



Haider Farooq

Mobile: +92 317 4008222 1234haiderfarooq@gmail.com
Islamabad Pakistan <https://github.com/haiderfarooq3>

Education

FAST National University of Computer and Emerging Sciences

Aug 2022 – Present

Majoring in BS (Data Science)

Relevant Courses: Programming Fundamentals, Object Oriented Programming, Data Structures, Introduction to Data Science, Fundamentals of Big Data Analytics, Advanced Statistics, Computer Organisation and Assembly Language

Projects:

Spotify Clone – Python

May 2024

- Developed a Spotify alternative with music recommendation, playback, and streaming using a sample of the FMA dataset, Python for feature extraction, **MongoDB** for data storage, and **K-means clustering** for model training.
- Deployed an interactive music streaming web app with real-time recommendations using **Flask** and **Apache Kafka**, showcasing personalized music recommendations and real-time streaming.

Automated Kafka Setup for Real-Time Product Insights

April 2024

- Extracted insights from Amazon metadata using **Apriori** and **PCY** algorithms, real-time data streaming, and MongoDB integration for scalable analysis.
- Automated Kafka setup** with a bash script, showcasing the potential of streaming analytics for actionable product insights.

Text Mining and Analysis Pipeline

March 2024

- Text Data Processing and TF-IDF Vector Generation:** Cleaned and standardized textual data using NLTK, generated TF-IDF vectors, and normalized them using NumPy to ensure effective document comparison.
- Scalable Analysis with Hadoop:** Utilized Hadoop clusters for scalable data processing, implementing MapReduce to calculate word frequencies, IDF values, and TF-IDF scores, enabling efficient analysis of large datasets.

Advanced Image Retrieval System using LSH

March 2024

- Image Processing and LSH Implementation: Preprocessed CIFAR-10 images, performed color analysis, generated textual shingles, and used **Minhashing** and Locality-Sensitive Hashing (**LSH**) for efficient similarity detection.
- Web Application Integration: Built a **Flask** web app to handle image uploads and display similar images using the LSH algorithm, ensuring fast, scalable retrieval and improved user engagement.

Sentiment Analysis and Preprocessing of Customer Reviews

Feb 2024

- Text Preprocessing: Loads reviews from a JSON file, removes stop words and punctuation, and converts text to lowercase using a comprehensive stop words list.
- Sentiment Analysis: Analyzes reviews with predefined word sets, categorizes sentiment, and saves results with sentiment scores to a text file.

Additional Skills and Experience

- ApacheHadoop, Kafka & Spark (Beginner Level)
- Machine Learning
- PowerBI
- MongoDB
- C++
- SFML
- HTML
- Assembly Language