

Sujeet Kumar

Data Analyst | Software Developer Intern

✉ sujeetsmpatel@gmail.com ☎ 7870107176 🌐 Sujeet2003 🔗 sujeetsmpatel.vercel.app

🌐 linkedin.com/in/sujeet-kumar-005561286/ 🏠 kaggle sujeet2003

EDUCATION

B.E., *HKBK College of Engineering*, CGPA - 8.8 11/2021 – present | Bengaluru
12th, *Shershah College*, Percentage - 80% 06/2018 – 04/2020 | Bihar
10th, *Central English School*, Percentage - 80% 05/2013 – 04/2018 | Bihar

EXPERIENCE

Software Developer Intern, *Tata Technologies (Onsite)* 09/2024 – 11/2024 | Bengaluru

- Developed an offline private ChatGPT using **Python**, **GenAI**, and **PostgreSQL**, integrating it with a secure local database for **100+ employees**.
- Implemented **multi-mode chatbot** functionality (**Basic**, **General**, **Summarize**), enhancing efficiency and user satisfaction.
- Added **chat history** features to improve engagement and continuity in information retrieval.

SKILLS

Languages: HTML | CSS | Javascript (Basic) | Python | SQL

Libraries & Framework: Django | Numpy | Pandas | Matplotlib | Seaborn | NLTK | Scikit-Learn

Others: VS Code | Jupyter Notebook | Git & github | Statistics | Probability | Machine Learning | Natural Language Processing | Data Structure | Data Analytics | EDA | PostgreSQL | MySQL

PROJECTS

Quora Question Pair Detection, *Python | Feature Engineering | NLP | Machine Learning* 📄

- Developed a duplicate question detection system using NLP and machine learning, achieving **80% accuracy** on **400,000+** Quora question pairs.
- Performed advanced text preprocessing and exploratory data analysis (EDA) to enhance model **performance by 5%**.
- Enhanced the model's capability to detect duplicate questions, reducing redundancy in large-scale datasets and improving search efficiency for end-users.

Network security System, *Python | ML | AWS | Github* 📄

- Implemented an **MLOps pipeline** for network security with ETL, logging, preprocessing, model training, and **AWS S3** deployment.
- Deployed Dockerized applications on **AWS EC2** for efficient real-time **network security predictions**.
- Automated **CI/CD** workflows using **GitHub Actions** for seamless integration and reliable deployment.