

Week 3: ML Algorithms + Project 3 (IPL Prediction Model)

Topics Covered:

- Popular ML Algorithms explained with examples:
 1. **Linear Regression** (predicting house price with size)
 2. **Logistic Regression** (predicting pass/fail in exam)
 3. **Decision Trees** (asking yes/no questions to reach a decision, like “20 Questions” game)
 4. **Random Forest** (group of trees voting for answer)
 5. **KNN** (finding nearest friends with similar taste)
- Accuracy, Confusion Matrix, Overfitting explained with very simple terms.

Hands-on Project: IPL Match Winner Prediction Model

- Dataset: Past IPL matches data (teams, players, score, venue).
- Model: Decision Tree / Random Forest.
- Goal: Predict match winner.

1. Popular ML Algorithms (Explained Simply with Examples)

Machine Learning works using **algorithms** (methods/recipes).
Let's learn the most common ones with fun examples.

◆ (a) Linear Regression

- **What it does?** Finds a straight-line relationship between input and output.
- **Example:** Predicting house price using its size.
 - Bigger house → higher price.
 - Graph makes a straight line.



The more hours you study → the more marks you score.

◆ (b) Logistic Regression

- **What it does?** Predicts outcomes that are **Yes/No** (binary).
- **Example:** Predicting if a student will pass/fail in an exam.
 - Input: study hours
 - Output: Pass (1) or Fail (0)



If you studied 2 hours → fail. If you studied 10 hours → pass.
The model draws an “S-curve” instead of a straight line.

◆ (c) Decision Trees

- **What it does?** Makes decisions by asking step-by-step **Yes/No questions**.
- **Example:** Deciding what to eat.
 - Is it morning? Yes → Breakfast.
 - Is it evening? Yes → Dinner.

 **Student Example:**

It's like the game "**20 Questions**", where you keep asking Yes/No questions until you reach the answer.

◆ (d) Random Forest

- **What it does?** Uses **many Decision Trees together**. Each tree gives a prediction, and the majority wins.
- **Example:**
 - 10 friends guessing your exam marks.
 - One or two may be wrong, but if most say **75 marks**, that's the final answer.

 **Student Example:**

It's like **voting** in a class. One student may be wrong, but when many students vote, the result is more accurate.

- ◆ **(e) K-Nearest Neighbors (KNN)**

- **What it does?** Predicts something based on its nearest neighbors (similar cases).
- **Example:**
 - If your new classmate likes cricket, football, and gaming, he will probably join the “sports group.”

📌 Student Example:

Think of **finding friends** in college. You sit with those who have similar interests.

★ 2. Accuracy, Confusion Matrix, and Overfitting

✓ Accuracy

- Accuracy tells how many predictions were **correct**.
 - Example: If model predicts 8 out of 10 correctly → 80% accuracy.
-

✓ Confusion Matrix

A simple table that shows:

- **True Positive (TP)**: Model predicted “Diabetes: Yes” → Correct.
- **True Negative (TN)**: Model predicted “No” → Correct.
- **False Positive (FP)**: Model predicted “Yes” → Wrong.
- **False Negative (FN)**: Model predicted “No” → Wrong.

📍 Student Example:

It's like your teacher marking an answer sheet and checking how many are right/wrong.

✓ Overfitting

- When model **memorizes training data** instead of learning patterns.
- It performs well on training but badly on new data.

📍 Student Example:

If you **mug up** answers from the book (training), but in exam (testing), slightly new questions come, you fail. That's overfitting.

3. Hands-on Project: IPL Match Winner Prediction Model

Problem:

We want to predict which team will win an IPL match using past match data.

Dataset:

Contains details of past IPL matches like:

- Team 1 & Team 2
 - Players
 - Venue (stadium)
 - Toss winner
 - Runs scored
 - Match result (winner team)
-

Steps to Build the Model

1. Import Dataset
 - Load IPL dataset in Python.
2. Preprocess Data
 - Convert team names and venues into numbers (encoding).
 - Handle missing data.
3. Train/Test Split
 - Train on 80% of matches, test on 20%.

4. Choose Model

- Use **Decision Tree** or **Random Forest** (since they work well for categorical data like teams).

5. Train the Model

- Teach the model patterns from past match outcomes.

6. Test the Model

- Check accuracy on test matches.

7. Make Predictions

- Give new match details → Model predicts which team will win.
-

Example for Students

👉 Just like cricket fans guess “**Which team will win today?**” based on past performance (home ground, players, toss result), our ML model does the same but with **mathematical logic and data**.

By End of Week 3:

- You will know **important ML algorithms** with real-life examples.
- You will understand **accuracy, confusion matrix, overfitting** in very simple terms.
- You will complete **IPL Winner Prediction Model**.