

# Geetansh Yadav

| +91-9024418965 | yadavgeetansh007@gmail.com | linkedin.com/in/Geetansh | github.com/geetansshh

## EDUCATION

### Indian Institute of Information Technology, Nagpur

*B.Tech, Computer Science and Engineering*

– **CGPA: 8.73**

Nagpur, India

*Aug 2022 – Present*

### Central Board of Secondary Education (CBSE)

*Senior Secondary: 94.6% (2021) — Secondary: 94% (2019)*

## EXPERIENCE

### SMS Datatech

*Software Engineering Intern*

Tokyo, Japan

*Jul 2025 – Present*

- Design and implement an **AI-assisted web scraping** platform in **Python** using , **Crawl4AI**, and **Firecrawl** to extract **structured data** at scale.
- Extend the framework for **domain-specific** use cases (e-commerce, GovTech) with reusable spiders, schema normalization, and modular pipelines.
- Migrate legacy scrapers into a **unified Scrapy** codebase to reduce duplication and complexity, improving **maintainability** and **debuggability**.
- Optimize large-scale **ETL/data pipelines** for **reliability** and **performance**; improved crawl throughput and downstream ingestion stability.
- Harden the system for **high-traffic** sources through better error handling, configuration management, and testing.
- Use **GraphQL** APIs to pull **structured JSON** directly from backends—replacing brittle HTML/CSS selector scraping; implement **pagination**, query batching, and auth to boost **stability**, **throughput**, and **maintainability**.

*Skills: Python, Scrapy, Crawl4AI, Firecrawl, Web Scraping, ETL, Data Pipelines, Scalability, Code Quality*

### Mphasis

*Software Engineering Intern*

Noida, India

*Jun. 2024 – Aug. 2024*

- Built and optimized **user interfaces** for an interview management application.
- Prototyped an **AI-driven resume shortlisting bot**, improving candidate screening efficiency.

## PROJECTS

### Fake News Detection Using NLP

- Developed an **NLP-based** fake-news detection system for near real-time classification of deceptive content; implemented training/inference pipelines in **Python**.
- Evaluated multiple **machine learning** approaches to improve accuracy and adaptability to evolving misinformation patterns.
- Applied **continuous learning** strategies to maintain model effectiveness over time.

### Chikitsa: Medical Chatbot & Appointment Booking

GitHub

- Built a **chatbot** to handle medical queries and basic triage, returning concise health information.
- Implemented **interaction history** (7-day retention) so users can review previous conversations.
- Designed a **booking workflow** to schedule appointments directly via the chatbot interface.

## TECHNICAL SKILLS

**Languages:** Python, C, C++, SQL, HTML, CSS, JavaScript, TypeScript

**Frameworks/Libraries:** Scrapy, Crawl4AI, Firecrawl, Flask, React, TensorFlow, Keras, PyTorch, OpenCV, Seaborn

**APIs:** GraphQL

**Areas:** Web Scraping, Data Pipelines, ETL, Machine Learning, NLP, API Integration, Scalability, Debugging, Maintainability, Code Optimization

**Profiles:** CodeChef **3 (1652)** Geetansh007 — LeetCode Geetansh

## LEADERSHIP & ACTIVITIES

**Founder**, Pahal — community initiative supporting underprivileged students.

**Core Team Member**, Udyam (Entrepreneurship Cell, IIIT Nagpur) — Event Management Lead.