

# Sankaran S

sankaran.sangumuthu@gmail.com | +91 9677917815 | Tamil Nadu, India

sankaran-s2001.github.io/portfolio | linkedin.com/in/sankaran-s21 | github.com/sankaran-s2001

## Professional Summary

---

Junior Data Scientist with **1.5+ years** of proven experience building production-ready ML systems, RAG applications, and computer vision solutions. Demonstrated expertise in delivering **90%+ accuracy models** and automated document-processing pipelines solving real business problems across multiple industries. Successfully completed **20+ end-to-end data science projects** from concept to deployment, with strong proficiency in Python, SQL, machine learning, and modern AI frameworks. Seeking to contribute advanced analytics and automated ML solutions to drive data-driven decision making.

## Technical Skills

---

**Programming:** Python (Advanced), SQL (PostgreSQL, MySQL), JavaScript, HTML/CSS

**Machine Learning:** Scikit-learn, XGBoost, Random Forest, SVM, Linear/Polynomial Regression, Decision Trees, K-NN

**AI & Deep Learning:** NLP, Computer Vision, YOLO, OpenCV, PyTorch(Basics)

**Advanced AI:** LangChain, RAG Systems, HuggingFace, FAISS, Chroma, Vector DB, OpenAI API

**Analytics & Visualization:** Power BI, Advanced Excel, Pandas, NumPy, Matplotlib, Seaborn, Plotly

**Tools & Platforms:** Streamlit, Gradio, Docker, Git, Flask, FastAPI, Linux, ETL Pipelines

## Work Experience

---

**Junior Data Scientist**, F & R Data Analytical System & Solutions Pvt. Ltd. Jan 2024 – May 2025

- Developed computer vision models for brand logo detection achieving **92% accuracy**, processing **100K+ images weekly** using Python, OpenCV, and deep learning frameworks
- Built automated receipt extraction system handling **5,000+ documents daily**, reducing manual processing time by **75%** using Python and OCR technologies
- Performed comprehensive data labeling and annotation for **50+ brand categories**, improving model training efficiency by **40%** through optimized workflows
- Collaborated with analytics team to deliver comprehensive reports, increasing **client satisfaction by 30%** through reliable automation and real-time processing systems

**Data Analyst Intern**, Nobeth Analytics

Jul 2023 – Dec 2023

- Developed telecom customer churn prediction model achieving **85% accuracy** on **100,000+ customer records** using Python and scikit-learn
- Created interactive Power BI dashboards reducing report generation time from **8 hours to 2 hours weekly** through automated data visualization
- Identified key churn factors through statistical analysis, resulting in **15% improvement** in customer retention strategies and comprehensive data pipeline development
- Built automated reporting systems processing **500K+ row datasets** using SQL optimization techniques for real-time business intelligence

## Projects

---

### SmartPDF AI-RAG System

GitHub

- Built AI-powered document Q&A system with LangChain & HuggingFace integration for advanced document intelligence
- Implemented context-aware responses using PDF processing, vector embeddings, and memory-enhanced conversations with FAISS vector database

### Object Detection Recognition System

GitHub

- Developed real-time object detection web application using YOLOv8 and OpenCV achieving accurate detection for **80+ object classes**
- Built interactive Streamlit interface with image upload, automated object counting, and visualization charts for comprehensive object analysis

### Machine Learning & Analytics Suite

- Created comprehensive ML models: IPL Score Predictor, Coffee Shop Revenue Predictor, Diabetes Risk Predictor achieving **80-85% accuracy**
- Developed recommendation systems (movie & book engines) with content-based filtering, collaborative filtering, and OMDB API integration

### Advanced Power BI Dashboard Collection

- Built professional dashboards: Data Professional Survey (500+ professionals), E-commerce Sales Analytics, Healthcare Analytics, Employee Attrition Analysis
- Conducted large-scale data analysis: US Traffic Accidents (2016-2023), NYC Airbnb Market Analysis, Global Layoffs SQL Analysis (2020-2023)

## Education

---

**B.Tech Mechanical Engineering**, Kalasalingam Academy of Research & Education      2019 – 2023

Strong foundation in applied mathematics, statistics, and analytical problem-solving. Developed technical skills in data analysis and computational methods with projects involving statistical modeling and engineering analytics.

## Certifications

---

**IBM Data Analyst Professional Certificate** (June 2025)

[View Certificate](#)

11 comprehensive courses: Introduction to Data Analytics, Python for Data Science AI & Development, Data Analysis with Python, Data Visualization with Python, SQL for Data Science, IBM Data Analyst Capstone Project

**Python for Data Science** – Simplilearn (2024)

**SQL Basic & Intermediate** – HackerRank (2024)

**SAP Certified Application Associate** – Procurement with SAP ERP 6.0 EhP7