

Safiullah Khokhar

📍 Lahore, Punjab, Pakistan

✉️ safiullahkhokhar7@gmail.com

☎️ +923497496965

🌐 [in/safiullah-khokhar](https://github.com/Safiullahkhokhar)

🌐 <https://github.com/Safiullahkhokhar>

SUMMARY

AI/ML Engineer with a background in machine learning, data preprocessing, and model implementation. Experienced in developing and optimizing machine learning models utilizing Python and frameworks such as TensorFlow, PyTorch, and Scikit-learn. Capable of handling huge datasets via Pandas and NumPy, visualizing insights, and developing AI solutions for complicated issues. I'm passionate in AI-driven innovation and optimizing machine learning workflows for real-world applications.

EXPERIENCE

Data Analyst Lead Internship (Remote)

YoungDev

April 2024 – May 2024

- Built and optimized a customer churn prediction model.
- Analyzed and structured large datasets to extract actionable insights.
- Worked closely with business and data teams to optimize model performance.

Machine Learning Internship (Remote)

Mentorness

May 2024 – June 2024

- Processed, cleaned, and formatted diverse datasets using Pandas and NumPy.
- Built and optimized ML models, including regression and classification models.
- Visualized data using Matplotlib and Seaborn for insightful analytics.
- Developed a Random Forest-based time series prediction model.

PROJECTS

Predictive Modeling for Disease Diagnosis

April 2024 - May 2024

- Collected and preprocessed data from diverse healthcare sources, ensuring data quality and integrity.
- Investigated how select features to enhance model performance, including handling missing data, normalization, and categorical variables.
- Implemented and tuned models using Python libraries such as Scikit-Learn, TensorFlow, and Pandas.
- The system provided data-driven insights, supporting clinicians in making more informed decisions based on patterns and trends observed in the data.
- The project fostered collaboration between data scientists and medical professionals, bridging the gap between technology and healthcare.
- Deployed the model into a user-friendly application for real-time diagnosis, improving patient outcomes and facilitating early disease detection.

Scrape YouTube using YouTube Data API

Nov 2023 – Dec 2023

- Collect metrics such as subscriber count, view count, upload date, and more to gain insights into channel performance.
- Retrieve information like video title, description, publish date, category, and comment count for comprehensive video analysis.
- Customize the scraping process to target specific channels, video categories, or keywords, ensuring you gather the most relevant data for your objectives.
- Implement robust error handling and rate limiting mechanisms to adhere to YouTube Data API usage guidelines and maintain smooth operation.
- Insightful visualizations (charts, graphs) using libraries like Matplotlib or Seaborn to effectively communicate the extracted data's story.
- Time series analysis techniques to identify trends in channel growth, video views, and other metrics over time.

EDUCATION

BSCS

Shaheed Zulfiqar Ali Bhutto Institute of Science and Technology (SZABIST)

• Karachi, Pakistan • CGPA 3.1 • [In Progress Aug 2025]

CERTIFICATIONS

Data Analyst with Power BI

DataCamp • 2023

Machine Learning Specialization

Coursera • 2024

SKILLS

Technology: Python, SQL, Git/GitHub, Power BI, Google Sheet & Excel

COURSEWORK

Artificial Intelligence

SZABIST

Data Science

SZABIST