SUMIT KUMAR

Data Engineer

+91-7549980508 sumit749284@gmail.com linkedin.com/in/smaxiso github.com/smaxiso

Education

B.Tech. in Computer Science and Engineering 2017 – 2021

National Institute of Technology Patna (CGPA: 7.79) *Patna, Bihar*

- **Relevant Coursework:** Operating System, Computer Networks, Object-Oriented Programming, Database Management System, Data Structures & Algorithms

Experience

Tata Consultancy Services Mar 2023 – Present

Data Engineer *Bangalore, India*

- Developed a scalable Data Migration Framework 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 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 Argo Framework for report generation.
- Technologies used: Python, SQL, Apache Spark, GCP.

NIT Patna May 2020 – July 2020

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.

Image Classification using CNNs | *Python, Machine Learning, Digital Image Processing, CNN*

- Led a team of 4 in developing a machine learning-based CNN model for image classification, utilizing the CIFAR10 dataset.
- Achieved a high accuracy rate of 87.44% on the training dataset and 82.5% on the testing dataset.

YouTube Spam Comment Filter | *Python, Machine Learning, Binary Classification*

- Developed and implemented a machine learning algorithm for binary classification of spam and non-spam comments on YouTube videos.
- Achieved an accuracy rate of 96.21% through training on multiple datasets.

Technical Skills

Programming Languages: C, C++, Python, Java

Databases: MySQL, BigQuery

Developer Tools: Git, GitHub, IntelliJ IDEA, VS Code, Jupyter Notebook, PyCharm, GCP, HDFS, ETL, CI/CD, Jenkins

Concepts: Data Engineering, Software Development, Machine Learning, Cloud Computing, Data Migration, ETL Processes, Data Warehousing, Real-time Data Processing, Agile Methodology, Big Data Analytics