

SUMIT KUMAR

Data Engineer

+91-7549980508

sumit749284@gmail.com

linkedin.com/in/smaxiso

github.com/smaxiso

Education

B.Tech. in Computer Science & Engineering
National Institute of Technology Patna (CGPA: 8)

2017 – 2021
Patna, Bihar

Experience

Tata Consultancy Services

Mar 2023 – Present

Data Engineer

Bangalore, India

- Developed a scalable ETL Framework for Data Migration for a leading global online payments company using Python, AWS, GCS, and BigQuery.
- Reduced data migration time by 20%, improving scalability by 30%.
- Technologies used: Python, AWS, GCS, BigQuery

Tata Consultancy Services

Jan 2024 – May 2024

Data Engineer

Bangalore, India

- Implemented optimizations in the Lynx Framework, resulting in a 35% improvement in data linkage accuracy and efficiency.
- Optimized LSH (Locality-Sensitive Hashing) algorithm, reducing approximate nearest neighbor search time by 40%
- Streamlined feature aggregation pipelines, reducing processing latency by 25%.
- Technologies used: PySpark, Scala, APSS (All Pair Similarity Search), BigQuery, GCP (Dataproc, GCS), LSH

Tata Consultancy Services

Aug 2021 – Jan 2023

Data Engineer

Bangalore, India

- Built on-demand merchant reports, increasing data accuracy by 15%.
- Decreased report generation time by 25%.
- Contributed to the Argo Framework for report generation.
- Technologies used: Python, SQL, Apache Spark, GCP

NIT Patna

May 2020 – July 2020

Data Scientist Research Intern

Patna, India

- Developed a real-time forest fire detection system using Python-based machine learning algorithms and fuzzy logic
- Achieved an accuracy rate of 90% in predicting the likelihood and severity of forest fires.
- Technologies used: Python, machine learning, fuzzy logic

Projects

Joint Image Compression & Encryption | Python, Digital Image Processing, JPEG2000, RC4, PIL Library

- Created an algorithm for joint image compression and encryption using lossless JPEG2000 and RC4 encryption.
- Achieved a compression ratio of 5.2, 99.69% NPCR, and 47.63% UACI in processed images.

Digital Image Compression | Python, Machine Learning, Digital Image Processing, K-means, PCA, PIL Library

- Designed an algorithm for digital image compression using K-means clustering and PCA.
- Achieved a compression ratio of 2.8 with 55-70% compression and a PSNR of 30 and above.

Technical Skills

Programming Languages: Python, C++, C, Java, Shell/Bash

Databases: MySQL, BigQuery, Oracle

Frameworks: PySpark, Apache Spark, Django, React

Developer Tools: Git, GitHub, Jupyter Notebook, CI/CD, Jenkins, Airflow

Concepts: ETL, Data Migration, Data Warehousing, Real-time Data Processing, Data Analytics, Cloud Computing, Machine Learning, Unix Systems, Generative AI, Agile Methodology, HDFS