

Ramesh Naik

ramesh.naik13@gmail.com | [LinkedIn](#) | +1 (930) 333 4993 | github.com/Ramesh2019

TECHNICAL SKILLS

Machine Learning & AI: Deep Learning, NLP, LLM, Predictive Modeling, Generative AI, Statistics, A/B Testing, Causal inference
Programming & Frameworks: R, Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch), LangChain, DataLake
Data Engineering & Pipelines: Data Preprocessing, ETL, Apache Airflow, PySpark, Hadoop, Docker, Kubernetes, Delta, DBT
Cloud & Visualization: Google Cloud Platform (GCP), AWS (EMR, Lambda, S3, Glue, EKS), Azure, Databricks, Power BI, Tableau
Databases & Storage: PostgreSQL, Redshift, Snowflake, MongoDB, DynamoDB, Neo4j, NoSQL, TSQL, Github

PROFESSIONAL EXPERIENCE

Data Engineer

O'Neill School of Public and Environmental Affairs

Nov 2024 – Ongoing

Bloomington, IN, US

- Engineered AI-powered **data pipelines** to process and classify over 270,000 nonprofit mission statements, utilizing LLMs for advanced feature extraction and automated organizational classification, accelerating data curation efforts.
- Boosted ETL pipeline efficiency** by 35% by migrating data transformation logic from Pandas to SQL, resulting in faster query execution, streamlined updates to normalized schema, and improved responsiveness for ad-hoc analytical queries.
- Conducted large-scale data integrity** assessments on 275,000+ tax records using SQL-based validation logic, identifying critical gaps and improving the consistency of cross-dataset joins.

Data Engineer

Sutherland Global Services

Nov 2021 – Jul 2023

Mumbai, MH, India

- Built an **end-to-end ML workflow** to predict loan defaulters across delayed payment categories (30, 60, 90+ days) using historical loan data, implementing LightGBM for **financial risk modeling and classification**.
- Authored **PostgreSQL** scripts to transform quarterly raw loan data and integrated **Apache Airflow and Apache Spark** to automate ETL, feature selection, statistical analysis, data visualization, and dataset splitting for **risk analytics**.
- Trained **machine learning and deep learning** models to predict the failure of over 10 components within CT and MRI machines, utilizing historical maintenance records, sensor data, and field reports for **Philips Healthcare**.
- Implemented machine learning models such as Random Forests and Gradient Boosting Machines (LighGBM) and deep learning approaches for classification.
- Maintained a **precision of over 87%**, achieving predictions that facilitated maintenance and reduced equipment downtime.

Data Analyst

Eclerx Services Ltd.

Jun 2019 – Oct 2021

Mumbai, MH, India

- Engineered a backend system for an EdTech venture on **AWS**, covering data gathering and personalized career, learning content, and job role recommendations for over 10,000 user profiles.
- Harnessed machine learning models for targeted career and course suggestions, enhancing user engagement and learning outcomes by **tailoring recommendations** to individual profiles.
- Revamped a chat analytics platform analyzing 12,000+ customer interactions to **extract key insights on customer** sentiment, agent performance, and sales engagement.
- Fine-Tuned an NER model to **tag entities like customer issues and agent greetings**. Identified trends in complaints, promotions, and service efficiency, resulting in a 20% increase in customer satisfaction.

PROJECTS

Distributed Data Processing and Visualization Pipeline

- Built a scalable data pipeline by deploying Apache Spark on Jetstream cloud to process 10GB+ data, integrating PySpark workflows for daily transformation of 500K+ records, and storing 1M+ entries in a sharded MongoDB cluster with high availability.
- Enabled insightful analytics by generating 10+ weekly statistical visualizations using Seaborn, accelerating trend discovery and enhancing decision-making efficiency by 35–40%

Weather Data ETL Using Kafka

- Architected a scalable 3+ **multi-node Kafka** infrastructure with country-wise topics, **enabling parallel ingestion** and improving real-time streaming efficiency by ~20%.
- Programed Kafka producers and consumers to stream weather metrics into **PostgreSQL**, supporting ETL on 100K+ records daily.
- Configured Kafka brokers, topics, and consumer groups to **ensure high ingestion reliability** (~99%), enhancing pipeline stability.

EDUCATION

Indiana University Bloomington

Masters (MS) in Data Science: (CGPA: 4.0 / 4.0)

Aug 2023 – May 2025

Bloomington, IN, US

University of Mumbai

Bachelors of Engineering (BE) in Computer Science

Jul 2015 – May 2019

Mumbai, MH, India

CERTIFICATION

- AWS Certified AI Practitioner
- Alteryx Designer Core Certification
- Microsoft Certified: Power BI Data Analyst Associate
- AWS Certified Machine Learning Engineer – Associate
- Astronomer Certification for Apache Airflow Fundamentals
- Databricks Certified Data Engineer – Associate (InProgress*)