

# Soham Sharma

+91-8580612408 | sharma.soham04@gmail.com | [LinkedIn](#) | [GitHub/sxham04](#) | [Kaggle/sxham04](#)

---

## TECHNICAL SKILLS

Programming Languages: Python (Advanced), Java, SQL, R, Django  
ML/AI/Data Science: Scikit-learn, Pandas, NumPy, Keras, OpenCV, NLP  
Tools & Platforms: Git, Docker, Jupyter Notebooks

---

## PROJECTS

### P2P File Transfer System with Ring-Topology Swarming | [\[GitHub Link\]](#)

Technologies Used: Java (Sockets, Multithreading), NIO, TCP/IP, Object Serialization

- Developed a **distributed system using Java Sockets** to partition files into 100KB chunks, enabling decentralized file sharing across multiple concurrent nodes.
- Implemented a **Ring-Topology Swarm** with multithreaded upload/download handlers, allowing peers to exchange missing data segments and reconstruct files autonomously.

### CPU Scheduling Simulator | [\[GitHub Link\]](#)

Technologies Used: Java (OOP), Data Structures (Lists, Queues), Swing (GUI)

- Engineered a **Java-based simulation engine** for 6 core CPU scheduling algorithms (FCFS, SJF, SRT, Round Robin, and Priority), providing precise calculations for waiting and turnaround times.
- Developed a **timeline-event system** to generate Gantt chart data and integrated a Java Swing GUI to visualize real-time process execution and performance metrics.

### Twitter Sentiment Analysis Engine | [\[GitHub Link\]](#) | [\[Kaggle link\]](#)

Technologies Used: Python, Scikit-learn, NLTK, TF-IDF Vectorization, Matplotlib/Seaborn

- Developed a **comprehensive NLP pipeline** using NLTK for text normalization (tokenization, stopword removal, lemmatization) and TF-IDF vectorization to process and classify 75,000+ tweets into four sentiment categories.
- Benchmarked **5 machine learning models** (Logistic Regression, Naive Bayes, Decision Tree, Random Forest, and SVM), achieving a peak accuracy of **94.6%** with Random Forest through hyperparameter analysis and performance evaluation.

### Customer Segmentation Engine | [\[Kaggle Link\]](#)

Technologies Used: Python, Scikit-learn, PCA, Pandas, Matplotlib/Seaborn

- Engineered an **unsupervised learning pipeline** to segment 8,900+ credit card users into distinct behavioral profiles by implementing median imputation, log transformations, and standard scaling.
- Optimized clustering performance using PCA for dimensionality reduction and the K-Means algorithm, identifying the ideal cluster count via the Elbow Method to reveal actionable spending and credit usage patterns.

---

## EDUCATION

DIT University | Dehradun, Uttarakhand

Bachelor's of Technology (CSE) | 2023-2027 | CGPA : 8.57

---

## CERTIFICATIONS AND AWARDS

- AI Foundations Certificate | Issuer: Oracle Learning
- Deep Learning Specialization (5 Courses) | Issuer: Coursera (DeepLearning.AI)
- Participated in youthopia hackathon, organised by the college