

Predicting Football Match Outcomes Using Logistic Regression

Team MSD-07
Ayush
Rahul
Sandhya

Model Selection

Model Chosen:

Logistic Regression

Reason:

- Ideal for predicting categorical outcomes (Win / Not Win)
- Simple and interpretable
- Works well with binary or multi-class classification problems
- Helps estimate the probability of a home win





Why Logistic Regression?

- Our target variable (FTR: Full-Time Result) has 3 outcomes (H, D, A)
- We transformed this into a binary classification:
- 1 = Home Win (H)
- 0 = Not Home Win (D or A)
- Logistic regression allows us to estimate probabilities and interpret how each feature affects the outcome
- Easy to extend to multi-class if needed later





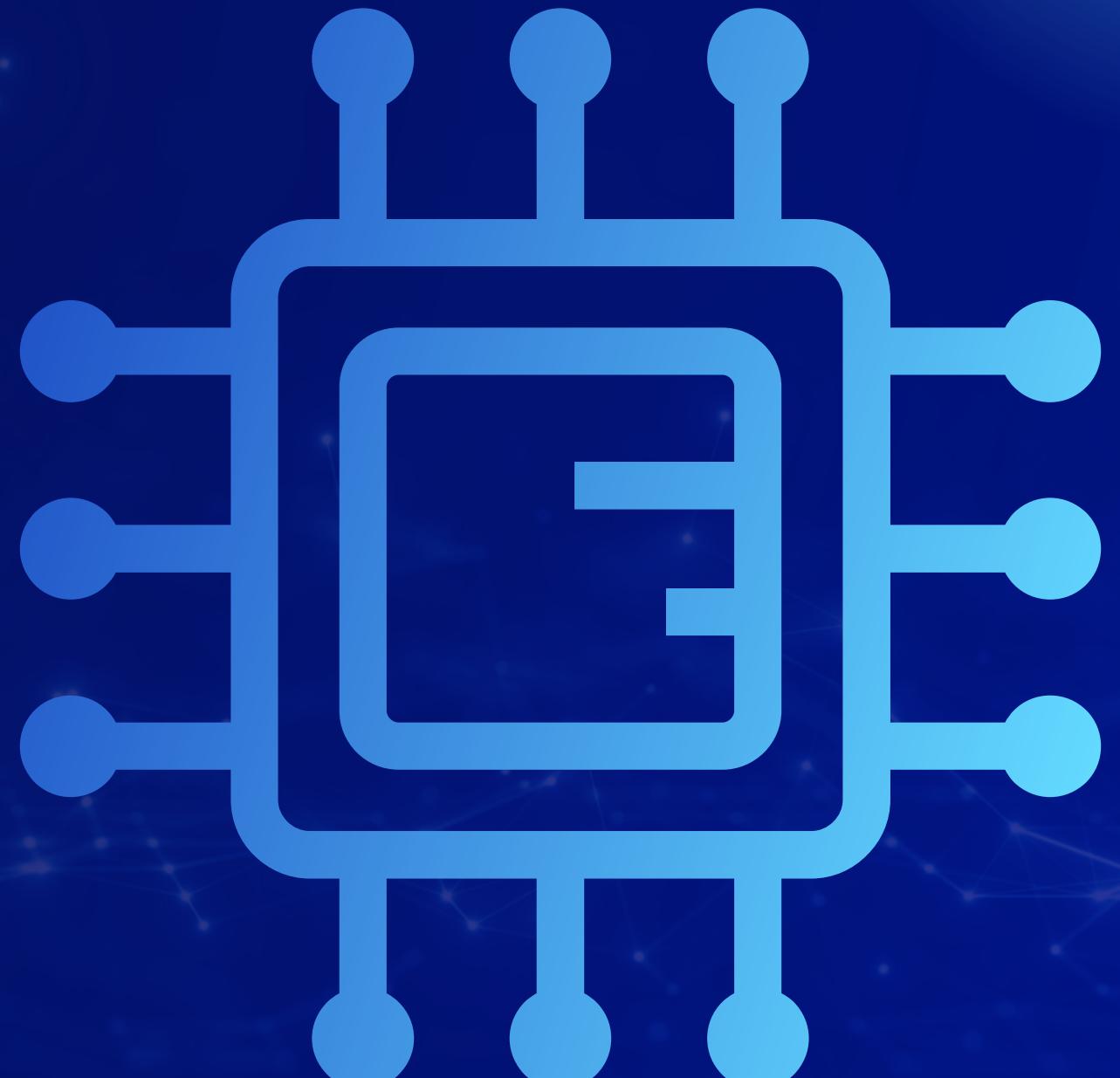
Feature Selection

Chosen Features:

1. HomeGoals – Number of goals scored by the home team
2. AwayGoals – Number of goals scored by the away team
3. GoalDifference – Calculated as HomeGoals – AwayGoals
4. (Optional) Wk – To observe performance patterns across the season

Why These Features?

- These variables showed strong patterns in our EDA
- Goal difference especially had a clear impact on match result
- They're numerical and directly relevant to team performance





Hypotheses

- 🔍 H1: Matches with a higher goal difference are more likely to result in a home win
- 🔍 H2: On average, home teams score more goals than away teams
- 🔍 H3: When the home team scores 2 or more goals, the chance of winning increases significantly



**THANK
YOU**

