

# Ankit Tiwari

Data Science & ML Developer

 [unkit.site](http://unkit.site)  [github/anktw](https://github.com/anktw)  [kaggle/ankittfzd](https://kaggle.com/ankittfzd)  +91 789789xxxx  at79xxxx@gmail.com  Ayodhya, India

## EDUCATION

---

### Bachelor of Technology (Computer Engineering)

2022 – 2026

S.I.E.T Nilokheri, Haryana afw kurukshetra University, Kurukshetra 

CGPA: 7.5

## PROJECTS

---

### House Price Prediction

08/2025 – 09/2025

*Python, NumPy, Pandas, Seaborn, scikit-learn*

- Predict house prices to improve valuation decisions for real-estate.
- This is a ML pipeline that cleans and imputes **70+ features**, encodes categorical variables, scales numerical data, compares 10 regression models **using 7-fold CV**, and deploys a **Gradient Boosting Regressor** for final price prediction.
- Achieved **high R<sup>2</sup> performance via 7-fold cross-validation**, enabling reliable automated price predictions
- Reducing manual valuation effort by **~80%+** through model-driven estimation.

### Handwritten Digit Classification Using CNN

09/2025 – 10/2025

*TensorFlow*

- Recognizes handwritten digits automatically.
- **Trained on MNIST** by applying image preprocessing, convolution-pooling feature extraction, dense classification layers, and deployed a custom prediction function for classifying new handwritten digit images.
- Achieved **~99% test accuracy** with real-time image inference and confidence visualization.

### RAG-Based Patient Review Question-Answering Chatbot

10/2025 – 11/2025

*Generative AI + LangChain + Gradio*

- An AI system that helps hospitals instantly analyze patient reviews and answer operational queries using retrieved real-world feedback.
- Embeds hospital reviews, performs vector similarity search, and uses Gemini to generate context-aware answers with a Gradio chatbot interface.
- **Achieved 92%+ relevant response accuracy**, reduced review-analysis time by **~85%**.

### AutoEDA

08/2025 – Present

*Python, Pandas, NumPy, Streamlit, Seaborn, Scikit-learn*

- A tool for Automated Exploratory Data Analysis for quickly understanding datasets and reduce time spent EDA tasks.
- A modular Streamlit application that auto-detects categorical/numerical columns, computes summary statistics, **generates interactive plots** and provides in-app preprocessing (**missing-value strategies**, encoding, scaling, outlier detection & handling).
- Cut typical manual **EDA time by 60%**, **sped up preprocessing workflows by 2–3x** through GUI automation, and immediate export of cleaned datasets.

## CERTIFICATES

---

- Data Science & Machine Learning (GeeksForGeeks)

- React Foundations (Vercel) 

- Python Programming with DS&A(YBI Foundations) 

- Next.js App Router Fundamentals (Vercel) 

## INTERESTS

---

Quantitative Research, Systems Programming, Distributed Computing, Open source.