Sami ur Rehman Khan
Software Engineer

sami.u.r.khan10@gmail.com ✉

+92-336-4590272 📱

Lahore, Pakistan 📍

linkedin.com/in/sami-ur-rehman-khan 🔗

I'm a Computer Science graduate passionate about Big Data, AI, Data Warehousing, and Soware Development. Eager to work for a company where I can contribute my skills and passion, especially my data-driven approach, to develop cutting-edge products and services that fuse several domains.

EDUCATION

Master of Sciences in Data Science
FAST NUCES

08/2025 - Present   *Lahore, Pakistan*

Bachelor of Sciences in Computer Science
University of Central Punjab

Lahore, Pakistan

PROJECTS

Live Data Sentiment Classifier (Reddit)
- Developed a Python-based pipeline to scrape real-time Reddit comments, perform sentiment analysis using **VADER** , and train an **ML classifier** (Logistic Regression with TF-IDF). Automated data storage in **SQLite** and implemented model persistence with **Joblib** .

Hostel Finder Assistant Voice Bot
- Developed an intelligent voice bot using OpenAI and Vapi AI to assist users in finding hostels based on their location and preferences. Designed and integrated a voice-based interaction system to provide a seamless and user-friendly experience. The bot ensures fast, accurate information retrieval, allowing users to get real-time responses and personalized recommendations for hostel options. This project highlights the effective use of AI and natural language processing to enhance user convenience and decision-making.

Data Analysis and Machine Learning with Python
- For a Programming for Big Data (PBD) project, utilized Python libraries such as Pandas, NumPy, Matplotlib, and Seaborn for data preprocessing, analysis, and visualization. Cleaned and transformed datasets to extract meaningful insights, created visual representations, and applied machine learning algorithms like K-Means for clustering and Linear Regression for predictive modeling. Focused on improving data quality and understanding patterns within the data.

Real-Time Twitter Data Streaming and Storage Using Apache Hadoop
- For the final Programming for Big Data (PBD) project, implemented real-time data streaming using Apache Hadoop, Apache Flume, and the Twitter API. Configured the flume.conf file with API keys obtained from the Twitter Developer platform to collect live tweets based on specific keywords. The streaming data was ingested and stored in Hadoop Distributed File System (HDFS) for further processing and analysis. This project demonstrated the integration of big data tools to manage and store large-scale, real-time data streams effectively.

E-Commerce Android Application(FYP)
- Designed and developed a feature-rich e-commerce Android application for the final year project using Java and Google Firebase as the backend. The application catered to both local and branded buyers, allowing users to browse and purchase products seamlessly. A standout feature enabled users to request items not available in the catalog directly from suppliers, ensuring a personalized shopping experience. The application focused on user-friendly navigation, real-time data synchronization, and efficient product management through Firebase integration.

VOLUNTEER EXPERIENCE

Plantation Drive
Lahore

Participated in a mass-scale plant drive held on the outskirts of Lahore. Successfully planted 150 plants within a matter of hours.

SKILLS

Attention to Detail   Team Player
Quick Learner   Adaptive   Logic Builder

TOOLS

Power BI

Jupyter Notebook

MS Excel

CERTIFICATES

Mastering AI and Prompt Engineering (01/2025)
Completed a comprehensive course on AI and prompt engineering, focusing on building and optimizing AI-driven solutions through effective prompt design and interaction with large language models.

Python Developer (05/2025)
https://api2.sololearn.com/v2/certificates/CC-HX3BZ8DJ/image/png?t=638830686589841000

Machine Learning (06/2025)
Completed a short course where I revised concepts such as data preprocessing, model evaluation, and feature engineering. The course focused on supervised learning techniques, with hands-on practice in algorithms like linear regression. Toward the end, we explored deep learning, including artificial neural networks (ANN) and convolutional neural networks (CNN), along with their practical applications.

LANGUAGES

Python
Professional Working Proficiency

SQL
Professional Working Proficiency

JAVA
Professional Working Proficiency

C++
Professional Working Proficiency

INTERESTS

Outdoor Sports   Travelling
Exploring Nature   Data Analytics
Data Warehousing   Business Intelligence