

MAHFUJAR RAHMAN

Kandi, Pirgacha – 5450, Rangpur, BD

☎ +8801747430802 | ✉ mrnoyon.cse@gmail.com |  [LinkedIn](#) |  [GitHub](#) |  [GitLab](#)

Python Developer with 3+ years of experience building web crawlers and scalable data pipelines. Skilled in overcoming scraping challenges such as IP rotation, CAPTCHA solving, and JavaScript-rendered content, and familiar with Scrapy. Quick learner and problem solver, with a strong track record of adapting to new tools and delivering efficient, production-ready solutions.

Work Experience

Python Data Engineer at **LOBSTR.IO**, France (Remote)

Oct, 2022 – Present

- **Developed and optimized** *Python-based web crawlers* to automate large-scale data extraction.
- **Handled anti-bot systems** (*Datadome, PerimeterX*) using proxy rotation & CAPTCHA-solver.
- Enhanced crawler frameworks with **modular, reusable components** for faster development and deployment.
- **Designed and maintained** ETL pipelines using *PostgreSQL & SQLAlchemy* to ensure data integrity and scalability.
- Utilized **AWS S3** for secure data storage, backup, and integration with data workflows.
- **R&D** to solve complex data engineering challenges.

Junior IT Executive at **WSDA New Zealand**, Dhaka (Onsite)

Aug, 2022 – Oct, 2022

- **Collected and organized data** to support business development.
- **Validated and verified data** for accuracy and integrity in decision-making.
- **Managed databases**, ensuring up-to-date and accessible records.
- **Contributed to value production** and distribution strategies for optimized processes.

Education

- B.Sc. in Computer Science and Engineering with CGPA 3.12/4.00
University of Rajshahi (Rangpur Engineering College) 2017 – 2022
- H.S.C. in Science with GPA 4.88/5.00
Devi Chowdhurani Degree College, Rangpur 2015 – 2016

Skills

- **Programming:** Python (Advanced), SQL, JavaScript, Regex
- **Web Crawling:** Requests, lxml, BeautifulSoup, Anti-bot handling (*Datadome, PerimeterX, CAPTCHA-solving services*), Proxy/IP rotation, JavaScript-rendered sites
- **Data & Tools:** SQLAlchemy, PostgreSQL, AWS (S3), Git, Linux (*Ubuntu/WSL2*)
- **Other:** ETL Pipelines, Data Quality Assurance, Problem-Solving

Related Projects

Arogga Web Crawler

 <https://github.com/mahfujars/arogga>

- **Tech Stack:** Celery, SQLAlchemy, PostgreSQL, Regex, JSON, Requests, lxml
- **Data:** Scraped and cleaned product data (Name, SKU, Price, etc.) across all categories.
- **Error Handling:** Managed pagination, rate limits, and 404 errors for uninterrupted scraping.
- **Optimization:** Leveraged Celery for parallel tasks, improving speed, with PostgreSQL for efficient storage.

Sia.Ch Web Crawler (*Interview test at lobstr.io*)

 <https://github.com/mahfujars/sia>

- **Tech Stack:** SQLAlchemy, MySQL, Regex, JSON, Requests, lxml
- **Data:** Scraped and cleaned member & office data (Name, Member, Name-Member association).
- **Error Handling:** Managed pagination, rate limits, and 404 errors.
- **Security:** Decrypted phone numbers using crypto.cipher AES.

SweetChat (*Academic project*)

 <https://github.com/mahfujars/sweetchat-nextjs>

- **Role:** Project leader
- **Tech Stack:** Next.js, React, Styled-Components, Material-UI, Firebase.
- **Features:** Real-time messaging for a university project.

Misc

- **CONTESTANT** of **The 2019 ACM-ICPC Asia Dhaka Regional Contest** (Onsite)
 - **Certificate:** https://github.com/mahfujars/mahfujars/blob/main/MAHFUJAR_ICPC19.pdf
- **TOP EARNER** in Web Design and Development (WDD) at **LEDP Project 2020, ICT Division**
 - **Certificate:** <https://ledp.ictd.gov.bd/certificate/?regid=376778>
- **Assistant Country Director (ACD)** at International Youth Summit
- **Campus Ambassador** at ICT Olympiad Bangladesh