Samarjit Debnath

Software Engineer
Agartala, Tripura, India
contactsamarjit100@gmail.com +91-8415049068 LinkedIn GitHub

Experience

Optum India
DECEMBER 2021 - PRESENT, NOIDA, UTTAR PRADESH, INDIA

Data Scientist (Machine Learning Engineer)
MARCH 2024 - PRESENT, NOIDA, UTTAR PRADESH, INDIA

- Architected a model score file profiler tool that automated processing using Airflow, Azure Databricks, and Spark, reducing manual time from 1-2 weeks to under 15 minutes. The tool triggers emails for discrepancies and stores output in Snowflake, enabling a scalable, low-latency production launch.
- Refactored the Disease Severity rule-based model using Factory and Null Object design patterns for five sub-models, with easy scalability for more. This modernization improved maintainability, boosted runtime by 45%, and optimized space by streamlining intermediate data writes.
- Crafted a scalable pipeline integrating clinical records with claims data for model evaluation, resulting in a 2.5-3% increase in member identification for disease event predictions.

Software Engineer (Machine Learning Engineer)
DECEMBER 2021 - MARCH 2024, NOIDA, UTTAR PRADESH, INDIA

- Constructed data-model pipelines with Spark, Azure Databricks, and Apache Airflow using the unifAI framework, achieving 98.5% accuracy in reproducing historical ML runs compared to legacy systems.
- Provided input on data engineering and feature engineering, cleansing data in monthly and weekly pipelines, and extracting features from over 200 million healthcare records.
- Developed a tree-based ML model using CatBoost, predicting emergency room usage within 6 months, resulting in $2M cost savings per quarter.
- Developed an NLP solution for address matching with a 90% similarity threshold, and utilized an LLM to fill in missing entries across state and global tables.
- Contributed to developing retrospective validation tools for machine learning models, handling over 150 million records.

Kinara, Inc. / Software Engineer
JUNE 2021 - OCTOBER 2021, HYDERABAD, TELANGANA, INDIA

- Contributed to the software tools design team for the ARA-1 chip, including developing utilities for memory read/write speed testing and application software for measuring Neural Network Processor performance.
- Engineered a System API trace tool that monitors API calls and tracks their execution status.
- Authored multiple test scripts using Python3 for the test suite, supporting internal and production releases.
- Optimized and debugged existing software, adding new features such as plugins and utility tools to enhance both legacy and newly created applications.

Academics

National Institute of Technology, Agartala / B. Tech (Electronics and Communication Engineering)
JULY 2017 - JUNE 2021, AGARTALA, TRIPURA, INDIA

Skills

- **Programming Languages**: C/C++, Python3
- **Tools & Packages:**
  - **Data Engineering and AI/ML:** Big Data (Spark, Hadoop, HDFS), Azure Databricks, ETL, Machine Learning (Regression, Classification, Clustering, Tree-based Algorithms), Deep Learning (ANN, CNN, RNN, LSTM, BERT), Numpy, Pandas, Scikit-Learn, Tensorflow, NLP Fundamental, LLM Tools (LangChain, Gemini), MLflow
  - **Software Development and Deployment:** FastAPI, Streamlit, Object-Oriented Design, GitHub, Docker, Jenkins, Airflow, Ngrok
- **Database**: SQL, Snowflake, Vector Database Fundamentals (ChromaDB)
- **IDE & Code Editors**: VS Code, Vim, Anaconda Jupyter, PyCharm
- **Environments**: Linux - Ubuntu/CentOS, macOS, Windows

Project

- **Intelligent Query Engine [Link]**
  A Modern AI-Powered Query System
  - Developed an LLM-based question-answering system for the Tees store, utilizing Google Gemini, Hugging Face embeddings, and ChromaDB for efficient vector storage and retrieval.
  - Integrated LangChain framework to query MySQL inventory, sales, and discount data, delivering accurate and context-aware results.