Abhishek Kumar Harendra Gupta

Vasai, Palghar, Mumbai, Maharashtra| +91-7678041491 |

1289shek2004@gmail.com GitHub | LinkedIn

Career Objective

Aspiring Data Scientist with hands-on experience in Python, R, SQL, visualization, and ML. Passionate about data-driven decision-making, social impact, and continuous learning. Actively engaged in realworld projects, remote internships, and research. Seeking opportunities to contribute to data-driven innovation.

Education Bachelor of Science in Information Technology

Shree L.R. Tiwari Degree College, University of Mumbai

Expected Graduation: 2026

Internship & Training Experience Data Science Intern Evoastra Ventures | *May 2025 – Present* Location: Remote/Hydrabad

- Engaged in end-to-end development of data science and AI projects involving web scraping, deep learning, computer vision, and natural language processing.
- Collaborated with mentors and cross-functional teams to deliver practical solutions aligned with real-world business use cases.
- Applied Python, machine learning, and deep learning frameworks to extract insights, automate workflows, and build scalable models.

Key Projects:

- Web Scraping Car Details: Scraped and cleaned structured data (make, model, price, year, mileage) from online car listings using BeautifulSoup and Pandas.
- Image Captioning Using Deep Learning: Developed a CNN-LSTM-based image captioning model using Keras, generating meaningful descriptions for images.
- Mice Classification via Protein Expression: Built classification models (Random Forest, SVM) on biomedical data to classify mice into 8 groups based on protein expression profiles.
- Facial Recognition System: Created a secure facial recognition solution using OpenCV and deep learning for employee identification in real time.

- Automated Resume Analysis Tool (Ongoing): Building an NLP tool using spaCy and Scikit-learn to extract candidate info and match resumes with job descriptions.

Technologies: Python, Pandas, NumPy, Scikit-learn, TensorFlow, Keras, OpenCV, BeautifulSoup, Matplotlib, Seaborn, spaCy, NLTK, Streamlit, Git

Data Associate L1 – Data Science & ML (Accepted)
Infotact Solutions (Remote)

May 2025 -- Present

- Built EDA & ML pipelines, documented code, and dashboards
- Performance-based stipend and recognition opportunities

Data Analyst Intern — (Virtual Offer Accepted)
Elevent labs (Remote)

Jun 2025 – present (4 weeks)

- Selected for internship to assist in building dashboards and analyzing ecommerce performance data
- Expected tools: Excel, SQL, Power BI, Python

Data Analytics Trainee offered (Rejected)
MedTourEasy (Remote) 4 weeks

Certifications ✞ Essentials of Data Science with R Software – Module 1
IIRS-ISRO/NPTEL – Jan to May 2024
Certificate: Essentials Of Data Science With R Software - 1_ Probability And Statistical Inference.pdf
[NPTEL ID: NPTEL24MA18S1162300082]

✞ **Green Skills & Artificial Intelligence**
Edunet Foundation + IBM Skills Build + AICTE + Shell India – Jan to March 2025
Certificate: Edunet_workshop_certificate.pdf
Details: Participated in a data analytics workshop powered by IBM SkillsBuild.

✞ **DLLE Extension Work** – *University of Mumbai*
Participated in 120+ hours of Annual Extension Activities in 2024–2025, covering:
Information Technology & Industry Orientation

Entrepreneurship & Career Guidance
Women Empowerment
Population Education & Open Schooling.

Projects 1. Mice Protein Classification using Machine Learning

- Built a classification model to identify experimental treatment groups based on protein expression levels in mice.
- Performed detailed preprocessing, feature selection, and EDA on a biomedical dataset.
- Applied multiple models including Random Forest, SVM, and Gradient Boosting; optimized using Grid SearchCV.
- Achieved high accuracy and F1-score; results were interpreted using confusion matrix and feature importance plots.

2. CO_2 Emission Analysis Project

- Conducted an in-depth exploratory analysis of global CO_2 emission trends across countries and sectors.
- Cleaned and merged multi-source datasets to create comprehensive visualizations using Matplotlib and Seaborn.
- Identified top emitting nations and year-over-year trends to support environmental policy discussions.
- Built regression models to understand the relationship between GDP and CO_2 output.

3. Customer Segmentation using RFM and K-Means

- Segmented customers of a retail dataset based on Recency, Frequency, and Monetary value using RFM analysis.
- Applied K-Means clustering to group customers and identify high-value and at-risk segments.
- Created interactive visualizations using Power BI and Python for marketing team insights.
- Suggested targeted marketing strategies based on segment behaviour.

4. Food Waste Management and Reduction Analysis

- Analyzed country-wise food waste data from households, retail, and food services.
- Visualized disparities in waste levels using heatmaps and bar plots.

- Built classification models (Decision Tree, Logistic Regression) to identify key food waste drivers.
- Proposed data-driven recommendations to optimize food systems and reduce global waste. (Explore more at: GitHub)

Technical Skills

- **Languages:** Python, R, SQL, Machine Learning, AI
- **Libraries:** Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn, TensorFlow, Streamlit
- **Tools:** Anaconda, Power BI, VS Code, Tableau, Excel, Jupyter, RStudio, UiPath Studio
- **DBs:** MySQL, SQLite, PostgreeSQL, SQL Server
- **Other:** Basic Web Designing, MS Office, Git & Github, Scilab, Weka Software

---

Achievements & Participation

- **Organized "Data Quest" - Inter-College ML Model building Competition.**
- **Completed research project under DLLE on “Education of Specially-Abled Students”.**
- **Participated in college fests like Technova 2k25, Carpe-Diem 2.0 · Volunteered in college fests like Udaan, Carpe-Diem 2.0**

Languages

- English – Professional
- Hindi – Native
- Marathi – Conversational