

# Santosh Chapagain

Chitwan, Nepal | +977-9744601007 | santosh.chapagain.datasci@gmail.com

[Portfolio](#) | [GitHub](#) | [LinkedIn](#)

Data Scientist with experience building ML pipelines in Python using Pandas, Scikit-Learn and TensorFlow. Developed regression and RAG-based systems with measurable performance gains. Focused on model deployment and real-world AI applications.

## Skills

- Python, Machine Learning, Deep Learning, GenAI, SQL, Fast API, [Basics: C, C++]
- Tools: Pandas, NumPy, Matplotlib, Scikit-Learn, TensorFlow, Lang Chain, Lang Graph, MySQL
- Techniques: Feature Engineering, Cross-Validation, Hypothesis Testing, A/B Testing, Model Deployment, Data Visualization, RAG, LLMs

## EDUCATION

- Bachelor in Computer Science and Information technology [2024 – 2028]  
Far-Western University | Dreams College
- XII (NEB) | Gurukul College [2022 – 2024]  
3.48 GPA | A Grade

## EXPERIENCE

### Tech Lead – Dreams IT Club [Dec 2026 – Present]

- Led and mentored a 5+ technical team, driving collaborative project development and strategic planning.
- Organized and managed 3+ technical events including workshops and competitions, coordinating cross-functional teams.

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

- Trained 50+ students on AI workshop, introducing participants to Large Language Models (LLMs) and practical AI applications.

## ACADEMIC PROJECTS

### Teaching RAG AI | [\[GitHub\]](#)

- Designed and implemented a RAG-powered assistant that identifies exact timestamps in teaching videos, reducing search time for learners.

### YouTube AI Chatbot | [\[GitHub\]](#)

- Developed a multilingual chatbot that detects multiple languages, translates to English, and provides detailed answers using RAG & Ollama models

### House Price Prediction | [\[GitHub\]](#)

- Built ML pipeline testing Decision Tree, Random Forest, Lasso, Gradient Boosting, and Ridge; best **Linear Regression achieved  $R^2 \approx 0.85 \pm 0.05$**  on 5-fold cross-validation.

## ACHIEVEMENTS/CERTIFICATION

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