RAHUL SAMANT

+91-9368884561 / rahulsamantcoc2@gmail.com / linkedin.com/in/ rahul-samant-kb37 / github.com/rahulsamant37 | Portfolio

OBJECTIVE

Data Scientist with expertise in **Python, AI-powered automation**, and advanced web scraping. Proficient in designing and implementing scalable, high-performance solutions to tackle complex data challenges, with a focus on leveraging cutting-edge **machine learning and deep learning methodologies**. Experienced in **delivering impactful,** data-driven results across diverse industries, including **predictive analytics, process optimization, and intelligent decision-making systems**.

WORK EXPRIENCE

Python Developer Intern
Infosys Springboard | September – Present

- To develop a robust and efficient **web scraping application** to automate **data extraction from various e-commerce platforms**, reducing manual effort and improving data accuracy.
- **Technologies: BeautifulSoup, Playwright, Streamlit, Flask, LangChain, ChatGenerativeAI**, LangServe
- Utilized **Playwright for web scraping**, BeautifulSoup for HTML parsing, **LangChain** for language model integration, and **ChatGenerativeAI** for extracting specific product details. Deployed the application using **Streamlit and LangServe for a user-friendly interface**.
- Implemented advanced scraping techniques to handle diverse web structures, achieving **an accuracy rate of 99%** in data extraction.
- Enhanced scraping speed by **30% using multithreading**, improving performance on high-traffic pages and reducing average load time.
- Enabled data extraction from over **50 unique websites** with complex HTML structures, leading to a **20% reduction** in manual data collection time for users.

PROJECTS

AI Research Assistant with Semantic Search | Lead Developer   (GitHub Link)
- Developed an **AI-powered research assistant** with semantic document search capabilities, processing and indexing **over 50,000 documents** with 95% search accuracy compared to traditional keyword-based systems.
- **Technologies: Python, PostgreSQL, pgvector, pgai, Ollama, Streamlit, Sentence-Transformers, TimescaleDB, psycopg2, Pandas**.
- Engineered a scalable system leveraging **384-dimensional embeddings** for semantic representation, achieving **40% faster** query response times through **IVFFlat indexing**, and integrated Ollama for advanced document analysis.
- **Achieved: 87% improvement** in search relevancy scores, **3x faster** document processing with batch operations, **92% accuracy** in content summarization, and **65% reduction** in query latency through vector optimization.
- **Enhanced** system performance through vector scaling, resulting in **99.9% uptime** and ability to handle **1000+ concurrent searches** with sub-second response time.
- **Future Work:** Planning to implement enhanced visualization features and advanced CRUD operations while expanding AI capabilities through broader pgai integration.

Neural Machine Translation System | Lead Developer   (GitHub Link)
- Engineered a **sequence-to-sequence neural translation model** for English-to-Spanish translation using **encoder-decoder LSTM architecture**, achieving high accuracy in language conversion tasks.
- **Technologies: Python, TensorFlow, Keras, NumPy, GloVe Word Embeddings, LSTM Networks, Seq2Seq Architecture**.
- Implemented sophisticated **data preprocessing pipeline** with custom tokenization and word embedding integration using **Stanford's GloVe 100-dimensional vectors**, resulting in improved translation quality.
- **Achieved: 90%+ accuracy** in basic translation tasks, integrated **20,000 training samples** with 18,000/2,000 train-test split, and designed system for **variable-length sentence processing.**
- **Optimized** model architecture through embedding layer customization and **bidirectional LSTM networks**, enabling efficient handling of complex sentence structures and context retention.
- **Future Work:** Implementing attention mechanisms and transformer architecture to enhance translation accuracy and context handling.

EXTRA CURRICULAR ACTIVITIES & ACHIEVEMENTS

- Actively contributed to renowned repositories, including **Gradio** and **Wagtail**, enhancing their functionalities and supporting community-driven development.
- Successfully completed **Hacktoberfest** by having **4 Pull/Merge Requests accepted**, demonstrating commitment to collaborative open-source contributions.
- Shortlisted among only 25 candidates for the prestigious Data Science program **at IIT Kharagpur**, a highly selective achievement in a competitive field**.**

TECHNICAL SKILLS

Languages & Tools: Python, SQL, Docker, Git, Flask, Streamlit
AI/ML: TensorFlow, Keras, LangChain, BERT, OpenAI, Gemini
Data Engineering: NumPy, Pandas, BeautifulSoup, Playwright, pgvector
Core Competencies: NLP, System Design, API Integration, OOP, Database Management
Soft Skills: Communication and Interpersonal Skills, Team Leadership, Analytical Thinking

EDUCATION

Punjab Engineering College	Chandigarh
Bachelor of Technology in Computer Science Engineering	2022 – Present
City Convent School (Percentage: 95.00%)	U.S. Nagar, Uttarakhand
Senior Secondary Education, CBSE	2020– 2021
City Convent School (Percentage: 90.00%)	U.S. Nagar, Uttarakhand
Secondary Education, CBSE	2018 – 2019