

# Sanjaikumar Balasubramaniyan

🌐 [www.sanjaibala.com](http://www.sanjaibala.com) | <https://www.linkedin.com/in/sanjai-bala> | <https://github.com/sanjai-11>

📍 San Francisco, CA, United States | 📞 +1 (812) 671-6737 | ✉️ [bsan@iu.edu](mailto:bsan@iu.edu)

## SUMMARY

Machine Learning Engineer specializing in large-scale recommendation, ranking, and personalization systems. Experience building LLM-powered discovery models, user engagement optimization pipelines, and scalable data-driven ML solutions. Strong background in NLP, behavioral modeling, and applied AI product development using Python, PyTorch, and cloud ML infrastructure. Passionate about designing intelligent content discovery and recommendation systems that improve user experience at scale.

## EDUCATION

### Master of Science in Data Science

GPA: 3.5/4

Indiana University - Bloomington, IN, USA

August 2024 – May 2026

**Coursework:** Applied Algorithms, Statistics, Advanced Database Concepts, Machine Learning, Data Mining, LLM, Applied Database Technologies, Engineering Cloud Computing

## WORK EXPERIENCE

### Research Data Scientist

January 2025 – Present

Indiana University

Bloomington, IN, USA

- Built **ML pipelines** analyzing multi-year nonprofit funding datasets (1M+ records) to **model behavioral allocation patterns and ranking signals**.
- Designed scalable preprocessing and feature engineering workflows **improving downstream modeling** reliability and recommendation signal quality.
- Performed **statistical modeling** and **user behavior analysis** to identify engagement drivers and allocation disparities.
- Developed **visualization dashboards** translating large-scale data insights into **actionable decision strategies**.

### Founding AI Engineer Intern

July 2025 – Nov 2025

riAI Capital LLC.,

Reno, NV, USA

- Developed **LLM-powered advisory research assistant** integrating prompt engineering, retrieval, and ranking logic improving research turnaround by **40%**.
- Designed reusable **prompt templates** and **evaluation metrics** to optimize recommendation consistency and model reliability.
- Built **benchmarking pipelines** and **adoption tracking** frameworks to measure engagement and **model performance**.
- Collaborated cross-functionally with product and compliance teams to productionize **ML-driven decision systems**.

### Data Analyst

April 2023 – July 2024

Systech Solutions

Chennai, TN, India

- Built **customer segmentation** and **recommendation clustering models** improving campaign targeting accuracy by **25%**.
- Automated ETL and feature engineering pipelines using Python and SQL reducing data processing latency by **35%**.
- Developed Power BI dashboards enabling real-time decision analytics **improving business response** time by 40%.
- Partnered with business teams to productionize **ML-driven analytics** workflows supporting revenue optimization.

## PROJECTS

### Scalable ETL Pipeline for Open Data Analytics | Python, SQL, Apache Airflow, Pandas, PostgreSQL, AWS

- Built an **automated ETL pipeline** using Python and SQL to ingest, clean, validate, and store large public datasets into a **relational database**, enabling reliable downstream analytics and reporting

### AI-Powered Recommendation & Conversational System | Python, JavaScript, Gemini API, NLP, Semantic Search

- Built an AI-powered conversational recommendation system using **Google Gemini LLM**, integrating semantic search, user intent modeling, contextual memory, and adaptive ranking logic to simulate **real-world content discovery** and **personalization flows**.

## TECHNICAL SKILLS

**Programming & ML:** Python, R, C/C++, SQL, JavaScript, PyTorch, Scikit-learn, NLP

**Machine Learning Domains:** Recommendation Systems, Ranking Models, NLP, Information Retrieval, Personalization

**AI / LLM:** Prompt Engineering, Retrieval Augmented Generation (RAG), Conversational AI, Embeddings, Semantic Search

**Data Engineering & Big Data:** ETL Pipelines, Feature Engineering, Kafka (Basic), AWS (S3, Lambda), Airflow

**Visualization & Analytics:** Power BI, Tableau, Matplotlib, Seaborn, Gephi

**Developer Tools:** Git, Docker, REST APIs, RabbitMQ, Jira, Scrum

**Soft Skills:** Communication, Analytical Thinking, Ideation, Decision Making, Time Management.

**Research Interests:** Recommender Systems, NLP, Ranking Algorithms, LLM Personalization, User Behavior Modeling