

SOHAM MUKHERJEE

Kolkata, India | +91 8240771772 | soumyapu31@gmail.com | [LinkedIn](#) | [Website](#) | [GitHub](#)

SUMMARY

Data Science (M.Sc.) student with experience in Machine Learning, Deep Learning, and a strong focus on AI. Skilled in Python, SQL, PowerBI, Statistical Analysis, Neural Networks, Computer Vision, and Natural Language Processing. Passionate about using AI to solve complex, data-driven challenges.

SKILLS

- | | | | |
|--------------------------|-----------------|------------------------|-------------------|
| • Power BI Desktop | • MySQL | • Data Visualization | • Neural Networks |
| • Power BI Service | • Python | • Data Storytelling | • Computer Vision |
| • Power Query (M- Query) | • Data Cleaning | • Statistical Analysis | • Problem Solving |
| • DAX | • Data Modeling | • NLP | • Quick Learner |

PROJECTS

Automatic License Plate Detector | [GitHub](#)

- License plate detection after fine-tuning a pre-trained *YOLOv8* model on our custom dataset.
- Text extraction using *Tesseract* OCR to recognize the license plate numbers.
- Visualizing detected plates and extracted text in the web interface built using *Streamlit*.

Hollywood Movie Recommender | [GitHub](#)

- Used content-based filtering and Natural Language Processing with cosine similarity to recommend similar movies based on genres, keywords, cast, and crew data.
- Integrated the TMDb API to fetch movie details and posters, with data preprocessing involving merging datasets, stemming, and text vectorization.
- Displays movie recommendations through a Streamlit web app.

Food Detector | [GitHub](#)

- Initial classification with a CNN baseline model.
- Transfer learning with EfficientNet and ResNet, and enhanced via data augmentation.
- Fine-tuned pre-trained models to get better results.

WhatsApp Chat Analyzer | [GitHub](#) | [URL](#)

- Preprocess WhatsApp chat data and display key statistics (total messages, words, media, links).
- Visualize user activity with timelines and heatmaps.
- Generate word clouds and analyze emojis using Python's WordCloud library.

Flights Dashboard | [GitHub](#)

- Built using Python's *Streamlit* library for an easy and seamless interaction.
- Fetches data from a real-time database hosted on AWS using Python's *mysql.connector* library.
- Provides a user interface for checking what flights are available from a given city to another city (eg., Kolkata to Bengaluru).

CERTIFICATIONS

- Python for Data Science & Machine Learning [*Udemy* ([URL](#))]
- TensorFlow Developer Certification [*Zero to Mastery* ([URL](#))]
- MySQL Bootcamp [*Udemy* ([URL](#))]
- Certification of Statistics & Probability using Python [*Infosys Springboard* ([URL](#))]

EDUCATION

Master of Science (M.Sc.) in Data Science & Analytics

(2023-present)

NSHM College of Management & Technology

Bachelor of Science (B.Sc.) in Data Science

(2020-2023)

iLEAD