

# Harsh Kumar

✉ [harsh.k22@iiits.in](mailto:harsh.k22@iiits.in) | **in** [Harsh Kumar](#) | **o** [harshrajput4343](#)

## EDUCATION

### Indian Institute of Information Technology, Sri City

Bachelor of Technology - Electronics & Communication Engineering; CGPA: 7.5

### D.A.V public School

CBSE(X); Percentage: 91.4%

Chittoor, India

Nov 2022 - June 2026

India

April 2009 - April 2019

## EXPERIENCE

### Research Assistant — IIIT Sri City

Aug 2025 – Nov 2025

Data Scientist — Hybrid Intrusion & Anomaly Detection

- Built end-to-end data science pipelines including secure data extraction, cleaning, feature engineering, and dataset validation for large-scale IoT and network traffic.
- Applied statistical analysis, machine learning, and deep learning models (CNNs, attention networks, XGBoost) to detect intrusions and anomalous behavior across high-dimensional telemetry data.
- Optimized preprocessing strategies, evaluation metrics, and decision thresholds to reduce false positives and improve classification reliability in operational environments.
- Tracked experiments, model versions, and performance using MLflow to ensure reproducibility, auditability, and scientific rigor in all results.

### CodeClause

Dec 2025 – Jan 2026

Data Science Intern

- Built Python-based data pipelines for cleaning, transforming, and preparing real-world datasets for machine learning workflows.
- Performed exploratory data analysis (EDA) and feature engineering to identify patterns, reduce noise, and improve model readiness.
- Trained and evaluated supervised learning models, producing reproducible notebooks and technical reports for business-driven analytics.

## PROJECTS

### Waste Detection System

(GitHub )

End-to-End Computer Vision Pipeline

- \* Built a production-grade waste detection system using **YOLOv5** with dataset annotation, preprocessing, augmentation, and model training for real-time object detection.
- \* Designed a modular **computer vision pipeline** in Python integrating OpenCV-based image processing, model inference, and post-processing for bounding-box filtering and confidence scoring.
- \* Optimized model performance through hyperparameter tuning and data augmentation, improving detection accuracy and generalization across varied waste categories.
- \* Containerized the entire pipeline using **Docker** and deployed it on **AWS** and **Azure** using CI/CD workflows for scalable and reproducible inference.
- \* **Technologies:** YOLOv5, Python, OpenCV, Docker, AWS, Azure, CI/CD

### Book Recommender System

(GitHub )

Collaborative Filtering Based Recommendation Platform

- \* Built a collaborative filtering-based recommender system using **Scikit-learn** and nearest-neighbor similarity to generate personalized book recommendations from user-item rating matrices.
- \* Engineered features by filtering high-activity users and books (200 books / 50 ratings threshold), merging Kaggle datasets, and preprocessing sparse data for reliable similarity computation.
- \* Developed an interactive **Streamlit** web application to serve real-time recommendations and deployed it on **AWS** using **Docker** for scalable access.
- \* **Technologies:** Python, Scikit-learn, Streamlit, Docker, AWS, Kaggle, GitHub

### AI-Powered Medical Chatbot

(GitHub )

LLM-Driven Healthcare Assistant with RAG and API Deployment

- \* Built an LLM-powered medical assistant using GPT-4 and Gemini with prompt engineering and retrieval-augmented generation over Pinecone vector search.
- \* Designed LangChain pipelines for document ingestion, semantic retrieval, and context-aware medical response generation.
- \* Implemented Flask-based inference APIs with caching, logging, and error handling for scalable deployment.
- \* **Technologies:** Gemini, GPT-4, LangChain, Pinecone, Hugging Face, Flask, AWS, GitHub Actions

## TECHNICAL SKILLS

**Languages:** Python, C/C++, JavaScript, TypeScript, SQL

**DS & ML:** EDA, Feature Engineering, Data Preprocessing, Classification, Regression, Imbalanced Learning, Cross-Validation, Hyperparameter Tuning, Model Evaluation

**ML Stack:** Scikit-learn, XGBoost, LangChain, Hugging Face, Transformers, LLMs (GPT-4, Gemini, Azure OpenAI), RAG

**Databases & Vector Stores:** PostgreSQL, MySQL, MongoDB, Pinecone, ChromaDB, Weaviate

**MLOps & Deployment:** FastAPI, REST APIs, Docker, CI/CD, MLflow, Model Deployment, Experiment Tracking

**Cloud & Tools:** AWS, Azure, Git, GitHub Actions, Streamlit, Power BI, NumPy, Pandas, Matplotlib, Seaborn

## ACHIEVEMENTS & CERTIFICATIONS

- Oracle Cloud Infrastructure 2025 Certified Generative AI Professional [Link](#)
- Complete Data Science, Machine Learning, Deep Learning, NLP Bootcamp — Krish Naik [Link](#)
- Solved 250+ problems on LeetCode/GfG with focus on algorithms, data structures, and problem-solving.

**IIIT SriCity Cricket Team Captain**

September 2024 - nov 2025

*IIIT Sri City Intercollegiate Team*

- \* Captained the IIIT Sri City cricket team in intercollegiate tournaments, leading match strategy and on-field performance.

**Events Coordinator at Abhisarga**

March 2024 - December 2024

*Annual College Technical Festival*

- \* Managed technical event logistics for 1,500+ participants across 15+ competitions, ensuring 100% on-time execution.