Saboor Ahmed

AI Engineer

Data scientist with expertise spanning from web scraping to exploratory data analysis, visualization, model training, and deployment. Passionate about AI, I develop custom Retrieval-Augmented Generation (RAG) systems integrated with large language models (LLMs), leveraging scraped data for intelligent applications.

suboorahmed2002@gmail.com

+92 331-8111331

Lahore, Pakistan

saboor2002.github.io/landing_page3/

linkedin.com/in/saboor-ahmed-2002sa

github.com/saboor2632

EDUCATION

Bachelors in Computer Science
National University of Computer and Emerging Sciences (FAST-NUCES)
2021 - 2025 *Lahore*

SKILLS

Python | Langchain | Scrappy | Bs4 | Qdrant | Flask | FastAPI | Streamlit | Supabase | SQL | PostgreSQL | LLM | NLP | RAG

PERSONAL PROJECTS

AI-Powered Job Query System with RAG

- Scraped job listings from multiple Pakistani job sites using **Scrapy and BeautifulSoup4** , then stored structured job data in **Qdrant (vector database)** . Leveraged **Hugging Face's BGE-small embeddings** for semantic similarity search, enabling efficient retrieval of relevant job postings. Integrated **Gemini Flash 1.5 Pro** as the LLM to process user queries and provide context-aware job insights. This **Retrieval-Augmented Generation (RAG) system** allows users to explore job opportunities in a conversational and intelligent manner.

AI-Powered News Aggregator & Recommender

- Scraped news articles from **Dawn News** and categorized them using a **classification model** built with **NLP techniques** like tokenization, stop-word removal, and **Word2Vec embeddings** . Stored categorized articles in **MongoDB** for efficient retrieval. Implemented a **personalized recommendation system** , linking user accounts with preferred categories to deliver relevant news via a user-friendly app interface.

Bank Customer Churn Prediction

- Developed a **churn prediction model** for banks using a **Random Forest Classifier** , achieving **86% accuracy** through **hyperparameter tuning** . Built a **Flask API** to deploy the model, enabling real-time predictions for customer retention insights.

Operational Implementation of Satellite-Based Fire Detection

- Implemented an **Enhanced Contextual Fire Detection Algorithm** to identify **thermal anomalies** using **MODIS satellite data** . Processed **MOD021KM** data to retrieve radiance values, converted them to **brightness temperatures** , and applied **cloud and water masking** to reduce false positives. Used **thresholding techniques** and **background characterization** to detect and classify fire pixels. Validated results against the **MOD14 fire product** , achieving accurate fire detection. The workflow is scalable for applications in **environmental monitoring, disaster response, and climate research** .

Solar Power Prediction

- Created a solar power prediction software using machine learning in a group project. We used random forest as the model and evaluated it through K fold cross validation.

CERTIFICATES

ETHGlobal Agentic Ethereum Hackathon Participant (02/2025)
*Participated in **Agentic Ethereum 2025** , an ETHGlobal event focused on blockchain, AI agents, and smart contract innovations. Engaged in **hacking, workshops, and keynotes** while submitting the project **"ETH0"** to the event. Developed expertise in **Ethereum-based AI agents and decentralized applications (dApps).***

Environmental Data Analyst – Internship at EH-DS Lab (FAST-NUCES Lahore) (06/2024 - 08/2024)
*Successfully completed an **internship at the Environment & Health Data Science (EH-DS) Lab, FAST-NUCES Lahore** , as an **Environmental Data Analyst***

Python for Data Science, AI & Development – IBM (Coursera) (06/2024)

Introduction to Relational Databases (RDBMS) - IBM (Coursera) (06/2024)

Databases and SQL for Data Science with Python - IBM (Coursera) (07/2024)

ETL and Data Pipelines with Shell, Airflow and Kafka - IBM (Coursera) (07/2024)

LANGUAGES

English
Full Professional Proficiency

Urdu
Full Professional Proficiency

INTERESTS

LLM | RAG | Gen AI | ETL