Solutions LLC*                                                                       *Richmond, VA*
- Managed requirements gathering and data ingestion, to improve data quality and optimize accordingly. Also Built and productionized a churn prediction ML model in Python, deployed using AWS services with orchestration via AWS Glue, resulting in a 9% increase in retention. And Automated churn model deployment on AWS with CloudWatch.

**Research Data Engineer**                                                     **May 2023 – August 2023**
*Kelly School Of Business(FADS) – Indiana University*                                    *Bloomington, IN*
- Created statistical models in R to analyze pricing strategies and provide insights for marketing decisions. Also conducted Market Basket Analysis using the Apriori algorithm, leveraging MongoDB to identify valuable product pairings.
- Analyzed customer purchase behaviors through cluster-based EDA in Python (pandas, seaborn), segmenting customers for targeted marketing. Developed a data validation pipeline in R, improving data accuracy and reducing errors by 20%.

**Data Scientist**                                                        **November 2021 – March 2022**
*Plotmydata*                                                                             *Hyderabad, India*
- Designed and deployed a tree-based ensemble machine learning model on predictive analytics for diverse use cases, ranging from academic performance prediction and classification of fake news, achieving accuracy over 95%, and developed an interactive Plotly-Dash dashboard to visualize model performance and validate its robustness.

## Projects

**City Service Request Analysis Dashboard** | *Python, Streamlit, PyDeck, AWS S3, AWS Lambda, AWS RDS*
- Developed an advanced dashboard to analyze and visualize service request data for the City of Bloomington, leveraging AWS services and utilized Python libraries to create interactive visualizations that aid in data-driven decision making.

**AppFusion a Personalized App Recommendation System** | *Python, MongoDB, Dash, Streamlit*
- Created an app that addresses cold-start problems for new or inactive users by suggesting customized mobile apps based on user behavior. The interactive interface and backend were developed using Python, MongoDB, and Streamlit.

## Technical Skills

**Programming/Scripting**: Python, SQL, R, Java, JavaScript, Matlab, C, TypeScript, HTML, CSS, Angular, React
**Cloud, Data, Databases and Analytics**: Microsoft Azure (Data Lake, Databricks, Data Factory), AWS (S3, EC2, Redshift, Glue, RDS), Snowflake, PostgreSQL, MySQL, MongoDB, Neo4j, Power BI, Tableau, Excel, SAS, Docker, JIRA
**ML and Data Science**: NLP (spaCy, NLTK, scikit-learn), Deep Learning, Neural Networks (PyTorch, TensorFlow), pandas, numpy, Vector Databases (ChromaDB, pgvector, FAISS), Search (Elasticsearch)
**Generative AI and LLM**: Large Language Models (OpenAI GPT, Claude), RAG Systems, Fine-Tuning, Prompt Engineering, Agentic Frameworks (PydanticAI, LangGraph, CrewAI)

## Education

**Master of Science, Data Science**                                           **August 2022 – May 2024**
*Indiana University*                                                                      *Bloomington, IN*
**Bachelor of Technology, Computer Science**                                    **June 2018 – May 2022**
*GITAM University*                                                                        *Hyderabad, India*

## Certifications

**AWS Certified Data Engineer – Associate**                                           [Certification Link]