

Sarah Shabbir

Data Scientist | Researcher | Analyst

Rawalpindi, Pakistan | Phone: +92 333 5428093 | sarahshabbir345@gmail.com

<https://www.linkedin.com/in/sarah-shabbir-3ba157234/>

SUMMARY

Analytical and research-driven Data Scientist with a strong foundation in machine learning, data analytics, and statistical modeling. Completed a Computer Vision FYP, “AutoCop” involving data annotation, image preprocessing, and model development. Gained hands-on experience in Healthcare Data Science through an internship at “RADS”, applying data-driven approaches to healthcare research and analytics. As the Founding Vice President of the AU Data & AI Society, led initiatives promoting AI, data science, and technical learning. Passionate about transforming data into actionable insights and solving real-world problems through AI-driven solutions.

EDUCATION

Air University, Islamabad

Bachelor of Science in Data Science

Islamabad, Pakistan

Sep 2022 – June 2026

Relevant Coursework: Data Analysis, Machine Learning, Deep Learning, Design Algorithms, Software Engineering, Artificial Intelligence, Statistics & Probability, Multivariate Statistics & Reliability Analysis, Big Data Analysis, Data Warehouse

WORK EXPERIENCE

Research & Development Solutions

Data Science Research Intern

Islamabad, Pakistan

July 2025 – Sep 2025

- Contributed to **IRB-aligned research projects**, supporting data-driven healthcare and socioeconomic studies.
- Analyzed **large-scale PSLM and PBS datasets** comprising thousands of survey records using Python to generate insights on Pakistan's health and education expenditure trends.
- Cleaned, validated, and performed exploratory data analysis (EDA) on **multi-variable national survey datasets**, improving data quality and supporting evidence-based research outcomes.

Freelance Data Scientist

Junior Researcher and Consultant

Remote

2024 – 2025

- Performed **Sentiment Analysis and Text Mining** using Natural Language Processing (NLP) techniques to extract insights from customer reviews, social media data, and feedback forms.
- Cleaned, preprocessed, and transformed **unstructured text data** using Python, enabling accurate analysis and feature extraction for machine learning workflows.
- Applied **Exploratory Data Analysis (EDA)** and sentiment classification techniques to identify customer trends, preferences, and pain points.
- Generated actionable insights and presented findings through reports and visualizations, supporting data-driven business decision-making and stakeholder recommendations.
- Utilized **Python, Pandas, NumPy, NLTK, Data Wrangling** and **machine learning techniques** to analyze textual datasets and improve analytical outcomes.

Air University

Computer Vision Intern

Islamabad, Pakistan

2023 – 2024

- Collaborated with faculty members on **PhD-level research** in **Computer Vision** and **Machine Learning**.
- Conducted data annotation and preprocessing to support model development and evaluation.
- Served as a Research Writer, contributing to research documentation, literature reviews, and technical reporting.

Techparade Software House

Social Media Manager

Remote

October 2023 – February 2024

- Managed and updated brand social media accounts.
- Developed **content strategies** aligned with business goals.

TECHNICAL SKILLS

Programming & Databases | Python, SQL, C++, R, C, MySQL, Git, GitHub

Data Science & Machine Learning | Machine Learning, Deep Learning, Computer Vision, Natural Language Processing (NLP), Sentiment Analysis, Data Mining, Predictive Modeling, Statistical Analysis, Exploratory Data Analysis (EDA), Feature Engineering, Classification, Regression, Clustering, Data Annotation, Scikit-learn, Pandas, NumPy

Visualization & Business Intelligence | Power BI, Tableau, Looker Studio, Data Visualization, Dashboard Development, MS Excel, Seaborn, Matplotlib

Tools & Platforms | Linux (Ubuntu), Jupyter Notebook, RStudio, Canva, Lucidchart, Visual Paradigm

PROJECTS

Daraz Sentiment Analysis Model

- Developed an **NLP-based Sentiment Analysis** model using 22,000+ **Daraz** customer reviews to evaluate customer satisfaction and product perception.
- Extracted, cleaned, and **processed 22K+ records** using Python, applying text preprocessing, feature extraction, and sentiment classification techniques.
- Generated actionable insights from customer feedback, supporting data-driven analysis of consumer trends and purchasing behavior.

Plant Disease Detection using Computer Vision

- Developed a **deep learning model** using CNN to detect plant diseases from leaf images (sourced through **kaggle**).
- Performed image preprocessing and classification using computer vision techniques.

Multivariate Analysis of Social Influence

- Conducted **multivariate statistical analysis** to examine relationships between demographic factors, conformity, and autonomy.
- **Applied Python**, data visualization, and **correlation analysis** to identify significant patterns and behavioral trends.
- Generated research-driven insights supporting evidence-based conclusions and academic findings.

FYP – AI Based Traffic Violation Detection System

- Developed an AI-based Traffic Violation Detection System using computer vision and **deep learning** techniques to automatically identify traffic rule violations.
- The project was designed to address the limitations of the existing **Safe City infrastructure** by improving detection accuracy and automation.
- The system detects violations such as helmet absence, seatbelt non-compliance, and pillion riding from near to real time traffic footage for smarter traffic monitoring using **YOLOv8**.
- Achieved an overall model **accuracy of 85%**, demonstrating performance comparable to existing Safe City infrastructure systems.

CERTIFICATIONS

- Introduction to Big Data with Spark and Hadoop Coursera
- Introduction to Data Visualization with Power BI Datacamp

ACHIEVEMENTS

- Exhibited Sentiment Analysis Project at DataFest 2024 Pakistan Bureau of Statistics
- Formed a Data and AI Society Air University, Pakistan

PUBLICATIONS

D. Hamid, C. A. Ul Hassan, A. U. Yasin, S. Shabir, T. Akram and M. A. Aslam, "Retaining the Best: How AI and IoT Can Combat Employee Turnover," 2024 2nd International Conference on Computing and Data Analytics (ICCDAA), Shinas, Oman, 2024, pp. 1-7, doi: 10.1109/ICCDAA64887.2024.10867326.