Shamik Biswas

A Lead AI Engineer with expertise in NLP, Generative AI, LLM, GPT, ML, Project Management and Finance.

shamikb.work@gmail.com

+91 8431 908166

Portfolio

Technical Skills

Python, Machine Learning (ML), Deep Learning (DL), Pytorch, Natural Language Processing (NLP), Transformers (LLM), RLHF, ChatGPT, Llama 3, Claude, Mistral, Prompting, Generative AI, LangChain, LlamaIndex, Weaviate, Chroma, AWS, SQL, Linux

Experience

- ➢ *July 2024: Code and Theory (Lead AI Engineer)*
- ➢ *December 2020 – May 2024: London Stock Exchange Group (Sr. Data Scientist)*

  - ❖ *Project: Extractive Summarisation, ESG Recommender System, Automatic Speech Transcription and Financial Metrics Insights Engine*
    - Delivered projects worth up to an estimated $2 Million in yearly revenue
    - Performing the role of a product development manager, planning (Agile Methodology) and managing all application development requirements and presenting the development work/application to key business stakeholders and senior leadership
    - Building (Research Extensively, Design & Implement) end to end LLM/ML applications with Microservice architecture
    - Building applications with/without ChatGPT through prompt engineering
    - Built an extractive summarisation engine for financial reports for Investors, Investment Banks, Hedge funds, Private Equity etc
    - Built a ESG Recommender System to provide key insights on ESG focused companies and the same data was used to compute ESG scores for building ESG Indices such as FTSE Developed ESG Indices, FTSE Emerging ESG Indices, FTSE All-Share ESG Indices
    - Worked on Speech to Text for Automatic Speech Transcription for Corporate Disclosures such as Earnings Call, Corporate Conferences...
    - Built a customisable Financial Metrics Extractor engine to extract Key Financials from millions of documents and use them for financial insights
    - Built a custom annotation tool for annotating financial transcripts and annotated complex financial documents
    - Trained LLM/transformer models (BERT, RoBERTa, GPT, etc.) amongst other ML/DL models on AWS as well as on various resource configurations in WSL
    - Used AWS SageMaker to build and deploy transformer models
    - Used model compression techniques (PEFT, LoRA, QAT, GPTQ, QLoRA) to *Quantize* transformer models and reduced model size by converting them to *ONNX*
    - Used ONNX Runtime for faster inference
    - Containerised production model for deployment using Docker
    - Regularly contribute to technical documentation for a Wiki
    - Participating in knowledge sharing sessions to train other data scientists on contemporary technology
    - Mentoring and training fellow data scientists, following best practices without compromising quality, while adhering to challenging deadlines
    - Making key decisions in rolling out new value-added features for machine learning tools

- ➢ *April 2018 – December 2020: Kantar Analytics (Sr. Data Scientist)*

  - ❖ *Project: Client Knowledge Centre*
    - ❖ *Client:* **Volkswagen and Mazda**
    - Managing a project worth $1.2 Million/year in revenue

- Built Client Knowledge Centre to drive customer-centric decisions and leverage it for gaining insights and reporting
- Delivered several machine learning (ML) algorithms for processing Big Data and modelling on Python using Azure DevOps, which resulted in an increase of QoQ Sales and leveraged the said analysis to maximise ROI
- Automated reports as well as developed Big Data processes on Python, increasing efficiency by 80%
- Delivered weekly, monthly, and quarterly reports to help key decision makers (CXO) use it as dip stick read on business pulse, monthly course correction and budget setting respectively

❖ *Project: Lead Generation*
  ❖ *Client:* Eureka Forbes
- Lead the CES team to generate hot leads for cross and up sell opportunities for the existing customer base achieved through ML algorithms (XGB, Random Forest, MLP, Ridge/Lasso Regression, SGD, SVM)
- Identified potential customers to cross sell other products
- Improved campaign response and find out effective campaign strategies
- Identified key factors, which affect the propensity to up sell and maximised Customer Lifetime Value

➤ *April 2017 – April 2018: Tropical Research Services (Data Scientist)*

❖ *Project: Commodity Pricing in Spot and Futures*
  ❖ *Client:* Archer Daniels Midland, Bunge Limited, Cargill, Louis Dreyfus
- Scrutinised the Supply and Demand scenario of Agri commodities in Indian as well as global context to predict future trend of production, demand to derive the price direction and intensity of the price movement
- Developed country and state-wise balance sheet for sugar and coffee. Constant watch on weather development across the globe and its impact on the crop calendar
- Developed relationship with market intermediaries involved in value chain, constant interaction with clients for requirement gathering, improvising on research update, and cementing healthy relations with the clients resolving their business queries
- Prepared daily, weekly, and monthly reports providing outlook on commodity prices through fundamental research

Certifications

- Generative AI with Large Language Models (LLM)
- Building Systems with ChatGPT API
- ChatGPT Prompt Engineering for Developers
- Langchain for LLM Application Development
- Langchain building Chat Applications
- NLP with Classification and Vector Spaces
- Applied Machine Learning in Python
- Applied Text Mining in Python
- Machine Learning in Python with Scikit-learn
- Pytorch Essential Training
- Using Python for Web Services
- Six Sigma Green Belt
- PLENTY MORE IF INTERESTED

Education

➤ *March 2015-17 - MBA - IMT*
    Specialisation in Analytics & Strategic Management

➤ *December 2013 – B.Sc. & B.A (Integrated) – Sri Aurobindo International Centre of Education*
    Specialisation in Computer Science, Mathematics, Chemistry, English and French

Language Skills

➤ *Professional Level:* English & French (Writing/Speaking)
➤ *Intermediate Level:* German (Writing/Speaking)