

Ashish Agarwal

Graduate Research Assistant @ SAIL LAB in UNH || **Data Engineer in UNH**
(203) 781-6025 : agarwal.ashish.singhal@gmail.com || <https://linkedin.com/in/agarwalashishsinghal/>
<https://CRLannister.github.io/> || <https://github.com/CRLannister>

SUMMARY

Passionate data science graduate student looking to advance understanding and expertise. Technically proficient in Statistical Analysis, Hypothesis testing, Descriptive statistics, Text processing, Sentiment analysis, A/B Testing, Anomaly detection, Time Series Analysis, Hyperparameter tuning, Regression analysis, Supervised learning, Unsupervised learning, Reinforcement learning, Deep Learning, Natural Language Processing, LLM RAG, Python, R, Pandas, Polars, Dask, Numpy, Tensorflow, Scikit-Learn, SQL, NoSQL, Excel, AWS Lambda, Redshift, Kafka, Azure, git, Docker, S3, EC2, Airflow, Django, Colab, Bash, Linux, MLops, LLMops, MasterControl, Netsuite, Matplotlib, Seaborn, Plotly, Altair, Power BI, QLIK, Tableau and MicroStrategy.

EDUCATION

Master of Science, Data Science (University of New Haven) GPA: 4.0 [West Haven, CT :- Aug 2023 – Expected Dec 2024]
Bachelor's in computer engineering (Institute of Engineering) GPA: 3.2 [Pulchowk, Lalitpur, Nepal :- Nov 2016 – Sep 2021]

RELATED EXPERIENCE

Data Engineer Intern (North East Scientific) [Waterbury, CT :- June 2024 – Aug 2024]

- Led the development of data pipelines using Mastercontrol and Netsuite APIs, automating data extraction, transformation, Validation and storage in MySQL database and Excel Files, reducing external dependencies.
- Developed dashboards and reports for Inventory, Sales and Production insights using QLIK Sense Cloud.
- Developed a fully operational Retrieval-Augmented Generation (RAG) system, leveraging open-source LLM models for document processing. Created comprehensive knowledge bases to streamline new employee training, enhance manufacturing practices, and support efficient information retrieval.

Graduate Research Assistant (Secure and Assured Intelligent Learning Lab) [West Haven, CT :- March 2024 – Present]

- Acquisition and analysis of electroencephalographic (EEG) signals in brain-computer interface (BCI) systems.
- Conducted statistical analysis employing advanced techniques such as time-frequency analysis via wavelet transforms, blind source separation using independent component analysis and canonical correlation analysis, signal denoising with empirical mode decomposition, and machine learning algorithms like SVM and deep neural networks for pattern recognition and classification.
- Setting up and configuring essential resources, including GPU acceleration, Docker containerization, databases for data management, Datalore and WordPress websites for web development and deployment.

Data Engineer (University of New Haven) [West Haven, CT :- Oct 2023 – Present]

- Engineered ETL/ELT pipelines on AWS utilizing Lambda, EC2, S3, Redshift, CloudWatch, integrating with PowerBI.
- Implemented OCR for image data feature extraction, fine-tuned deep learning models leveraging Hugging Face libraries.

Associate Software Engineer (LIS Yomari) [Lalitpur, Nepal :- April 2021 – May 2022]

- Migrated on-premises data warehouses to AWS, implementing secure data transfer, infrastructure automation with scalable ETL/ELT pipelines using Apache Spark, Kafka, and AWS services like Batch, Lambda, Kinesis, Airflow, S3, Athena and Redshift.
- Implemented advanced data modeling techniques as star/snowflake schemas, changing dimensions, and data partitioning.
- Optimized performance through caching, partitioning, indexing and developed robust MicroStrategy reports and dashboards.

Data Science Intern (Tootle) [Lalitpur, Nepal :- Jan 2021 – March 2021]

- Conducted research on ride-sharing platforms in Nepal, developing LRFM algorithms and customer segmentation models to analyze user behavior and identify target segments.
- Developed real-time dashboards for data visualization, leveraging MongoDB for NoSQL data storage, Django for web development, and Python for data processing and analysis.

RELEVANT PROJECTS

RAG-based Knowledge Management System for Training and Information Retrieval

Waterbury, CT

North East Scientific

June 2024 – Aug 2024

- Developed a Retrieval-Augmented Generation (RAG) system by scraping and cleaning data from the company's website and specification documents, converting them into markdown format.
- Built a Chroma database incorporating LLM models such as Llama3, Phi3, and Qwen2 with quantization techniques to optimize GPU usage.
- Employed advanced prompt engineering to ensure accurate, non-hallucinated responses by reranking documents and interfacing with the database.
- Designed a user-friendly web interface, similar to ChatGPT or OpenWebUI, enabling user context tracking and delivering reliable, query-specific responses for training and operational purposes.

Retail Sales Inventory and Traffic ETL and Reporting

Lalitpur, Nepal

LIS Yomari [Client- Ralph Lauren]

July 2021 – March 2022

- Architected a robust ETL/ELT pipeline on AWS, ingesting data from TrueVUE APIs, S3 buckets, GCP, and FTP servers, leveraging EC2, EMR, Kinesis, Glue, Lambda, S3, Airflow, Redshift, CloudWatch, CloudFormation, and IAM.
- Ingested streaming data, processed and transformed it, orchestrated workflows, stored data in S3 and Redshift, implemented monitoring, logging, and automated infrastructure provisioning.
- Developed metadata objects, metrics, and attributes based on specifications in MicroStrategy, conducted data validation and performance testing, designed dashboards and dossiers tailored to client requirements.

EXTRACURRICULAR ACTIVITIES

Helping Hands Nepal, HENN

Kathmandu, Nepal

Volunteer and Technical Supervisor

May 2017 – July 2019

- Implemented Linux-based low-resource distros in remote schools of Nepal through HENN initiative.
- Created tailored educational materials in Nepali language and distributed them through locally hosted web pages.