# Rajagopal Somaskandan

*Chennai, India* | *+91 7200288242* | *rajagopal-somaskandan.netlify.app* | *somaskandan931@gmail.com* | *https://linkedin.com/in/somaskandan-rajagopal* | *github.com/Somaskandan931*

## PROFESSIONAL SUMMARY

Applied Machine Learning Engineer specializing in explainable ML and NLP systems, with experience building production-ready fraud detection, semantic search, and ranking pipelines using XGBoost, LightGBM, and transformer-based models. Strong focus on model evaluation, SHAP-based explainability, and FastAPI deployment for real-world business use cases.

## TECHNICAL SKILLS

**Programming & Databases** : Python, Java, SQL (PostgreSQL), MongoDB
**Machine Learning** : Supervised and Unsupervised Learning, Feature Engineering, Model Training, Hyperparameter Tuning, Evaluation Metrics
**ML Algorithms** : Linear Regression, Logistic Regression, Decision Trees, Random Forest, XGBoost, LightGBM
**NLP & Generative AI** : Transformers, spaCy, NLTK, Sentence Embeddings, Semantic Search, Retrieval-Augmented Generation (RAG), Prompt Engineering, LangChain, ChromaDB
**Data Analysis & Visualization** : Pandas, NumPy, Matplotlib, Plotly, Streamlit, Power BI
**APIs, Deployment & Tools** : FastAPI, Flask, REST APIs, Docker, Git/GitHub, MLflow, AWS (Basics)

## EDUCATION

| | | |
|---|---|---|
| Sathyabama Institute of Science and Technology | **B.Sc. Computer Science (AI Specialization)** | First Class | *2025* |
| Sathyabama Institute of Science and Technology | **M.Sc. Computer Science** | | *2027* |

## PROJECTS

**PaperMind: AI-Powered Academic Paper Recommender** | *OpenAI Embeddings, FAISS, FastAPI, React, Semantic Scholar API* 2025
- Built a full-stack semantic search system for academic papers using OpenAI embeddings and FAISS for vector similarity search.
- Implemented FastAPI backend with vector indexing and GPT-powered relevance explanations for real-time paper retrieval.
- Developed a responsive React frontend with Tailwind CSS, enabling topic-based queries and PDF uploads for semantic matching against paper abstracts.

**EXPLAIN: Multilingual Fake News Detection with Hybrid Routing** | *XLM-RoBERTa, IndicBERT, FastText, PyTorch* 2025
- Developed a multilingual fake news detection system for 86k+ samples across 30+ languages, using automatic script detection and model routing for English, Indic, and low-resource languages.
- Fine-tuned XLM-RoBERTa achieving **91.7% accuracy and 0.914 macro-F1** , outperforming TF-IDF + Logistic Regression by 7.4 points.
- Consolidated three heterogeneous datasets (BharatFakeNewsKosh, IFND, FakeReal) with automated preprocessing, language detection via FastText, and standardized labeling for reproducible pipelines.

**FinGuard Pro: Explainable AI for Financial Fraud Detection** | *Python, SQL, XGBoost, SHAP* 2025
- Detected financial fraud with **94% recall and ROC-AUC 0.995** by architecting an end-to-end Python/SQL/XGBoost pipeline on imbalanced UPI transactions.
- Enabled transaction-level audit with SHAP-based explainability for reporting and auditing purposes.
- Reduced false positives by implementing rule-based AML screening using velocity and structuring patterns.

**SourceUp: Explainable Procurement Intelligence Platform** | *Sentence-BERT, FAISS, LightGBM, FastAPI, Redis* 2025
- Improved supplier sourcing efficiency with a semantic search engine using Sentence-BERT embeddings and FAISS.
- Increased result relevance with LightGBM learning-to-rank model achieving **NDCG@10 0.7439 (+9.2%)** .
- Delivered low-latency retrieval for 10k+ suppliers using FastAPI microservices and Redis caching.

**CareerGenie: Resume-to-Job Matching System (RAG)** | *Sentence-Transformers, ChromaDB, LaTeX* 2025
- Enhanced job matching with Retrieval-Augmented Generation combining semantic retrieval and LLM reasoning.
- Enabled query- and location-aware search with embedding-based indexing using Sentence-Transformers and ChromaDB.
- Streamlined ATS-ready resume creation by converting structured profiles into LaTeX resumes.

## EXPERIENCE

**Data Analyst Intern** | *India Advocacy*                    *July 2024 – October 2024*
- Cleaned, filtered, and structured **over 1M web records** and curated 15–20k high-quality datasets to support targeted outreach campaigns.
- Built automated data extraction pipelines using Java and Selenium to improve data collection efficiency.
- Worked with real-world data quality issues, implementing validation checks aligned with business usability requirements.

**Security Trooper (National Service)** | *Singapore Armed Forces*                    *January 2020 – January 2022*
- Supported security operations, operated authorized transport vehicles under regulated operational environments.
- Assisted Ministry of Health COVID-19 task force initiatives through structured data collection and reporting.

**Sales Associate (Part-Time)** | *Bata*                    *February 2022 – August 2022*
- Worked in a high-volume retail environment, managing customer interactions and daily operational responsibilities with strong communication, time management, and teamwork skills.

## PUBLICATIONS & CONFERENCES

**FinGuard Pro: Explainable AI for Financial Fraud Detection** | *ICAAIC 2025*                    *ISBN: 979-8-3315-6587-9*
**Real-Time Explainable Fake News Detection using BERT-LSTM** | *ICASET 2025*

## CERTIFICATIONS

**AI and Machine Learning** | *SpaceZee Technologies*                    *June 2024*
**Python Core** | *SpaceZee Technologies*                    *February 2024*