

Sunil Samson Suresh

Hyderabad, India +91-9182757352 sunil.samson@outlook.com

SUMMARY

Senior Data Scientist with 5+ years building and deploying production grade LLM and ML systems across insurance, immigration tech, and legal domains, specializing in GenAI, agentic systems, and end to end delivery on Azure and Databricks. Leads technical decisions across cross functional teams, mentors junior members, and sets engineering standards for how AI features are scoped, built, and shipped. Strong foundation in Python, PyTorch, TensorFlow, SQL, Docker, Kubernetes, and FastAPI, backed by MLOps practices, open source contributions to LlamaIndex, and an IEEE publication.

EXPERIENCE

11/2025 – present
Hyderabad, India

Senior Data Scientist VALUEMOMENTUM SOFTWARE SERVICES

- Lead fraud and severity data modelling for Allied World's Claims Analytics team across 9,000 Program Business Claims, building signals and features with claims stakeholders to flag high-risk claims early, allowing adjusters to intervene before costs escalate
- Serve as primary technical liaison between data science and business stakeholders to evaluate technical feasibility of discovery sessions, present findings in plain language, and tie model outputs to concrete business decisions
- Lead end-to-end transformation of document triage, severity prioritisation, and intervention decisioning, replacing manual workflows with automated ML pipelines to deliver measurable gains in throughput, accuracy, and cost

11/2023 – 10/2025
Tallinn, Estonia

Data Scientist JOBBATICAL OÜ

- Led the technical vision for platform Data Science and GenAI integration, partnering with Product and Engineering to set engineering standards for scoping, building, and shipping AI-powered features
- Architected end-to-end autonomous decision workflows using GPT, Mistral and Claude by leveraging LangGraph and Llama-Index framework and open-source models, deployed as production-grade FastAPI microservices to power real-time automation at scale
- Designed a conversational RAG based agentic system for the message composer backed by MongoDB vector search, grounding LLM outputs in live case data to ensure accurate, status specific responses and eliminate stale outputs
- Built OCR and Computer Vision pipelines to parse identity documents across multiple countries, using vision models to extract and validate structured PII fields

10/2021 – 10/2023
Hyderabad, India

Jr. Data Scientist APP ORCHID INC

- Fine tuned transformer models including BERT, RoBERTa, and LegalBERT for multi class NLP tasks such as entity recognition, clause categorisation, and document labelling, optimising accuracy across diverse legal corpora through systematic experimentation
- Built end to end NLP pipelines for structured information extraction from large legal document corpora, covering entity tagging and liability clause identification, directly powering downstream compliance and decision systems for enterprise clients
- Applied Bayesian probabilistic modeling using Pyro for fraud risk forecasting and uncertainty quantification, producing calibrated confidence estimates that enabled stakeholders to make risk based decisions with explicit uncertainty bounds

01/2021 – 08/2021
Hyderabad, India

Data Scientist Intern CORTEVA AGRISCIENCE

Developed time series forecasting models to predict crop lifecycle events, pest outbreaks, and disease risks using historical and environmental data, built predictive analytics pipelines integrating weather, satellite imagery, and temporal signals for early warning systems, and performed feature engineering on temporal and geospatial data to improve forecasting accuracy.

🎓 EDUCATION

2017 – 2021
Manipal, India

Mechanical Engineering (Spec: Data Science) | Bachelors of Technology
Manipal Institute of Technology
GPA: 8.28/10

🧪 SKILLS

– LANGUAGES

PythonSQL

– MACHINE LEARNING

Deep Learning	NLP	Statistical Inference	A/B Testing
Bayesian Modelling	XGBoost	Scikit-learn	PyTorch
TensorFlow	OpenCV		

– GENAI

RAG Architecture	Agentic AI	Prompt Engineering	LlamaIndex
LangGraph	PydanticAI	Transformers	

– CLOUD

Microsoft Azure	Azure Databricks	Google Cloud Platform	PySpark
MongoDB	Kubernetes		

– MLOPS AND ENGINEERING

FastAPI	REST APIs	Microservices	Docker
Kubernetes	MLFlow	CI/CD	GitHub Actions
Code Reviews			

📄 CERTIFICATIONS

Databricks Certified Machine Learning Associate
Databricks
<https://credentials.databricks.com/4b2dea3a-3ff4-4f66-bff0-7aff5c015f7e#acc.3oLYISHO>