Arnab Manna

LinkedIn | GitHub | Hackerrank

Location: Kolkata, West Bengal, India
Email: mannaarnab2001@gmail.com | Mobile: +917063039253

SUMMARY

I am a skillful **Data Scientist with 1.5 years of experience in different analytics domain**. Proficient in **data analysis**, skilled in **high-volume data cleaning**, and adept at **optimizing SQL queries.** Experienced in using **AWS and Azure cloud services**. Strong programming skills in **Python,** with expertise in **statistical modeling and machine learning.** Seeking opportunities to leverage my expertise and skills to make a meaningful impact.

EXPERIENCE

Associate AI Engineer   Nov 2023 – Present
DeepQ-AI Buisness Solutions   *Remote - Bengaluru, Karnataka, India*

- Developed a **handwritten bank cheque parser** with a custom model in **Azure Document Intelligence.**
- Collaborated with domain experts to create a synthetic dataset for credit card customers.
- Designed prebuilt templates for **customer churn, segmentation, and fraud detection**, reducing onboarding effort by **80%.**
- Created a topic modeling template using **LDA** and the advanced **BerTopic** framework.
- Built a Q&A chatbot with **GPT-3.5 turbo** for structured datasets, capable of generating code, algorithms, and graphs.
- Created a Q&A chatbot on **FAQ fianancial document** using **RAG** architecture with LangChain framework.
- Explored **AI Agents** and developed an advanced agent which can interact with pdf with **PhiData** framework.

Data Scientist   Mar 2023 – Aug 2023
AarogyaAI Innovations Pvt. Ltd.   *Remote – Bangaluru, Karnataka, India*

- Improved performance of ML model by **6.5%** from the previous person in validating MTB drugs.
- Generated **30k** vcf files for MTB drugs using **AWS Batch processing** and **reduced cost and time by 5%, 48 hours**.
- Developed **Hierarchical clustering** model to cluster out lineage of covid microorganism.
- Created checkpoints and workflows to hand over the pathogen pipeline to Bio-info and R&D team.

EDUCATION

Pondicherry University   Kalapet, Puducherry, India
Master of Science in Statistics   *Jun 2021 – May 2023*
  * Cumulative GPA : **8.27/10.0**

Haldia Government College   Haldia, West Bengal, India
Bachelor of Science in Statistics   *Jan 2018 – Apr 2021*
  * Cumulative GPA : **8.17/10.0**

SKILLS & COURSE WORK

- **Course Work**: Data Mining, Database Management, Data Visualization, Time Series Analysis, Regression Analysis, Estimation and Testing of Hypothesis
- **Programming Languages and Databases**: Python, R, C, MySQL, MongoDB
- **Cloud Services**: AWS, Azure
- **Python Libraries:** Numpy, Pandas, Scikit-Learn, Seaborn, Matplotlib, Plotly, Optuna, BerTopic, Tensorflow, Keras, nltk, spacy, PyPDF2, PyMuPDF, Evidently
- **Machine Learning:** Linear Regression, Logistic Regression, Random forest, PCA, LDA, AdaBoost, XGBoost
- **NLP:** Text data preprocessing, Word2Vec, TF-IDF, Tokenization, Word Embeddings, BERT and Transformer
- **Gen AI:** GPT-3, GPT-3.5 turbo, Llama2-7B, Mistral-7B, RAG, Indexer, Prompt Engineering, LangChain, Azure Prompt Flow, Groq, AI-Agents, PhiData, Ollama

CERTIFICATIONS

- Microsoft Excel for Data & Buisness Analytics by Udemy
- Python for Data Science and Machine Learning Bootcamp by Udemy
- The Complete SQL Bootcamp: Go from Zero to Hero by Udemy.