Training Data vs Testing Data

Student Details

Name : Menda Chandu

Roll Number : 23A51A4432 |

Department : CSD

Under the Guidance of DR K.Swaroopa, Professor of CSE(DS), AITAM,TEKKALI

Why Data Matters in Machine Learning

What is Machine Learning?

Teaching computers to learn from patterns without explicit programming

Why Split the Data?

Essential technique to build reliable, accurate models

The Foundation: Quality Data

Machines Learn from Data

Algorithms identify patterns and relationships

Quality = Accuracy

Better data leads to more reliable predictions

Poor Data Risks

Inaccurate data produces wrong results

Training Data: The Learning Phase

What It Does

Teaches the model by example

How It Works

- Learns patterns and relationships
- Builds internal model structure
- Analyses features and outcomes

Typical Size

70–80% of total dataset

Real-life analogy: Studying textbook and practice questions before an exam

Testing Data: The Evaluation Phase

1

Never Seen Before

Completely new examples the model hasn't encountered

2

Measure Performance

Tests how well the model generalises to real situations

3

Typical Size

20–30% of total dataset

Real-life analogy: Writing the actual exam paper under test conditions

Why Splitting Data is Crucial

01

Evaluate True Performance

Tests accuracy on unseen data

02

Prevent Overfitting

Stops model memorising instead of learning

03

Ensure Real-World Use

Validates reliability in practical applications

Real-World Example: Spam Detection

Training Phase

Labeled emails: spam or legitimate

Model learns patterns in headers, content, sender

Testing Phase

Never-before-seen emails

Classifies as spam or not spam

Key Differences at a Glance

Training Data

Learning phase

Builds the model

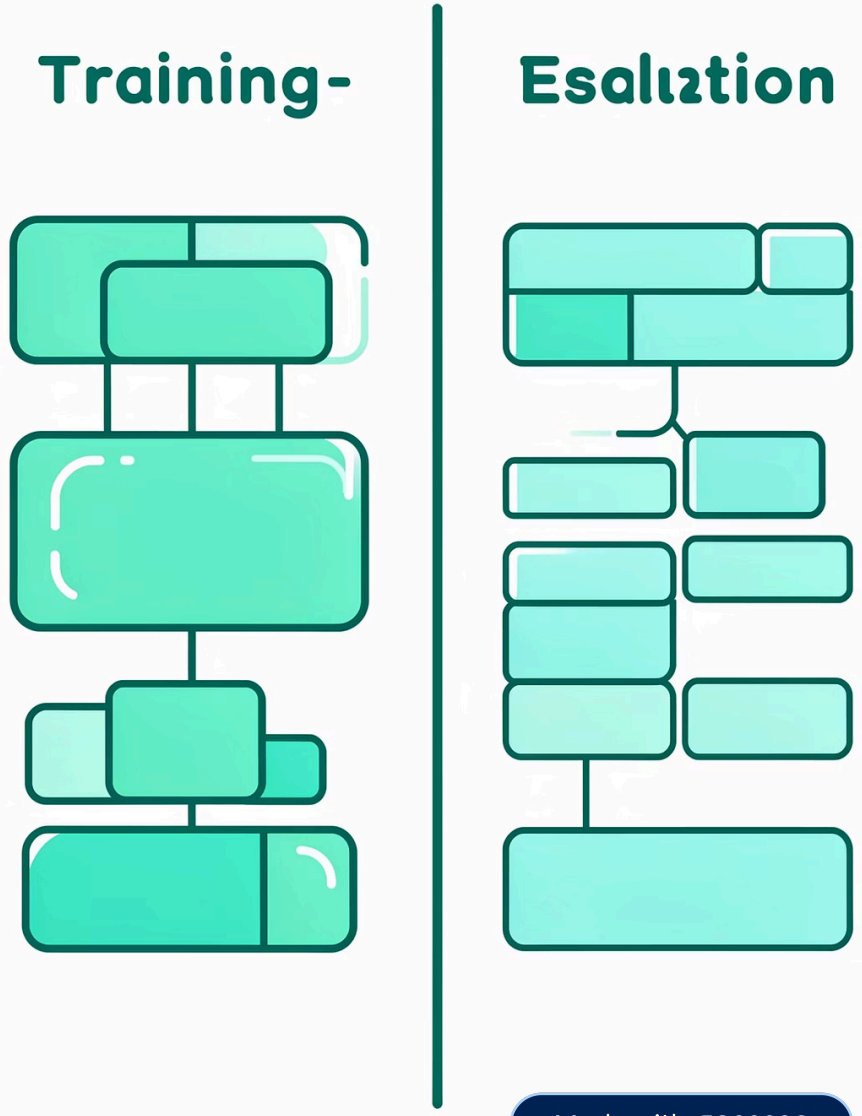
Large portion

Testing Data

Evaluation phase

Measures accuracy

Small portion



Key Takeaways

Both Are Essential

Training teaches, testing validates—neither works alone

Improves Reliability

Proper splitting ensures models work in real scenarios

Foundation Skill

Master this before exploring advanced ML techniques