

SHRUTIK PATIL

Nagpur-440002 India

+91 8483032099 ✉ shrutikp70@gmail.com [in linkedin](#) [github](#)

EDUCATION

Indian Institute of Technology Kharagpur

B.Tech + M.Tech in Chemical Engineering

Nov. 2020 – June 2025

Kharagpur, West Bengal — GPA: 7.2 / 10.0

EXPERIENCE

Darzee

Software Engineer Intern

May 2024 – August 2024

Patiala, Punjab

- Developed website with React.js, JavaScript, CSS, integrating 20+ reusable components that reduced time by 25%
- Optimized mobile applications using Flutter, implementing data-driven UI enhancements that increased user session
- Collaborated cross-functionally to implement A/B testing for UI elements, driving data-driven design decisions.

PROJECTS

Clutch Performance Detection in T20 Cricket | *Python, Streamlit, Pandas, Seaborn, Logistic Regression* **January 2024**

- * Developed win probability model using logistic regression on 15K+ deliveries (84.7% ROC-AUC)
- * Quantified clutch performance via Bayesian inference and WP delta, improving player ranking accuracy by 37%
- * Processed 1M+ deliveries from Cricsheet and ESPN using pandas, pyyaml with 92% phase labeling accuracy.

BooksThatFeelLikeThis | *Python, SentenceTransformers, spaCy, FAISS, CLIP, Reddit API* **December 2024**

- * Developed a scalable data ingestion pipeline for Reddit data, reducing preprocessing overhead by 60% for NLP tasks
- * Designed FAISS pipeline with MiniLM-L6-v2 embeddings for personalized book recommendations via text/image
- * Achieved 91.3% precision on title extraction via a regex + spaCy/Hugging Face NER, validated against OpenLibrary
- * Enabled multimodal querying with CLIP (ViT-B/32) image embeddings, matching visuals to books with 87% accuracy

Modeling COVID-19 Transmission Dynamics | *Course Project, IIT Kharagpur* **October 2024**

- * Designed forecasting pipeline using LSTM and SIR models, improved accuracy by 26% over baseline ARIMA models
- * Trained neural networks on 200K+ time-stamped, achieving an MAE of 0.21 on infected case forecasts over a 30-day
- * Simulated stochastic transmission volatility, replicating observed spread patterns in less than 90% of test regions

AI-Powered YouTube Video Summarization System | *Python, OpenAI GPT-3.5, NLP, API Integration* **June 2024**

- * Developed end-to-end NLP pipeline extracting transcripts from 100+ YouTube videos using YouTubeTranscriptAPI
- * Engineered text preprocessing workflow with rpunct library to restore punctuation in unstructured transcripts.
- * Enhanced GPT-3.5 summarization pipeline using prompt engineering, improving video transcript relevance by 40%
- * Automated content digestion process, reducing average video summarization time from 10 minutes to 30 seconds

COMPETITION / CONFERENCE

Data Analytics — Inter Hall — IIT Kharagpur **Mar 2022 - Apr 2022**

- * Objective: Predict inventory levels of warehouses for a month using dataset of past 2 years with ML, Time Series models
- * Decomposed data using seasonal decomposition; identified linear seasonality without trend across SKU time series
- * Switched from ARIMA with MAPE of 48.63% to LSTM, improving forecast accuracy and achieving MAPE of 1.31%

COURSEWORK & SKILLS

Languages: Python, SQL, JavaScript, HTML/CSS

ML Frameworks/Architecture: scikit-learn, Pytorch, spaCy, Transformers

Data Tools: Pandas, NumPy, Matplotlib, Seaborn, FAISS, OpenLCA

Developer Tools: Jupyter, VS Code, GitHub, Linux, Google Colab

CERTIFICATIONS

Python for Data Science and Machine Learning

Natural Language Processing in Python — Maven Analytics