Soutrik Chowdhury

Machine Learning Engineer

soutrik1991@gmail.com 8101288487 Bangalore,India
4th April 1991 github.com/soutrik71
linkedin.com/in/soutrikchowdhury

OVERALL PROFILE

- Nearly 8 **years of experience** as a **Machine Learning Engineer**, excelling in both individual contributor (IC) roles and leading **Data Science** and **Machine Learning Engineering** teams.
- Expertise in **conceptualizing, building, and deploying robust machine learning solutions** that seamlessly integrate with complex business requirements.
- Proficient in **data pre-processing, feature engineering, model selection, and hyperparameter tuning**, with a strong focus on **large language models (LLMs), SLMs and prompt engineering**.
- Skilled in leading machine learning frameworks such as **TensorFlow**, **PyTorch**, and **Scikit-learn**, with emerging expertise in tools like **LangChain**.
- Extensive experience in managing **end-to-end ML pipelines** using tools like **DVC** for version control and **CML** for continuous model training and integration.
- Hands-on expertise in **AWS** services for scalable and secure cloud-based deployments, including infrastructure optimization and automation.
- Proficient in **distributed computing technologies** and leveraging **Azure Cloud Services** for building scalable solutions.
- Strong proficiency in **GitHub** workflows, fostering efficient collaboration and deployment.
- Proven ability to deploy large scale ML systems using FastApi , Docker and Kubernetes
- Demonstrates **critical thinking**, **problem-solving skills**, and **attention to detail**, with a commitment to staying ahead of emerging technologies and methodologies.
- Passionate about delivering **innovative machine learning solutions** using cutting-edge technologies to drive impactful business outcomes and advance the field of AI.

PROFESSIONAL EXPERIENCE

04/2023 – present
Bangalore, India

AB InBev
LLM Application Engineer

1. Product Digital Analyst

- **Developed an end-to-end NL to SQL pipeline** to empower Business Intelligence (BI) teams with efficient, accurate access to structured data using natural language queries.
- **Objective:** Simplify database interactions for **CXO-level reporting** and decision-making.
- **Leadership:** Led a team of Data Scientists and ML Engineers.

Key Contributions:

- Built a multi-layered system for NL to SQL conversion:
- Added a **complex query decomposition module**, breaking down intricate questions into sub-queries for independent processing.
- Designed **different routes for simple and complex queries** using a **fine-tuned BERT model** for query classification.
- Enhanced precision through a **fine-tuned BERT model for NER extraction**, enabling granular search in the Redis cache by identifying key entities in user queries.
- Optimized performance with a **Redis caching layer** for rapid query execution and cost reduction.
- Delivered an extensible, fault-tolerant and **microservice** based system that supports diverse queries and maintains session-based context for seamless follow-ups.

- Designed and deployed endpoints using **FastAPI** integrated with **Redis** and **RabitMQ**, hosted on a VM via **Docker** for scalable and efficient performance.

Outcome: Streamlined BI workflows with a scalable, robust solution for NL to SQL conversion, enabling fast, accurate data retrieval and actionable insights.
This condensed version includes all the critical technical contributions and key features while keeping it concise and impactful.

2. Product Digital Assistant

- **Developed an LLM-based application** to streamline information retrieval for Logistics and HR teams from unstructured data sources.
- **Leadership:** Led a team of Data Scientists and ML Engineers to deliver a scalable, efficient solution.
- **Key Contributions:**
  - **Document Classification:**
    - Implemented **YOLO-based page classification** to categorize documents as simple or complex.
    - Processed simple documents with a **regular parser** and complex ones with **GPT-4** for extracting tables and charts.
  - **Data Storage and Retrieval:**
    - Generated vector embeddings using **embedding models** and indexed them in **Redis Vector Database (RedisVL)** for hybrid search.
    - Used **ColBERT** for precise distance-based retrieval and re-ranking to enhance search relevance.
  - **Query Handling:**
    - Enabled **NER-based query processing** for granular searches and GPT-powered **chain summarization** for detailed responses.
    - Supported seamless follow-up queries through session history and contextual continuity.
  - **Scalability and Optimization:**
    - Accelerated embedding generation with the **Ray framework** and incorporated a **real-time feedback mechanism** to iteratively improve query accuracy.
    - Delivered an extensible, fault-tolerant and **microservice** based system that supports diverse queries and maintains session-based context for seamless follow-ups.
    - Designed and deployed endpoints using **FastAPI** integrated with **Redis** and **RabitMQ**, hosted on a VM via **Docker** for scalable and efficient performance.

Outcome: Delivered a robust, user-friendly system that significantly improved data access and decision-making efficiency for Logistics and HR teams.

3.**Agentic Framework** using LangGraph

- Developed an agentic framework leveraging LangGraph to enable multi-tool dynamic orchestration, tackling multi-hop questions and dynamic state management.
- Implemented adaptive workflows for query refinement and sub-task chaining across different api bound tools to ensure comprehensive answers.
- Ensured session-based context handling for seamless follow-ups and optimized execution for real-time performance.

10/2021 – 03/2023
Bangalore, India

AB InBev
ML Engineer
Duplicate Invoice Detection Engine

- **Led a high-performing team** to develop a robust solution for detecting duplicate invoices, reducing financial risks, and enhancing operational efficiency.

Key Contributions:

- Built an **advanced semantic/text-matching engine** to identify duplicate invoices, improving fraud prevention and financial integrity.
- Integrated a **machine learning layer** to enhance predictive accuracy and reduce false positives, streamlining invoice processing.
- Designed a scalable solution on **Azure ML Studio**, adhering to best coding and architectural practices.

- Implemented **DVC for data versioning**, **MLFlow for model lifecycle management**, and **Deepchecks/EvidentlyAI** for data quality assessments. Enhanced model interpretability with **Shapely framework**.
- Architected CI/CD pipelines on **Azure DevOps/GitHub Actions** for automated Azure ML pipeline deployment, accelerating delivery.
- Instituted **Azure Application Insights** for performance monitoring, enabling proactive bottleneck resolution.

Cashflow and PNL Forecast Framework

- **Designed a versatile forecasting framework** integrating models like **ARIMA**, **Prophet**, and **LSTM** for accurate predictions.

Key Components:

- Developed an **anomaly detection and correction framework** using **Prophet + ADTK** and **regularized ARIMA**, improving forecast reliability.
- Implemented **parallelized and vectorized computing** on **Databricks clusters** for scalable, high-performance forecasting.
- Architected a **temporal regression framework** leveraging **Light GBM** and **XGBoost**, addressing large-scale forecasting challenges with intelligent grouping and parallelized training.
- Automated feature extraction pipelines, optimizing precision for product clusters.
- Designed for seamless operation on **Azure ML Studio** or **Databricks**, enabling platform flexibility.
- Managed development and deployment workflows via **Azure DevOps** for efficiency and scalability.

Additional Skills and Experience

- Delivered multiple **POC projects** showcasing innovative solutions.
- Proficient in **Power BI** and **Tableau** for impactful data visualizations.
- Created an open-source framework (**Docker, GitHub Actions, MLFlow, PostgreSQL**) to mimic **Azure ML ecosystems** for on-premise clients.
- Developed **FastAPI templates** for quick plug-and-play analytics solutions.

11/2020 – 09/2021
Bangalore, India

o9 Solutions
Senior Data Scientist

- Implemented end-to-end Demand Planning and Sensing solutions, leveraging advanced statistical algorithms, machine learning algorithms, and deep neural networks
- Successfully executed multiple proof-of-concept (POC) projects aimed at addressing complex forecasting challenges,including hierarchical forecasting

06/2018 – 11/2020
Bangalore, India

Fractal Analytics, Bengaluru
Data Scientist

- Distinguished as a highly accomplished Data Scientist/Decision Scientist,possessing a wealth of expertise in effectively framing and solving intricate business problems.
- Proficiency extends to data manipulation and exploration, as well as developing cutting-edge data science and analytical algorithms and solutions.
- Furthermore, have excel in seamlessly deploying sophisticated analytical models into production environments, ensuring their optimal utilization and integration within complex operational systems

09/2015 – 12/2016
India

Life Insurance Corporation of India Ltd
Business Development Officer

07/2014 – 07/2015
Durgapur, India

Deys Power System (P) Ltd.
Junior Engineer Maintenance

🎓 EDUCATION

2017 – 2018
Kolkata, India

PGDBA: Data Science & Machine Learning
Praxis Business School

2010 – 2014 **Electrical Engineering**
Suri, India West Bengal University of Technology

SKILLS

- Python
- Docker
- SQL
- MLOps
- Gradio
- AI-Agents
- Pytorch
- FastAPI
- Github Actions
- Hugging Face
- Minikube
- Git
- LangChain
- LLM & SLM
- AWS
- EKS

CERTIFICATES

Deep Learning Specialization
by Coursera

LangChain: Chat with Your Data
by Deeplearning.ai

Docker Training Course for the Absolute Beginner
by Kode Kloud

EMLO
The School of AI

ChatGPT Prompt Engineering for Developers
by Deeplearning.ai

Functions, Tools and Agents with LangChain
by Deeplearning.ai

FastAPI in Hard Way
by Udemy

LangChain for LLM Application Development
by Deeplearning.ai

Github Actions

ERA-V2
The School of AI