Sugam Mishra

📞 +1-6055928649 | ✉ mishrasugam11@gmail.com

🔗 https://www.linkedin.com/in/sugam-mishra-684b521b7/ | ⌨ github.com/infinitysugam

PROFESSIONAL SUMMARY

Data and AI Engineer with **3+ years** of experience building scalable data pipelines and automating financial and marketing analytics workflows. Specialized in **big data systems** using Apache Spark, HDFS, and AWS. Currently pursuing a Master's in Computer Science with research in **Graph RAG** and **video comprehension**.

EDUCATION

Master of Computer Science   January 2024 – Present
GPA 4.0   *South Dakota State University,Brookings,South Dakota*

Bachelor of Computer Science Engineering   June 2017 – May 2021
GPA 8.9/10   *KIIT University,Bhubaneswar,India*

EXPERIENCE

Student Programmer   May 2024 – August 2024
South Dakota State University   *Brookings, South Dakota*

- Developed interactive websites using **C# and the .NET framework**, showcasing proficiency in front-end and back-end development for college projects.
- Created a range of console applications and interactive forms to support various functionalities, enhancing user experience and application usability.
- **Tools: .NET, C#, API**

Graduate Teaching Assistant   January 2024 – Present
South Dakota State University   *Brookings, South Dakota*

- Guided and mentored 60+ undergraduate students in **C and C++ programming**, helping improve lab assignment quality and conceptual understanding. Assisted with grading and held weekly office hours to support student success.

Cloud Data Engineer   August 2023 – November 2023
GrowByData Pvt LTD   *Kathmandu, Nepal*

- Designed and implemented a **data warehouse solution** to store and manage marketing data across **100+ client accounts**, and **automated reporting dashboards** that replaced manual Excel-based workflows, reducing reporting time from **2 days to under 15 minutes** and improving accuracy and scalability for marketing analytics.
- Developed automated data extraction scripts for **10+ advertisement platforms** for **100+ clients** and syncing **hundreds of marketing campaigns daily**, enabling near real-time performance insights and reducing manual integration effort by over **90%**.
- **Tools: AWS, Distributed Processing, API, Python**

Big Data Engineer   May 2021 – August 2023
F1Soft Pvt LTD   *Lalitpur, Nepal*

- Led the design and implementation of a robust data architecture, building a **centralized data lake and data warehouse** for **9 distinct financial verticals**, enabling seamless handling of **large-scale transactional data** and supporting cross-domain analytics and reporting.
- Redesigned the legacy daily ETL pipeline that processed **1 billion+ financial transactions**, reducing processing time from **8 hours to 10 minutes** by migrating to a Big Data architecture using **HDFS** for storage and **Apache Spark** for distributed processing.
- Built a scalable system to process transactional data from **each bank branch** and **automated financial dashboards** like the **Balance Sheet** and **Profit & Loss Statement**, replacing a manual **Excel-based workflow** that previously took **2–3 weeks** and **5+ employees** to complete, reducing reporting time to **under 1 hour** and significantly improving accuracy and efficiency.
- **Tools: Apache Spark, Database, Apache Airflow,Hadoop**

Research Work

Video Comprehension   Ongoing
- Working on enhancing the DREAM1k dataset for improved model training in video comprehension.
- Developing innovative methods to evaluate model outputs, ensuring accuracy and relevance.
- Building a "chain of thought" architecture to improve model interpretability and performance..

Semantic-Aware Community Detection in Graph RAG   Ongoing
- Researching **semantic-aware community detection** to improve retrieval in Graph-based Retrieval-Augmented Generation (Graph RAG) systems.
- Developing techniques to represent knowledge graph nodes in **semantic vector space** using embeddings
- Implementing dynamic neighborhood expansion, pruning, and path-based prompting to enhance LLM-based question answering.

Skills

Programming Languages: Python, Scala, SQL, Bash
Big Data & Processing: Apache Spark, Apache Kafka, Hadoop, Airflow, ETL
Databases: MySQL, MongoDB, Database Modeling
Cloud & Tools: AWS, API Integration, Automation
Machine Learning & AI: Machine Learning, Artificial Intelligence, Knowledge Graphs

Volunteer Work

Rotaract Club   August 2021 – Present
General Member   *Lalitpur, Nepal*
- ∗ Organized and led multiple events, including fundraising and leadership initiatives to support community projects.
- ∗ Coordinated efforts to provide essential supplies to local schools, benefiting under-resourced students.
- ∗ Spearheaded environmental campaigns, promoting sustainability and awareness in the community.