

# Santosh Chapagain

+977-9761835424 | santosh.chapagain.datasci@gmail.com | <https://codewithsantosh.tech>  
<https://github.com/Santosh-Chapagain> | <https://www.linkedin.com/in/santosh-chapagain-041629259>

Data Scientist with strong expertise in Python, machine learning, and data analysis, currently pursuing a Bachelor's degree in Computer Science and Information Technology. Passionate about Generative AI and applying advanced AI techniques to solve complex problems and generate actionable insights.

## Skills

- Python, Machine Learning, Deep Learning, GenAI, SQL, HTML, CSS, Fast API, [Basics: C, C++]
- Tools: Pandas, NumPy, Matplotlib, Scikit-Learn, TensorFlow, Lang Chain, Lang Graph, MySQL

## EDUCATION

- Far-Western University | Dreams College [2024 - 2028]
- XII (NEB) | Gurukul College [2022 - 2024]  
3.48 GPA

## EXPERIENCE

### Tech Lead – Dreams IT Club [Jan 2026 - Present]

- Led the Dreams IT Club technical team
- Organized events (Workshops, Tech Competitions)

### AI Workshop Instructor [Dec-2025 – Dec-2025]

- Launched AI workshop event with hands-on coding using different LLM models.

## ACADEMIC PROJECTS

### Teaching RAG AI

GitHub: [https://github.com/SantoshChapagain/Rag\\_Based\\_Teaching\\_Assistant\\_AI](https://github.com/SantoshChapagain/Rag_Based_Teaching_Assistant_AI)

- Powered by RAG.
- Helps learners to quickly find the exact timestamps in teaching videos where a concept is introduced and where it concludes.

### YouTube AI Chatbot

GitHub: [https://github.com/Santosh-Chapagain/Youtube\\_AI\\_Chatbot](https://github.com/Santosh-Chapagain/Youtube_AI_Chatbot)

- Detect other language and translate to English.
- Processes them with Ollama model, and provides detailed answers using retrieval-augmented generation (RAG).

### House Price Prediction

GitHub: <https://github.com/Santosh-Chapagain/House-Price-Prediction>

- Developed a robust ML pipeline for house price prediction, testing multiple models (Decision Tree, Random Forest, Lasso, Gradient Boosting, Ridge) and selecting the **best-performing Linear Regression**, achieving  $R^2 \approx 0.85 \pm 0.05$  on 5-fold cross-validation.

## ACHIEVEMENTS/CERTIFICATION

- OCI Data Science Certification – Oracle Race Competition 2025
- Best UI/UX Winner – Ideathon
- Associated Data Scientist Certificate