

# DEBANJAN SAHA

+1 (781)-600-6019 | [saha.deb@northeastern.edu](mailto:saha.deb@northeastern.edu) | [LinkedIn](#) | [GitHub](#) | [Medium](#) | [Portfolio](#)

## SUMMARY

Experienced **Data Scientist** with a decade of expertise and a track record of leading **end-to-end ML and AI projects** while designing **cloud-agnostic AI solutions** for **real-time decision making**. Proficient in **multi-cloud platforms** such as **AWS, Azure, and GCP**, as well as **Python, natural language processing, machine learning, deep learning frameworks** and sophisticated analytics based on **machine learning** and **generative AI**. Seeking **Senior Data Scientist** or **Senior Machine Learning/AI Engineer** positions.



## PROFESSIONAL EXPERIENCE

- Principal AI Engineer | Nanobiosym | Cambridge, MA | Stealth Mode Healthcare Startup** 05/2023 – 12/2023
- Built COVID-19 **severity prediction** model using **LSTM**, analyzing patient data and improving **intervention response time** by 30%
  - Fine-tuned clinicalBERT** for **NER** on patient conversations, achieving F1 score of 0.87 in symptom extraction and severity assessment
  - Devised a **conversational AI chatbot** using Amazon **Lex** and **Polly**, increasing patient **engagement** by 40% and symptom collection
  - Orchestrated **feature extraction** in **AWS SageMaker Pipeline**, boosting model accuracy by 8% leveraging medical ontologies
  - Established **severity binning** for patient classification, triggering **SNS alerts**, reducing **response time** for **critical cases** by 20%
- Senior Data Scientist | Prescriber 360 | India | Marketing Science** 09/2021 – 08/2022
- Optimized media allocation through **MMMs**, boosting **campaign efficiency** by 20% using **regression coefficient analysis & optimization**
  - Developed **MTA models** to evaluate marketing channel effectiveness, resulting in 15% **ROI increase** by optimizing **channel touchpoints**
  - Enhanced advertising strategies using **advanced time-series analysis** with **hybrid LSTM** models, increasing **engagement** by 12%
  - Automated **KPI tracking** and campaign dashboards using **Power BI**, delivering real-time insights and reducing reporting time by 50%
  - Delivered **data-driven marketing insights** to senior leadership, improving campaign **decision-making efficiency** and **spends** by 10%
- Senior Data Scientist | Cognizant | India | Retail – Search & Recommendation** 10/2019 – 09/2021
- Analyzed **TB-scale** retail datasets using **SQL** and **Spark** on **AWS EMR**, identifying trends that boosted **search conversion rates** by 12%
  - Engineered **scalable ML pipelines** in AWS using **Sagemaker, Glue, Airflow** and **S3**, reducing **model deployment time** by 40%
  - Implemented **NLP solutions** in AWS, leveraging **Word2Vec embeddings** for **semantic search**, improving **query relevance** by 20%
  - Formulated **ranking** and **recommendation models**, improving **product search accuracy** by 15% and **customer engagement** by 10%
  - Accelerated **product search queries** using **LSH**, lowering **latency** by 25% and increasing **relevance** for better **user satisfaction**
- Data Scientist | Tata Consultancy Services | India | Banking & Finance – Anti Money Laundering (AML)** 12/2015 – 09/2019
- Authored **SQL queries** to extract transactional data for **SAR analysis**, enabling efficient identification of suspicious activities
  - Built ML models for **AML risk scoring** using **Python** and **SQL**, improving detection accuracy and **reducing false positives** by 30%
  - Migrated transactional data from **legacy SQL systems to cloud**, optimizing data processing and enabling **ML** for **AML** monitoring
- Associate Data Warehouse Engineer | ITC Infotech | India | Banking & Finance** 07/2013 – 11/2015
- Architected **snowflake/star schemas**, enhancing data model performance by 10%, ensuring optimal analytical capabilities
  - Implemented **materialized views** and **indexing**, reducing query response time 25%, enhancing precompute and speeding up queries

## TECHNICAL SKILLS

- Programming:** Python, Scala, PySpark, Matlab, C/C++, SQL(Postgres, MS-SQL, MySQL), HTML, CSS, React, Node.js
- ML Frameworks:** Scikit-learn, Tensorflow, TFX, TFDV, Keras, PyTorch, XGBoost, Spark ML, Huggingface, OpenAI, LLM, RAG, Langchain
- DevOps and MLOps:** Git, Docker, Kubernetes (AKS, EKS, GKE), Helm, DevOps, MLOps, LLMops, MLFlow, KubeFlow, ELK, Prometheus
- Data Visualization:** Tableau, PowerBI, Quicksight, Gantt Chart, Advanced Excel, Looker, Alteryx, Jupyter Notebook, Data Storytelling
- Cloud:** AWS (Amazon Web Services), Microsoft Azure, GCP (Google Cloud Platform), Sagemaker, Vertex AI, Azure ML, Azure Databricks
- Tools & Frameworks:** Git, GitHub, GitLab, JIRA, Terraform, Docker, Jenkins, Postman, VS Code, dbt, Talend

## CERTIFICATIONS

- [AWS certified Solutions Architect Associate](#) – Amazon Web Services 10/2023 – 10/2026
- [AWS certified Developer Associate](#) – Amazon Web Services 05/2024 – 05/2027
- [Databricks certified Generative AI Fundamentals](#) – Databricks 07/2024 – 07/2026
- [Data Science Career Track Certificate](#) – 365 Data Science 11/2022 – lifetime

## EDUCATION

- Master of Science in Data Analytics Engineering (GPA 4), Northeastern University, Boston, USA** 09/2022 – 08/2024
- Bachelor of Technology in Electronics and Communications Engineering, West Bengal Univ. of Technology, India** 08/2009 – 06/2013

## PROJECTS

- [Speech Emotion Recognition \(SER\)](#) | MLOps, GCP, Docker, Vertex AI, CNN, RNN 01/2024 – 05/2024
- [Legal Aid RAG Chatbot using FAISS](#) | NLP, Transformers, RAG, Langchain, LLM 01/2024 – 04/2024