

FATIMA SOHAIL

BS Data Science | Machine Learning | Computer Vision | Data Analytics
Gujranwala, Pakistan | +92 321 9641545 | fatimaa.cheemaa24@gmail.com
LinkedIn | GitHub | Kaggle

Professional Profile

BS Data Science undergraduate with hands-on experience in machine learning, deep learning, computer vision, data analytics, big-data processing, and data warehousing. Currently working on a research-oriented Final Year Project in few-shot 3D Gaussian Splatting and self-ensembling. Experienced with Python, PyTorch, scikit-learn, Apache Spark, SQL, R, CUDA, and QGIS across machine learning, large-scale analytics, and research projects. Seeking data science, machine learning, computer vision, analytics, and research opportunities.

Education

GIFT University, Gujranwala

Expected 2027

Bachelor of Science in Data Science

CGPA: 3.5/4.00

Relevant Coursework: Machine Learning, Deep Learning, Big Data Analytics, Data Mining, Data Visualization, Data Warehousing & BI, Parallel & Distributed Computing, Artificial Intelligence, and Database Systems.

Research & Final Year Project

Outlier-Aware Self-Ensembling for Few-Shot 3D Gaussian Splatting

2026–Present

Computer Vision, 3D Gaussian Splatting, Few-Shot Learning, Self-Ensembling, PyTorch

- Developing a research-oriented approach for improving few-shot 3D reconstruction and novel-view rendering under limited training views, with emphasis on outlier-aware learning and robust self-ensembling.

EfficientNet-GAT-Fuzzy Cervical Cytology Classification

2026

Python, PyTorch, EfficientNet-B0, Graph Attention Networks, CUDA, Grad-CAM

- Co-developed a five-class computer-vision pipeline on 4,049 SIPaKMeD images using stratified five-fold cross-validation, achieving 97.46% mean accuracy and 97.47% macro F1; conducted ablation experiments and Grad-CAM interpretability analysis.

Selected Data Science Projects

NYC Taxi Trip Analytics using Apache Spark

2026

Python, Apache Spark, Spark SQL, Parquet, Window Functions

- Analyzed 2.96M NYC Yellow Taxi trips using Spark DataFrames and SQL; cleaned the dataset to 2.87M valid records and reduced repeated-query execution from 14.03s to 1.25s through caching, an 11.21× speedup.

Hybrid Parallel Fraud Detection System

2026

C++17, OpenMP, CUDA, CMake, K-Nearest Neighbors

- Built Serial, OpenMP, CUDA, and Hybrid exact K-NN pipelines over 500,000 transactions; achieved a 17.84× controlled compute speedup at $K = 5$ using hybrid CPU–GPU execution.

Pakistan E-Commerce Customer Segmentation

2026

Python, Pandas, scikit-learn, RFM Analysis, K-Means

- Processed 584,524 e-commerce transactions and engineered RFM features; applied preprocessing and $K = 3$ clustering with a 0.36 Silhouette Score to identify high-value, potential, and at-risk customer groups.

Tour Management Data Warehouse

2026

PostgreSQL, SQL, Data Warehousing, Dimensional Modeling

- Designed relational and analytical database models using denormalization and Star, Snowflake, and Galaxy schemas for booking, customer, hotel, package, and payment analytics.

Technical Skills

Machine Learning & Computer Vision: Python, PyTorch, scikit-learn, 3D Gaussian Splatting, CNNs, transfer learning, supervised and unsupervised learning, K-NN, SVM, Random Forest, K-Means, PCA, cross-validation, feature engineering, Grad-CAM, model evaluation

Data Analytics & Visualization: Pandas, NumPy, SQL, EDA, statistical analysis, Matplotlib, Seaborn, Tableau, Power BI, Microsoft Excel, R, ggplot2, R Shiny, QGIS

Big Data & Data Engineering: Apache Spark, Spark SQL, Hadoop, Docker, PostgreSQL, MySQL, MongoDB, ETL, data preprocessing, data warehousing, dimensional modeling

Programming & Tools: C++17, CUDA C++, OpenMP, CMake, Git/GitHub, Kaggle, Google Colab, Jupyter Notebook, VS Code, Linux

Certifications & Activities

Google / Coursera: Google Advanced Data Analytics; Foundations of Data Science; The Nuts and Bolts of Machine Learning

Deloitte Australia: Data Analytics Job Simulation – Forage

Oct 2025

Hackathons: Harvard HSIL Hackathon (Apr 2026) | InnovateX 2.0 Code Sprint – Junior Participant (Aug 2025)