

Sanjaikumar Balasubramaniyan

Data Scientist

www.sanjai-11.github.io | (812) 671-6737 | bsan@iu.edu | [LinkedIn](#) | [GitHub](#) | [LeetCode](#)

TECHNICAL SKILLS

Programming: Python, R, SQL, C/C++

Technical Tools: PyTorch, TensorFlow, Tableau, PowerBI, Pandas, NumPy, Keras, Scikit Learn, Hugging Face, Computer Vision, Hypothesis Testing, Data Wrangling.

Cloud & Databases: PostgreSQL, MSSQL, MongoDB, Neo4j, Hadoop, Hive, AWS (S3, EC2, SageMaker, Glue, QuickSight), GCP (BigQuery, DataFlow, AutoML, VCP)

ML & DL Algorithms: Regression, Classification, Clustering, EDA, CNN, RNN, LSTM, MLPs, GRU, GANs

Developer Tools: VS Code, Docker, Postman, GitLab, GitHub, Bitbucket, JIRA

Certifications: Google Professional Data Analytics, Python for Data Science (NPTEL, Elite), ML by Harvard (edX)

PROFESSIONAL EXPERIENCE

Research Data Scientist

Indiana University, Bloomington, IN

Jan 2025 – Present

- Mapped funding networks for **157,000+ BIPOC nonprofit grants** to analyze funding trends and relationships.
- Overcame data limitations with advanced retrieval techniques for comprehensive analysis.
- Conducted network analysis to identify **bonding, bridging, and linking patterns** in grant funding.
- Classified grant recipients and funders by demographic focus, uncovering insights into funding equity for underserved communities.

Data Scientist Intern

Systech Solutions Inc, Chennai, India

Apr 2024 - July 2024

- Built **real-time Power BI dashboards**, enhancing decision-making speed and visualization quality by 40%.
- Automated data workflows with optimized **ETL pipelines**, cutting data processing time by 35%.
- Deployed machine learning models, **boosting customer segmentation** accuracy by 25% for predictive insights.
- Engineered **scalable data solutions**, enhancing cross-team operational efficiency by 30%.
- Conducted **Data Quality** assessment and **validation checks**, ensuring the accuracy and integrity of data pipelines across all operational teams.

Data Analyst

8Queens Software Technologies Pvt Ltd., Chennai, India

May 2022 – Jan 2023

- Streamlined real-time data processing by **transforming** datasets, improving accessibility and accuracy.
- Created interactive Tableau dashboards for sales, **driving a 15% improvement** in performance insights.
- Deployed supervised ML model for **diabetes risk prediction**, boosting predictive accuracy in **health analytics**.
- Delivered actionable insights through data-driven solutions, supporting strategic **decision-making** processes.

PROJECTS

AI Agent for Operating Backend System | LLM, Flask, SQLite, Google Gemini-pro

- Built an AI agent using Gemini-pro LLM, automating workflows with 50% improved efficiency and 80% fewer manual tasks at **95% accuracy**.
- Developed a Flask API with SQLite, cutting response times by 40%, supporting **1,000 queries**, and boosting backend efficiency by **30%**.

Advanced ANPR for Vehicle Management | YOLO, OCR, Python, SQL

Published in Adalya International Journal, [DOI: 10.37896/aj13.3/001](#)

- Built a **YOLO-based** model with **OCR** for license plate detection, achieving **96% accuracy** and enhancing blurry images by **85%** through advanced preprocessing.
- Developed a scalable system with **automated vehicle tracking**, entry/exit logging, and **92%** overall accuracy.

EDUCATION

Master of Science in Data Science

Indiana University, Bloomington, IN

Aug 2024 - May 2026

- Relevant Coursework: Statistics, Applied Algorithms, Advanced Database Concepts, Machine Learning, LLMs, Data Mining