

# Understanding Classification & Clustering

Hands-on with Teachable Machine & Orange Data Mining

Dr. Ebin Deni Raj

FACTS-H Lab, IIIT Kottayam



# Why This Workshop Matters

Data in humanities and social sciences is exploding:

- interview transcripts
- survey responses
- student feedback
- digital archives
- social media behaviour

**How do we understand this data without coding?**



# The Opportunity

Machine Learning today is not only for engineers.

In this workshop you will learn to:

- Build simple ML models using visual, no-code tools
- Classify text, images, or audio meaningfully
- Discover clusters and themes in documents
- Analyse data visually and intuitively



# Before We Begin: Myths vs Realities of ML

## Common Myths

- “ML is always accurate and objective.”
- “ML understands like humans do.”
- “More data automatically means better results.”
- “ML will replace human judgement.”



# Before We Begin: Myths vs Realities of ML

## Common Myths

- “ML is always accurate and objective.”
- “ML understands like humans do.”
- “More data automatically means better results.”
- “ML will replace human judgement.”

## Realities

- ML models learn patterns *only from the data they see.*
- ML is a tool, not intelligence.
- Poor data → biased results.
- Humans remain responsible for interpretation.



# Why Responsible ML Matters

Even simple models can cause:

- misclassification of people or communities,
- amplification of cultural bias,
- incorrect interpretations in research,
- flawed decision-making in classrooms or surveys.

**Responsible use of ML is essential — especially in critical domains.**



# Tools We Will Use

- **Teachable Machine** No-code tool from Google for image/audio classification.
- **Orange Data Mining** Visual workflows for classification and clustering.

These tools make ML accessible and insight-driven.



# Responsible Use of Teachable Machine

- Webcam-based classification can pick up biases.
- Lighting, background, skin tone, and angles matter.
- You are training the model — so your choices influence behaviour.

**Key Principle: “Better examples → fairer model.”**



# Responsible Use of Orange

- Text datasets carry cultural and language bias.
- Clustering may surface “themes” that are misinterpreted.
- Classification needs clean, well-labelled data.

**Orange does not decide truth. You do.**



# What You Will Be Able to Do After This Session

- Build your own classification model using your webcam
- Analyse text datasets visually (tweets, messages, reviews)
- Understand patterns without writing code
- Apply ML thinking to teaching, learning, or research

You will be ML-enabled — not ML-experts.



# Session Focus

By the end of this segment, you will:

- Have an intuitive idea of ML.
- Distinguish classification vs clustering clearly.
- See how these connect to TM and Orange.
- Be aware of responsible ML practices.



# What is Machine Learning?

## Working definition

ML builds systems that **learn patterns from data** instead of rules written by humans.

- Provide examples (data).
- Model learns a pattern.
- It predicts or groups new data.



# Two Fundamental Tasks

## Classification

- We already know the categories.
- Model learns to assign new data into them.

**Classification = Which label?**

## Clustering

- Categories are unknown.
- Model discovers meaningful groups.

**Clustering = Which group?**



# Responsible Reflection: Data Shapes Outcomes

- Classification reflects the labels you choose.
- Clustering reflects the structure hidden in your dataset.
- If the data is biased, your conclusions will be biased.

**Think critically: What story is your data telling — and what is it hiding?**



# Classification in Teachable Machine

- You show labelled examples (via camera/microphone).
- The model learns to distinguish between classes.
- Very sensitive to data quality and balance.

**More diverse examples → more robust model.**



# Clustering in Orange

- Algorithm groups items based on similarity.
- Interpretation is key: clusters do not have inherent meaning.
- You decide what each cluster represents.

**Clustering = discovery, not truth.**



# Connecting ML Tasks to Your Work

- Classify feedback sentiment in classrooms
- Automatically group interview transcripts into themes
- Identify patterns in survey responses
- Detect anomalies or unusual responses

ML becomes a lens for understanding — not a replacement for expertise.



# Ethical Pause: Avoid Overclaiming

- ML cannot detect emotions with absolute accuracy.
- ML cannot determine intent or morality.
- ML cannot understand culture or context the way humans do.
- ML cannot replace deep reading or qualitative interpretation.

**Use ML to support, not substitute, human judgement.**



# Group Activity: Apply + Reflect

Each group will:

- Select a mini-problem
- Decide whether it needs **classification** or **clustering**
- Build a workflow in TM or Orange
- Reflect on:
  - Why did the model behave this way?
  - What biases could be present?
  - How should a human validate the results?

Learning = Doing + Explaining + Questioning



# Questions?

Thank you.

We now move to live hands-on demonstrations.

