

## UTKARSH SHARMA

India | +91-8865035702 | [uditsharma1508@gmail.com](mailto:uditsharma1508@gmail.com)

LinkedIn: [linkedin.com/in/utkarsh-sharma-8b0365238](https://www.linkedin.com/in/utkarsh-sharma-8b0365238)

GitHub: [github.com/utkarshsharma2002us](https://github.com/utkarshsharma2002us)

Portfolio: <https://utkarshsharma2002us.github.io/>

## OBJECTIVE

MCA Student (2026) with a strong foundation in computer applications and a passion for software development, AI/ML, and Data Science. Skilled in **AI/ML, Python, MySQL**, Power BI, Tableau and Data visualisation. Open to entry-level opportunities in backend, AI/ML, Data Analyst and Data Science.

## EDUCATION

- Master of Computer Applications (MCA) | 2024-26 | Ajay Kumar Garg Engineering College, Ghaziabad, UP Affiliated to Dr. A.P.J. Abdul Kalam Technical University (AKTU)
- Bachelors of Computer Applications (BCA) | 2021-24 | Meerut Institute of Technology, Meerut, UP Affiliated to Chaudhary Charan Singh University (CCSU)

## SKILLS

- Programming Languages: Python, Java, SQL (MySQL)
- Data Science & ML: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn
- Generative AI & NLP: LLMs, RAG, Prompt Engineering
- Tools & Platforms: Git, GitHub, VS Code, MySQL Workbench, NetBeans
- Data Visualization: Power BI, Tableau
- Core Concepts: Data Structures & Algorithms, OOP, Software Development

## PROJECTS

### PDF Invoice Data Extraction & Automation System (Python, Tesseract, PyPDF2, OCR, API)

- Developed an automated PDF processing pipeline to extract and structure invoice data from text-based and scanned PDFs, implementing OCR fallback (pytesseract, pdf2image, OpenAI API) and regex-based parsing for accurate key extraction.
- Transformed unstructured invoice data into structured JSON and CSV formats, significantly reducing manual data entry effort and streamlining reporting and analysis workflows.

### RAG Book Assistant (Streamlit, LangChain, Mistral AI, Vector DB)

- Built a Retrieval-Augmented Generation (RAG) system to enable Q&A over PDF documents
- Implemented semantic search using embeddings and vector database (Chroma)
- Designed context-aware prompt pipeline to ensure accurate, document-based responses
- Improved information retrieval efficiency by eliminating manual document search

### Traffic Flow Prediction (Python, ML)

- Analysed historical traffic data (vehicle counts, time, and day) to identify congestion patterns and predict Traffic situations using a Random Forest Classifier.
- Performed data preprocessing and feature engineering (datetime conversion, encoding, handling missing values) to improve model accuracy and interpretability.

### Fun Chat Bot (Streamlit + Mistral AI)

- An interactive chatbot that maintains conversation memory and responds with a fun personality named Jassi.
- Chat interface using Streamlit, Conversation memory using LangChain messages, Powered by Mistral Large Language Model, Fun assistant personality ("Jassi"), Environment-based API key management

## Certifications

- Advanced Python Programming (Infosys Springboard)
- SQL Certification (Hacker Rank)
- Core Java Training of 8 weeks (Sunrise Technologies)
- AWS Fundamental (Udemy)

## Experience

### Software Development Intern (Amar Ujala Limited, Meerut)

February 2026 – ongoing

- Assisted in development, testing, and maintenance of internal Automation systems.
- Collaborated with team members to support ongoing Software Development Automation projects.