

# Sugam Mishra

☎ +1-6055928649 | ✉ mishrasugam11@gmail.com | 🌐 sugammishra.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

GPA 4.0

January 2024 – Present

South Dakota State University, Brookings, South Dakota

### Bachelor of Computer Science Engineering

GPA 8.9/10

June 2017 – May 2021

KIIT University, Bhubaneswar, India

## EXPERIENCE

### Student Programmer

South Dakota State University

May 2024 – August 2024

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

South Dakota State University

January 2024 – Present

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

GrowByData Pvt LTD

August 2023 – November 2023

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

F1Soft Pvt LTD

May 2021 – August 2023

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 Score (VCS)

Multimodal Intelligence Lab, South Dakota State University

- Developed the **Video Comprehension Score (VCS)** benchmark framework for evaluating multimodal video understanding using novel alignment metrics (**NAS**, **CAS**, **SAS**).
- Published the VCS toolkit and dataset to enable standardized evaluation of video-language model alignment.
- [Project Website: multimodal-intelligence-lab.github.io/Video-Comprehension-Score](https://multimodal-intelligence-lab.github.io/Video-Comprehension-Score)

### CLIP-CC Bench

Conference Submission — WACV 2026

- Co-authored a **benchmark dataset and evaluation suite** for assessing paragraph-level video description capabilities in Vision-Language Models (VLMs).
- Developed the **CLIP-CC Bench** dataset of 200 curated 90-second movie clips paired with expert-written, paragraph-length ground-truth descriptions.
- Proposed a **three-tier evaluation framework** combining N-gram overlap metrics, embedding-based semantic similarity, and **LLM-as-judge** assessment using the ACCR rubric (Accuracy, Completeness, Conciseness, Relevance).
- Benchmarked 14 state-of-the-art VLMs, revealing how **architecture and training design** affect discourse-level understanding more than model scale.
- Released standardized evaluation scripts, model outputs, and aggregation tools for **reproducibility and open research**.

### Semantic-Aware Community Detection in Graph RAG

Ongoing

- Developing a **semantic-aware community detection** framework to enhance retrieval and reasoning in Graph-based Retrieval-Augmented Generation (Graph RAG) systems.
- Designing an **edge scoring algorithm** that weights relations based on query-relation similarity, node context alignment, and hop decay to construct meaningful semantic subgraphs.

## PROJECTS

---

### Graph RAG Supply Chain Forecasting System

- Developed a **supply chain demand forecasting system** combining **Machine Learning**, **Large Language Models (GPT-3.5)**, and **Graph RAG** for contextual predictions with natural language explanations.

### Bitcoin Dashboard

- Developed a full-stack web application for managing and monitoring Bitcoin Core nodes with real-time blockchain data visualization and multi-wallet support
- Implemented block explorer functionality with transaction lookup, mempool monitoring, and node health tracking displaying network statistics and synchronization progress

### Stock Trading Platform

- Built full-stack trading platform with real-time market data streaming via WebSockets, automated order execution, and portfolio management
- Developed modular architecture with order management, risk management, and news aggregation modules integrated with OANDA API
- Implemented real-time risk controls with stop-loss automation and position sizing; containerized with Docker for scalable deployment

## 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.