Harsith Reddy Karne

425- 500-9089 ● harsithkarne@gmail.com ● www.linkedin.com/in/harsith-karne ● Portfolio

SUMMARY

A results-driven data professional and Master's candidate in Information Systems with a strong foundation in data engineering, machine learning, and statistical analysis. Proven expertise in building end-to-end data pipelines, developing predictive models, and leveraging large language models (LLMs) to create intelligent systems. Eager to apply a deep understanding of data-driven decision-making to solve complex problems and contribute to a data-focused team.

EDUCATION

University of Maryland, Robert H. Smith School of Business — **College Park, MD, USA**
Master of Science, Information Systems — May 2026

Osmania University — **Telangana, India**
Bachelor of Engineering, Computer Science — May 2024

TECHNICAL SKILLS

- **Programming & Data Science:** Python (Scikit-learn, Pandas, NumPy, TensorFlow, PyTorch), R, SQL, NoSQL, Java
- **AI & Machine Learning:** Predictive Modeling, Natural Language Processing (NLP), LLMs (Gemini), Regression, Classification, Clustering, Dimensionality Reduction (PCA), Recommender Systems
- **ETL & Big Data:** Apache Airflow, PySpark, Hadoop, Hive, Sqoop, Azure Data Factory, Data Warehousing, Data Modeling
- **Databases:** RedShift, PostgreSQL, MySQL, MongoDB, BigQuery, Oracle SQL
- **Cloud Technologies:** AWS (EC2, S3, RDS), Azure, Databricks, GCP
- **BI & Visualization:** Tableau, Power BI, Quicksight, Streamlit, Advanced Excel

WORK EXPERIENCE

Leftover Love, Inc. — **Oakland, MD**
Data Engineer Intern — June 2025 - Aug 2025

- Architected and deployed a robust ETL pipeline to centralize legacy data from 10+ disparate sources into a unified SQL database, improving data accessibility for over 100 users and reducing data retrieval times by 75%.
- Collaborated with executive leadership to translate business requirements into technical specifications, delivering a data system that enhanced operational decision-making efficiency by 40%.

AICTE (All India Council For Technical Education) — **Hyderabad, TG, India**
Data Engineer — May 2023 – July 2024

- Engineered end-to-end data solutions, integrating diverse sources into a central Data Lake and automating ETL processes, which reduced manual data processing workload by 20%.
- Spearheaded the automation of 5 critical security functions using RPA, cutting manual intervention by 50% and saving over 15 work-hours weekly.
- Improved overall system reliability by 30% through comprehensive root cause analysis of automation failures.

Palo Alto Networks — **Hyderabad, TG, India**
Data Analyst Intern — March 2022 – May 2022

- Developed and deployed predictive models using advanced statistical methods to analyze over 50,000 weekly security logs, enabling the early identification of potential threats up to 48 hours in advance.
- Created executive-level Tableau dashboards to visualize security trends, directly informing strategic investments and policy changes across the organization.

PROJECTS

AI-Powered Natural Language to SQL Agent (*Python, Gemini LLM, Streamlit, PostgreSQL, Microsoft AutoGen***)**

- Engineered an autonomous agent leveraging the Gemini API to translate natural language questions into executable SQL queries, successfully democratizing database access for non-technical users through a Streamlit-based UI.
- Implemented a multi-agent workflow using the Microsoft AutoGen framework to build a secure and reliable data retrieval pipeline, significantly improving the accuracy of dynamic query execution.

NBA Performance Analytics & Predictive Modeling (*Python, SQL, Tableau, Scikit-learn***)**

- Engineered an end-to-end data pipeline to process and analyze over 10 years of multi-source NBA data (2M+ records), identifying key performance indicators through advanced statistical analysis.
- Applied regression modeling to uncover strategic insights into player performance and team success, providing actionable recommendations that influenced team decision-making.

Gene Expression Classification for Cancer Diagnosis (*Python, Scikit-learn, XGBoost, Power BI***)**

- Increased a cancer diagnostic model's accuracy from 76% to 92% by conducting data quality checks, applying PCA for dimensionality reduction, and implementing hyperparameter-tuned Random Forest and XGBoost classifiers to identify patterns across 10 distinct tissue types.