

Kiran Rao

Snowflake | ETLs | Dashboards | LLMs | Hugging Face | SQL

sathyasunkojirao.k@northeastern.edu | 617-516-3845 | linkedin.com/in/kiran-ss | [Portfolio](#)

EDUCATION

Master of Science in Information systems | Northeastern University | Boston, MA

Sep 2024 – May 2026

Relevant Courses: DBMS, Data Science Eng., Big Data, Data Analytics

Bachelor of Technology in Computer Science & Engineering | PES University | Bangalore, India

Sep 2020 – May 2024

PROFESSIONAL EXPERIENCE

Northeastern University | Boston, MA

Teaching Assistant/ Data Science

Jan 2025 – Apr 2025

- Mentored 50+ students in **Data Analytics, Statistics, Machine Learning & Generative AI** providing guidance on core concepts
- Identified gaps in students' grasp of statistics, and machine learning concepts; designed targeted quizzes and labs
- Observed low engagement with **GenAI** concepts; developed stepwise walkthroughs for **GANs, VAEs, & Diffusion Models**
- Recognized inconsistent project quality; facilitated live coding, review sessions, increasing assignment completion rates by **15%**.

Indian Institute of Science | Bangalore, India

Data Science - ML Intern

Jun 2023 – May 2024

- Evaluated practices for EEG artifact detection, researched existing methods, identifying techniques that shaped the solution approach
- Identified inefficiencies in data collection, automated the process by building an integrated workflow in OpenViBE, enabling synchronized data capture, reducing data collection time by 40%
- Detected inefficiencies in EEG artifact detection, performed feature engineering on EEG signals to extract meaningful features
- Developed **CNN & Random Forest models** with advance denoising filters —notch, bandpass, boosting classification accuracy to 92%
- Observed lack of automation in Object detection, built a solution using YoloV5 with **KPIs**, improving accuracy & monitoring
- Noted absence of structured training data, curated a subset with annotated bounding boxes using Roboflow, enabling model training
- Flagged suboptimal detection rates, engineered a **Transformer** pipeline using **PyTorch**, improving detection accuracy by 75%
- Addressed occlusion-related recognition issues, applied **KMeans** and **GMM** clustering to accurately classify & group similar products, improving overall product classification accuracy by 15%
- Solved low-sample-size classification challenges using **few-shot learning**, achieving 85%+ accuracy

PROJECTS | <https://github.com/kiranss777/AI-Projects>

Q&A Assistant for NVIDIA 10-K and 10-Q Filings | Python, Langchain, Airflow, Pinecone

- Built an assistant to analyze NVIDIA's filings (10-K & 10-Q), enabling exploration of company strategy, risks, & financial health
- Automated data collection via web scraping, followed by Airflow ETLs for chunking, & embedding data into Pinecone vectorDB
- Detected absence of real-time data, integrated SerpAPI to fetch real-time data for hotels & flights enabling dynamic decision-making
- Exposed system via a FastAPI interface, supporting structured financial queries and generating analyst-ready summaries in real time

British Airways Review Analytics Dashboard | Python, Tableau

- Identified gaps in customer sentiment monitoring by analysing review patterns, feedback trends, designed 3 dashboards for **KPI reviews & sales forecasting**, enabling data-driven decision making that improved customer satisfaction scores by 25%
- Solved exploration limitations using **5 filters** including month, ratings, and regional segmentation, reducing analysis time by 50%
- Addressed strategic planning inefficiencies by translating review data into actionable insights via **Tableau** visualizations

SEC Financial Data Pipeline | Python, Airflow, Snowflake, DBT

- Engineered a data pipeline processing **3M+ row of SEC Financial data** stored it on AWS S3 and staged in **Snowflake** for processing
- Detected inconsistent reporting, automated **ETL** workflows in **Airflow** and created fact tables, **improving reliability by 20%**
- Reduced data inconsistencies by 30% by orchestrating validation and transformation pipelines with **DBT** resulting in reliable datasets

TECHNICAL SKILLS

Programming Languages

: Python, R, SQL

Data Science

: Regression, Classification, PCA, Time Series Analysis, XGBoost

Data Engineering

: Snowflake, Amazon S3, EC2, RDS, Azure Datalake, Databricks, Pinecone

AI technologies

: LangChain, LangGraph, Agentic AI, RAGs, Model Context Protocols

BI & Data Visualization

: Power BI, Tableau, Excel

CERTIFICATIONS

- Generative AI | Databricks
- Apache Airflow | Astronomer
- Model Context Protocol | Anthropic

PUBLICATIONS | <https://ieeexplore.ieee.org/document/10782852>

Advancing Medical Imaging: A Comparative Exploration of Generative Adversarial Networks for Chest X-ray Synthesis