

Rituparna Sharma

Email: sharmarituparna66@gmail.com

LinkedIn: [linkedin.com/in/rituparna-sharma](https://www.linkedin.com/in/rituparna-sharma)

Github: github.com/RituparnaSharma

Phone: +91-7002597915

Bengaluru, KA

Summary

Data Scientist/Engineer with 2.5 years in data analysis, ML, DL, and distributed computing, skilled in Python, PySpark, Docker, FastAPI, and AWS. Proven success in enhancing data quality, streamlining processes, and deploying scalable solutions with Terraform and AWS. Proficient in NLP and advanced techniques, eager to boost innovation and efficiency in data-driven projects.

Experience

Data Scientist

Lentra.ai

Dec 2022 - Present (1 year 3 months)

- Enhanced data quality for model training with Snorkel, showcasing data preprocessing skills.
- Utilized Hugging Face Sentence Transformers, YubiBERT, and LLM API for precision in financial data analysis.
- Led the creation of a system for classifying bank transactions, enhancing efficiency and accuracy. Developed scalable Python scripts, utilized deep learning for improved classification, and employed agglomerative clustering for customer segmentation, driving better marketing and customer service.
- Managed codebases with Git and Bitbucket, ensuring robust version control and collaboration.
- Streamlined deployment with Docker and developed REST API via FastAPI for enhanced model integration.
- Deployed and maintained ML models with AWS SageMaker, addressing data and model drift for optimal performance.

Data Engineer

Lentra.ai

Aug 2021 - Dec 2022 (1 year 5 months)

- Transformed a centralized Python application into a distributed system with PySpark, slashing processing times from hours to seconds.
- Utilized Sqoop for data ingestion into Hive tables, significantly enhancing operational efficiency.
- Developed modular Terraform scripts for efficient infrastructure management, streamlining deployment processes and enhancing system scalability.
- Utilized and applied AWS services (EC2, EMR, Airflow, S3, Lambda) to optimize data processing and storage solutions.
- Implemented ETL pipelines using AWS Glue for data transformation and incorporated AWS Athena to analyze the transformed data. Integrated with AWS Redshift for distributed data processing.

Data Scientist Intern

Lentra.ai

Feb 2021 - Aug 2021 (7 months)

- Utilized fuzzy matching and cosine similarity algorithms within a PySpark framework to meticulously process millions of data points.
- A recognition that accelerated my proficiency in PySpark and distributed computing.

Personal Projects

SmartPal

- Developed a specialized chatbot using Gemini Pro LLM and Gemini embeddings, ensuring precise and context-aware responses by employing ChromaDB for the strategic storage of personal professional knowledge.
- Implemented a tailored mini RAG model through LangChain, enhancing the bot's ability to handle diverse queries with customized information retrieval chains, elevating the accuracy and relevance of its responses in my professional field.

Text Summarization of Article

- This project is used to build a model on sentiment summation. This model uses the Hugging Face Transformer library to run extractive summaries. The advantage of using this model is the computation time and performance due to its smart batching technique.

Skills

- **Programming Languages:** Python, PySpark
- **Frameworks and tools:** Scikit-learn, Keras, Tensorflow, Langchain, Streamlit, Flask, FastAPI
- **Vector database:** Chromadb, FAISS
- **DevOps & Tools:** Docker, Git
- **Deployment platforms:** MLflow, AWS
- **AI/ML Techniques:** Finetuning with custom data, vector embedding, NLP, Neural network, Optimization, MLOPS.

Publications

A Survey on Twitter Sentiment Analysis

International Journal of Computer Sciences and Engineering

Awards and Achievements

- [Spot Award for Excellence](#), Lentra.ai
- Winner of Sentiment Analysis Hackathon
- Winner of Credit Consumption Prediction Hackathon

Education



GreyAtom

Data Science Master's program

2019 - 2020



Assam Science and Technology University - ASTU

Bachelor of Technology, Computer Science and Engineering

2015 - 2019

