

# Aayush Jha

aayushjha2431@gmail.com | +91 9987054087 | Dharwad

## CAREER OBJECTIVE

2nd-year B.Tech (CSE) student at IIIT Dharwad, aspiring ML Engineer. Skilled in Python, data cleaning, exploratory data analysis (EDA), and ML model building. Looking for Data Analyst / ML Internships (remote/part-time) to apply skills and grow in AI/ML.

## EDUCATION

### B.Tech,

Indian Institute Of Information Technology Dharwad

2024 - 2028

## TRAININGS / CERTIFICATIONS

### Intro To Machine Learning

Jul 2025

Kaggle, Virtual

Learned fundamentals of supervised learning, model evaluation, and data preprocessing.

### Data Visualisation: Empowering Business With Insights

Jul 2025

TATA (Forage Virtual Program), Virtual

Practical tasks in framing business scenarios, choosing visuals, and communicating insights effectively.

### Data Analytics & Forensic Technology Job Simulation

Jun 2025

Deloitte (Forage Virtual Program), Virtual

Completed simulation tasks on data analysis, forensic investigation, and real-world problem-solving.

### Prompt Design In Vertex AI

Jul 2025

Google Cloud, Virtual

Skill badge covering prompt engineering, LLM-based solutions, and AI workflows in Vertex AI.

### Pandas

Jul 2025 - Sep 2025

Kaggle, Virtual

Gained hands-on experience with pandas for data cleaning, merging, grouping, and exploratory data analysis.

## PROJECTS

### [Data Cleaning & Preprocessing ↗](#)

Aug 2025 - Sep 2025

- Cleaned and preprocessed raw datasets: handled missing values, removed duplicates, standardized categorical values and fixed data types.
- Implemented functions for bulk Excel/CSV processing and automated common file operations

### [Classification Models ↗](#)

Aug 2025 - Sep 2025

- Worked with Iris dataset.
- Implemented Different Classification Models - Logistic Regression, KNN, Decision Tree, SVM
- Evaluated the confusion matrix and different scores

### [EDA for titanic dataset ↗](#)

Aug 2025 - Sep 2025

- Did the Exploratory Data Analysis for the dataset Titanic.
- Found all the insights from the dataset

### [learnt and implemented Ensemble Learnings ↗](#)

Sep 2025

- Used a real life dataset - Bangalore EEG Epilepsy Dataset (BEED)
- Cleaned the dataset and did an EDA before implementing the models
- Used sklearn.ensemble for - Random Forest and Gradient Boosting
- Still ongoing for the integration of Streamlit

## Housing Dataset Price Prediction ↗

Aug 2025 - Present

- Implemented Linear Regression (with and without Polynomial Features) using scikit-learn.
- used dataset from sklearn.datasets
- Applied Pipeline, evaluated using: MSE and R square

### SKILLS

- |                 |                    |                 |
|-----------------|--------------------|-----------------|
| • Python        | • Pandas           | • NumPy         |
| • Matplotlib    | • Jupyter Notebook | • Data Cleaning |
| • Data Analysis | • Git              | • GitHub        |
| • scikit-learn  | • Machine Learning |                 |