Ankon Bhowmick

ankonbh@gmail.com, +919667055306
https://www.linkedin.com/in/ankon-bhowmick-2a8805193/
https://portfoliai-ankonbh.streamlit.app/
https://github.com/Ankon1708

Personal Summary

Highly motivated data scientist with a strong foundation in machine learning, data analytics, and NLP. As a recent graduate with an MSc in Data Science and Analytics at the University of Leeds, I bring expertise in exploratory data analysis and predictive modelling alongside being skilled in Python, R, and SQL. My analytical rigor and commitment to continuous learning equip me to deliver impactful insights and innovative solutions.

Education

1. **MSc Data Science and Analytics**, University of Leeds, (September 2023 - December 2024), 2:1 (UK)
2. **BTech Mathematics and Computing**, Delhi Technological University, (August 2019 – June 2023), CGPA – 8.33/10

Internships

1. **PwC**, Gurgaon, India, (June 2022 - August 2022)
   Intern/Trainee, Advisory- Technology Consulting
   - Collaborated with manager to perform an in-depth analysis of a substantial data dump containing over 50,000 SAP incident reports from a client company.
   - Implemented advanced data analysis techniques and Python programming to explore the data which contained 30 features, identifying both positive trends and potential areas of concern.
   - Deployed a decision tree model to predict successful incident resolution with an accuracy of 97.7%. Engineered modular code by organizing it into classes a dedicated prediction function for improved reusability.

Certifications

1. **Data Science Project-Based Learning and Internship Program Using R, Python Programming, And Tableau**, StepUp Analytics, Online, (December 2020 - February 2021)
2. **Microsoft Excel (Office 2019)**: Exam Reference #45320081

Skills

1. **Programming Languages**: Python, R, SQL
2. **Python Libraries**: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, OpenCV, PyTorch, TensorFlow
3. **R Libraries**: dplyr, ggplot2, caret, glm, rpart, timeSeries, rstanarm
4. **ML & DL Algorithms**: Linear Regression, Logistic Regression, KNN, Decision Trees, Naïve Bayes, K-Means, DBSCAN, Random Forest, PCA, Multi-level Regression, ANN, CNN, RNN, LSTM, Transformers, ARIMA
5. **NLP**: Text Preprocessing (NLTK), Embeddings, HuggingFace (Transformers, Datasets, PEFT), Vector Database (FAISS), Fine-tuning (BERT, T5, LoRA), GenAI (LangChain)
6. **MLOps**: Git, HuggingFace Hub, Weights & Biases, LangSmith

Academic Projects

1. **Predicting the Course of the Russo-Ukrainian War (Bachelor's Dissertation)**
   - Fine-tuned a BERT model on 1300+ news articles from the 2008 Russo-Georgian War to classify war intensity, and generated war-intensity predictions on 32,000+ articles from the Russo-Ukrainian war to create a timeline for analysing Russian war actions.
   - Python, Web Scraping, Selenium, Beautiful Soup, NLP, BERT, Hugging Face Transformers, PyTorch, Tensor Datasets, Text Classification, Data Preprocessing, CUDA, Data Visualisation
2. **Fine-tuning Flan-T5 for Article Title Generation**
   - Implemented LoRA to efficiently fine-tune Google's Flan-T5-base on 32,000+ samples for article title generation, achieving ~18% relative improvement in ROUGE-1 and ROUGE-L scores over pre-trained model, while maintaining near-identical performance (~1 point difference in ROUGE) compared to full fine-tuning.
   - Python, FLAN-T5, LLM, Fine-tuning, NLP, Seq2Seq, Hugging Face (Transformers, Datasets, PEFT, Hub), PyTorch, Pandas, NumPy, wandb, CUDA, Encoder-Decoder Transformer, NLTK, Kaggle
3. **Portfolio Query Assistant using RAG**
   - Deployed a retrieval-augmented generation (RAG) system using LangChain to process and query documents with Google GenAI (Gemini 2.5 Flash) and FAISS, and LangSmith for observability & evaluation.
   - Keywords: LangChain, RAG, Document Loaders, RecursiveCharacterTextSplitter, FAISS, Semantic Search, Vector Database, Python, Streamlit, LCEL, Google Gemini, LangSmith, HyDE, Multi-query search