



# Certification in Data Science and Artificial Intelligence Program

from E&ICT Academy, IIT Roorkee

In association with **masai**

# About E&ICT Academy, IIT Roorkee

The Electronics & ICT Academy (E&ICT), supported by MeitY, Govt. of India, aims to bridge the gap between industry demand and academic knowledge by delivering specialized, hands-on training and upskilling programs in emerging areas of the Information & Communication Technology and Electronics sector.

- **The Indian Institute of Technology (IIT) Roorkee:** A legacy of over 175 years. Established in 1847 as India's first engineering college, it became an IIT in 2001, excelling in engineering and technology.
- **Rankings:** IIT Roorkee consistently ranks among the top engineering and research institutions in India. In the NIRF India Rankings 2024, IIT Roorkee was ranked #6 in the 'Engineering' category and #8 in the 'Overall' category.
- **Strong Industry & Research Ecosystem:** With active collaborations, incubators, and innovation hubs, IIT Roorkee bridges academia and industry to drive real-world impact.
- **Industry-Focused Learning:** IIT Roorkee maintains strong ties with industries across its campus locations, providing students with opportunities for internships, real-world projects, and networking with business leaders.



# Why Choose This Course?

- **Prestigious Certification:** Receive a Certificate of Completion from E&ICT Academy IIT Roorkee, recognizing your achievement.
- **Campus Immersion:** An optional 2-day Campus Immersion for direct interaction with industry experts and peers.
- **Future-Proof Career Gateway:** Step into the rapidly evolving AI landscape by mastering Data Science and AI with hands-on expertise guided by top IIT faculty.
- **Advanced Curriculum:** Access cutting-edge business analysis content, engaging simulations, and practical evaluations. Focus on real-time project implementation for hands-on mastery.
- **Practical-Based Learning:** Work on hands-on projects and real-world AI scenarios, applying NLP and multi-agent frameworks to solve complex, industry-relevant problems.
- **Placement Opportunities:** Receive resume reviews, career coaching, and placement support to land your dream role. (Eligibility: 65% attendance & overall 70% marks; offered by Masai)

*\*Eligibility: Minimum 65% overall attendance and 35% in the cumulative evaluation score*

## What Will You Learn?

Drive your career forward by mastering Data Science and Machine Learning fundamentals. Learn to extract deep insights from complex data, build accurate predictive models, and become a portfolio-ready professional equipped to drive data-driven decision-making in any industry.

### Toolkit



python™



matplotlib



pandas



NumPy

# Course Details

**Course Duration**  
6 Months

**Time Commitment**  
8-10 hours per week

**Certification**  
E&ICT Academy, IIT Roorkee

## Course Curriculum

### Module 1: Programming & Data Foundations

- Python Fundamentals: Build comfort with variables, data types, conditionals, loops, functions, and core data structures to write programs that manipulate real data
- Notebook Workflow: Use Jupyter/Colab to run experiments, document analysis, and package work as shareable deliverables
- Working with CSV & JSON: Read/write structured data, validate schemas, and fix common issues like missing values, wrong types, and inconsistent formats
- API Fundamentals: Use Python to fetch real-world data via REST APIs, handle status codes, authentication keys, pagination, and convert responses into usable datasets

### Module 2: Data Analysis with NumPy & Pandas

- NumPy Essentials: Use arrays and vectorized operations for fast computation and numeric manipulation
- Pandas Core Operations: Perform filtering, sorting, grouping, aggregation, joins/merges, pivots, and date/time handling on datasets
- Data Cleaning & Preparation: Build repeatable workflows for missing values, duplicates, outliers, inconsistent categories, and messy text fields
- Exploratory Data Analysis (EDA): Ask structured questions, compute key metrics, and discover patterns that guide decisions or modeling

### Module 3: SQL, Visualization & Data Storytelling

- SQL Fundamentals: Query relational data using joins, aggregations, subqueries, and filters to extract analysis-ready tables
- Applied SQL for Analytics: Use window functions for ranking, cohort-style analysis, and running totals
- Visualization for Decision-Making: Learn chart selection, avoid misleading visuals, and build clear plots that answer real questions
- Storytelling with Data: Convert analysis into a structured narrative with context, evidence, insights, and recommendations

## Module 4: Maths & Statistics for Data Science (Applied)

- Math Foundations Used in DS: Learn transformations, scaling, logs/exponentials, vectors, and similarity intuition through practical examples
- Probability & Distributions: Build intuition for randomness and variability, and how distributions show up in real datasets
- Sampling & Estimation: Understand why samples mislead, how uncertainty arises, and how to reason about reliability of insights
- Hypothesis Testing & A/B Testing: Learn the logic behind experiments and how to avoid false conclusions
- Correlation vs Causation: Learn how misleading relationships happen and how to communicate limitations responsibly

## Module 5: Machine Learning

- ML Workflow & Problem Framing: Convert questions into ML problems, define success metrics, and set up training/validation/testing correctly
- Regression Models: Build and evaluate regression systems, interpret errors, and improve results with better features and preprocessing
- Classification Models: Build classification systems and evaluate with precision/recall/F1, confusion matrices, and threshold tuning
- Tree-Based Models: Train decision trees and understand ensemble intuition (random forests / gradient boosting) for stronger real-world performance
- Feature Engineering & Pipelines: Create robust preprocessing + modeling pipelines that are reusable and less error-prone
- Model Improvement & Debugging: Detect leakage, handle imbalance, perform basic tuning, and run error analysis to systematically improve

## Module 6: GenAI for Data Science

- Prompt Engineering for Applications: Design prompts that follow instructions reliably, produce structured outputs, and handle edge cases with consistency
- LLM APIs with Python: Integrate LLMs into Python workflows, manage failures gracefully, and design clean input-output contracts
- Embeddings & Vector Search: Convert text into embeddings, build semantic search, and retrieve relevant context from large document sets
- RAG (Retrieval-Augmented Generation): Build grounded Q&A systems by combining retrieval with generation, and reduce hallucinations through chunking and retrieval strategy
- Unstructured Data to Insights: Extract themes, summaries, and actionable insights from documents like PDFs, policies, support tickets, and internal knowledge bases

## Capstone Project

- Build & Evaluate: Deliver a working solution with measurable outcomes and documented trade-offs

# Projects

## Product-Image Classifier (CNN)

Flipkart's catalog team wants to auto-categorise product images into Apparel, Electronics, and Home. You train a small CNN on a public Indian e-commerce image dataset using free Colab GPU, push validation accuracy past 85%, and produce a confusion-matrix breakdown that names exactly which category pairs the model still fumbles – and why.

Computer Vision • TensorFlow / PyTorch • CNN Architecture • Colab GPU • Confusion-Matrix Analysis

### Decision Output

Trained classifier with documented confusion patterns the catalog team can route to manual review.

## Customer-Churn Predictor & API

Build a churn-prediction model for a D2C customer dataset. Engineer RFM features (recency, frequency, monetary), train a gradient-boosting classifier, deploy as a FastAPI endpoint, and write the model card – including the ethical considerations around retention targeting that any real product team will probe in your interview.

XGBoost • RFM Feature Engineering • FastAPI Deployment • Model Card • Ethical Framing

### Decision Output

A deployable churn predictor with a documented model card you can defend to a hiring panel.

*Please Note: "Due to the evolving nature of the industry expectations and partner institute feedback, some syllabus aspects may change. Any updates will be communicated during the Inauguration Session(s) or at the start of the relevant module".*

*The project deliverables or case studies for your batch may differ, as per the decision of the Institute or Course Faculty*

## Under the Guidance of



**Prof. Sanjeev Manhas**

Principal Investigator of E&ICT Academy, IIT Roorkee

Prof. Sanjeev Manhas, Principal Investigator of E&ICT Academy, IIT Roorkee, is a distinguished academician, researcher, and innovator with extensive experience in academia, research, and industry. He holds a Ph.D. in Electronics and Computer Engineering from De Montfort University, UK, and an M.Tech. from IIT Madras. He has worked at Micron Semiconductor and the Institute of Microelectronics, Singapore, contributing to advanced DRAM and CMOS process integration. His research spans nanoscale devices, device-circuit co-design, novel memories, MEMS, sensors, and CMOS process technologies.

# Meet our Instructor Team



## **Prof. Tharun Kumar Reddy Bollu**

**Assistant Professor, IIT Roorkee**

Machine Learning | Brain-Computer Interfaces | AI Healthcare

PhD from IIT Kanpur; ex-IIT Dhanbad and AI Architect in USA; published in leading IEEE journals.



## **Prof. Pawan Goyal**

**Professor, IIT Kharagpur**

10+ years | Natural Language Processing | Machine Learning | AI Systems

PhD from IIT Kanpur with publications in leading journals and conferences; recognised for contributions to AI and NLP research.



## **Sriram Desai**

**Senior Software Engineer, PayPal**

Backend Systems | Financial Platforms | Distributed Systems

Backend engineer at PayPal; ex-ByteDance, Amazon, Prime Video; built financial and content systems with scaling and compliance expertise.



## **Shubhendu Shishir**

**Head of Engineering, Simplismart**

GenAI Infrastructure | Distributed Systems | Payments

Leads GenAI infrastructure at Simplismart; ex-Amazon, Uber; built conversational agents, catalog systems, and instant payment systems.

*\*Note: Instructors teaching your batch may change based on specific module expertise, or circumstances such as personal emergencies, unavailability etc.*

# Admission Process



## Clear Qualifier Test

Clear the qualifier test to be eligible for the program



## Complete Selection & Enrolment Session

Only shortlisted candidates go through the counselling process



## Secure your seat

Learn from India's top educators and stand out from the crowd

# Fees Structure

<b>Application Fee</b> Non-refundable	<b>₹99</b>
--	------------

	Option 1	Option 2
	<b>Upfront</b>	<b>EMI</b> (Through our NBFC partners)
<b>Registration Fee</b> Non-refundable	<b>₹4,000</b>	<b>₹4,000</b>
<b>Program Fee</b> Non-refundable	<b>₹48,000</b>	<b>₹6,133 x 9 months</b>
<b>Total Fees</b> Non-refundable	<b>₹52,000 + GST</b>	<b>₹59,197 + GST</b>

\*GST at 18% extra, as applicable

**In association with masai.**

---

WhatsApp: +91 8792974750

Email: [eictitrprograms@masaischool.com](mailto:eictitrprograms@masaischool.com)