

15.095: Machine Learning Under a Modern Optimization Lens

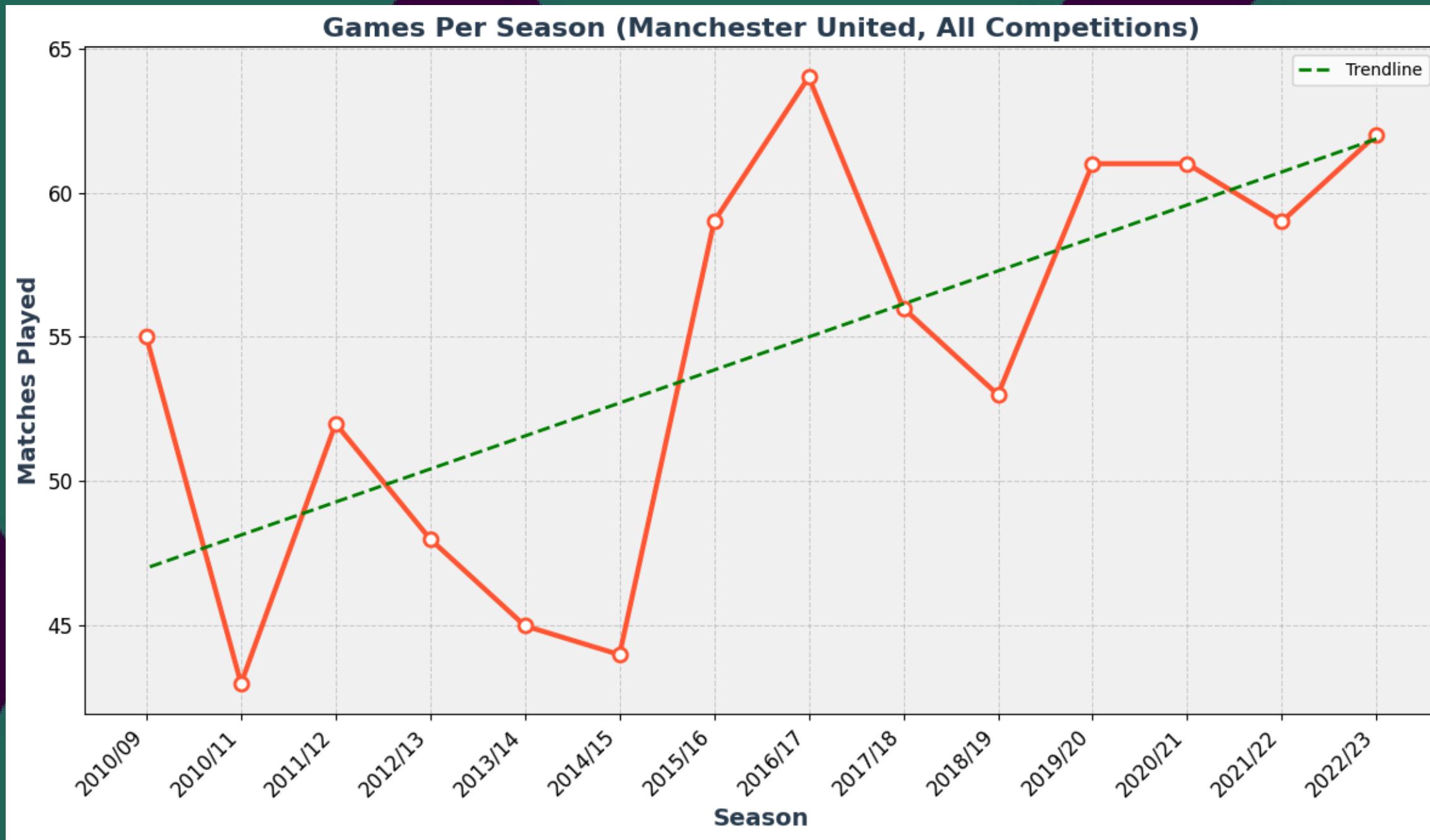
# Decision-Support System for Optimal Player Substitutions

In Professional Soccer Matches

Malcolm Gaynor, Atharva Navaratne

*MIT Sloan School of Management*

# Motivation - Why Substitutions Matter



Games Per Season has Increased Significantly Over the last ~10 Years\*

Strategic Impacts:

- Increased Transfers to Boost Squad Depth
- Increased Emphasis on Substitutions

Substitutions can be Key in Deciding Game Outcomes!

\*Does Not Include UEFA Competitions' New Format - Increases Number of Games Even More!

# Overview of the Data

Football Events Dataset ([Kaggle\\*](#))

Consists **940K+** “Events”.

Match Data: Goals, Attempts, Fouls,  
Corners, etc., Pre- and Post-  
Substitution

Substitution Data: FIFA Ratings ([Kaggle\\*](#))

for Players Subbed In and Out.

Fantasy Premier League Data ([EPL API\\*](#))

**Detailed Player Performance Metrics:**  
Minutes, G/A, Penalties, Clean Sheets,  
etc.

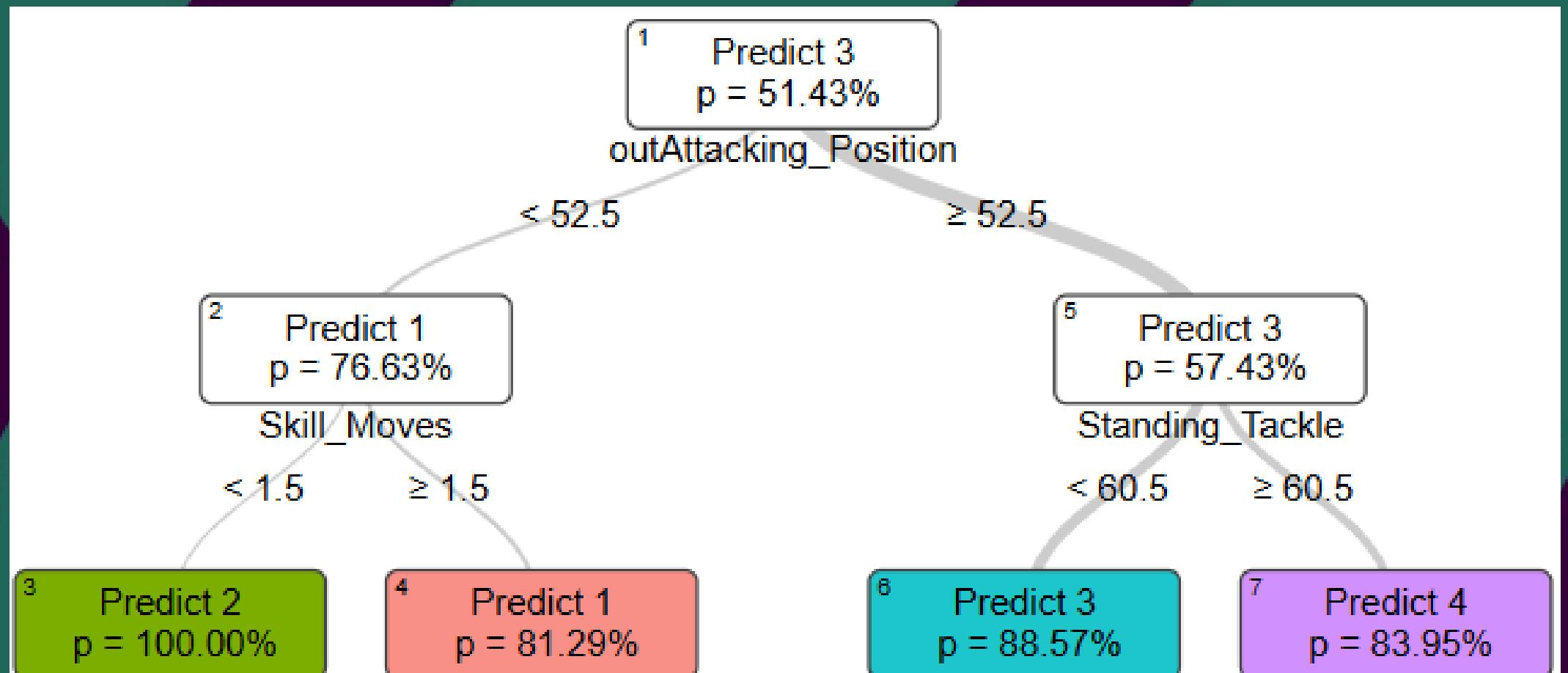
Dynamic Dataset: Updated After Every  
Game - at most Weekly.

# Clustering Substitutions

*Classifying the Types of Substitutions (Treatments)*

Classifying Substitutions into **4 Clusters** - OCT to Interpret Clusters

- 86% OOS Accuracy (Baseline Accuracy ~51%)
- Depth Restricted to 2 for Interpretability



**Substitution-Type Intuitively Classified** Into the Following:

1. Offensive In, Defensive Out
2. Defensive In, Defensive Out
3. Offensive In, Offensive Out
4. Defensive In, Offensive Out

# Optimal Prescriptive Trees

*When to “Prescribe” what “Treatment” (Cluster)*

OPT Fitted on **Pre-Substitution Match Details**: Considers the Game Dynamics  
Treatment: Substitution Cluster Used

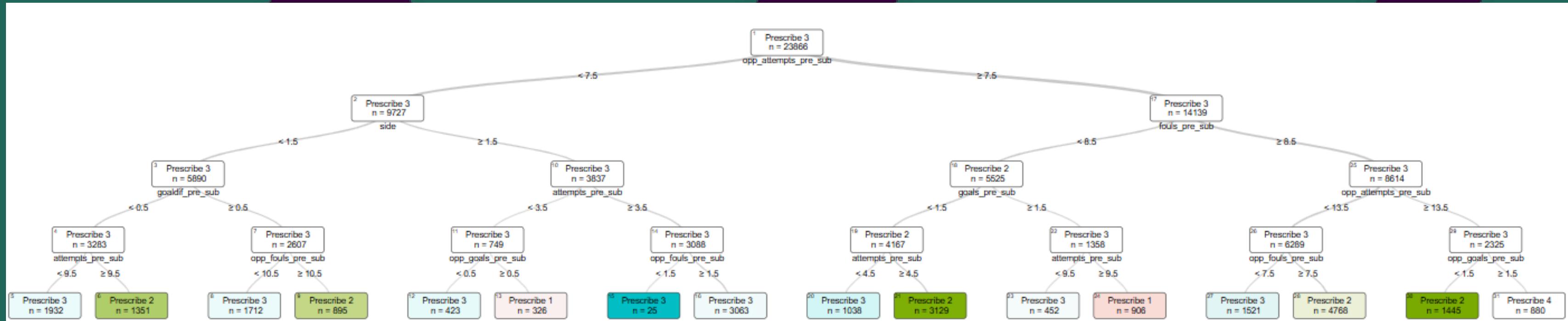
Outcome: **Post-Sub Goal Differential** (Goals Scored - Goals Scored by Opposition)

Regularization-Inspired Metric to Balance Goals, Attempts, and Game Scenario

- When **Losing**: Goal Differential +  $\lambda$ (Shots Attempted)
- When **Winning**: Goal Differential -  $\lambda$ (Shots Conceded)
- When **Drawing**: Goal Differential +  $2\lambda$ (Shots Conceded - Shots Attempted)

# Sub-Cluster Insights

*Are the Clusters Consistent with Logic?*



5% Cluster 1\* (Attacker In, Defender Out), 4% Cluster 4 (Defender In, Attacker Out)

49% Cluster 2 (Defender In, Defender Out), 43% Cluster 3 (Attacker In, Attacker Out)

Consistent with Logic: **Like-for-Like Substitutions are Significantly More Common!** The OPT prescribes Unbalanced Substitutes only in Very Specific Scenarios.

\*Offensive In, Defensive Out - Probably More Drastic than Cluster 4, hence occurs Almost Exclusively in Home Games!

# Drastic Prescriptions

*Digging Deeper into when the Prescriptions are Drastic... and Why?*

Cluster 4 (Deffensive In, Offensive Out):

- Tons of Opponent Shot Attempts
- Too Many Fouls
- Opponent has Scored Multiple Times

Defense has been Insufficient  
**Needs Reinforcement**, even at an  
Attacking Cost.

Cluster 1 (Offensive In, Defensive Out):

- Both Teams have Few Attempts
- Away Game (Crowd Hostility!)
- Opponent Has Already Scored

Defense is playing well, but you need to score and the Crowd is Hostile. **Throw Caution to the Wind** and Try to Get Goals!

# Cluster Individual Players

*What Players do we have on the Bench?*

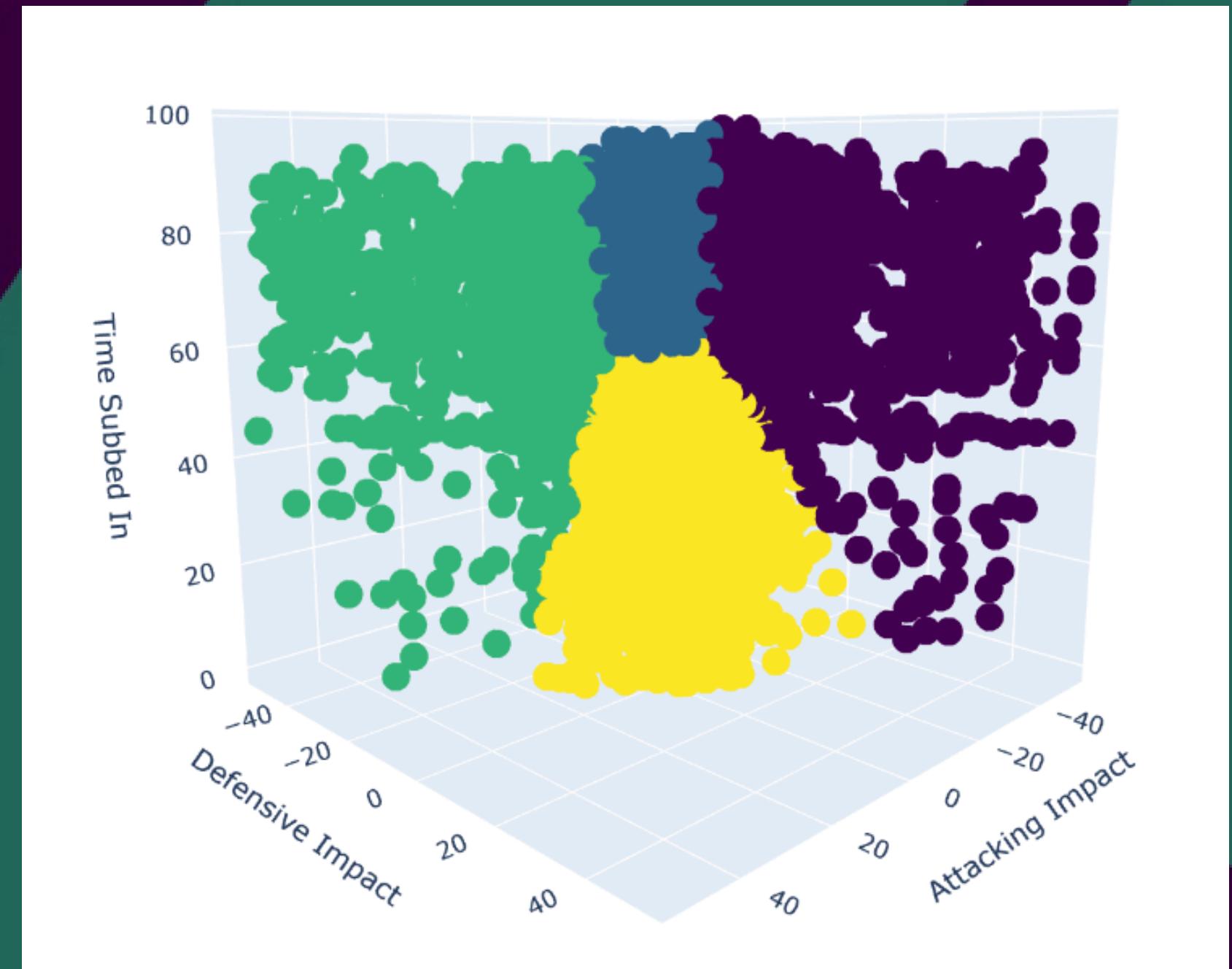
Define Attacking Impact and Defensive Impact scores as a weighted average of relevant FIFA ratings - used for Clustering.

**Interpretation** (Based on OCT):

- Early Subs - Balanced (Cluster 1)
- Late Subs - Attack-Minded (Cluster 2)
- Late Subs - Defense-Minded (Cluster 3)
- Late Subs - Balanced (Cluster 4)

From the OPT, we know What Type of Player to Sub In - Identify Individual Cluster.

Just around ~23% of the Substitutions happened Before the 45th Minute - 93% of These Were Due to Injury/Fatigue (No Strategic Reason). So we focus on Substitutions After the 45th Minute!



# From Cluster to Player

*Enough about Clusters... tell me who to Sub In!*

For Attackers/Midfielders:

- Probability of **Scoring a Goal** after being subbed in; higher is better.
- XGBoost (OSR<sup>2</sup>. ~91%)

For Defenders:

- Probability of **Conceding a Goal** after being subbed in; lower is better.
- XGBoost (OSR<sup>2</sup>. ~79%)

Regularization:

Customized Penalty (Age and MP) for “Tough” Games.\* If ODI > 3, Objective is:

- $(\text{Goals Scored} - \text{xG}) - \lambda(\text{Age} + \text{Minutes Played})$

Baseline Model: Goals Scored around this time/Goals Conceded Against the same teams **Last Season**. Does not Account for Form, ICT, FDR, Set Pieces, or Fixture Congestion.

Sort Arrays according to a Logical Order and **Prescribe the First Player** in the Array!

\*Historically, against tough opposition, managers don't pick players who haven't played a lot of games so far this season, or players who are too young and therefore inexperienced.

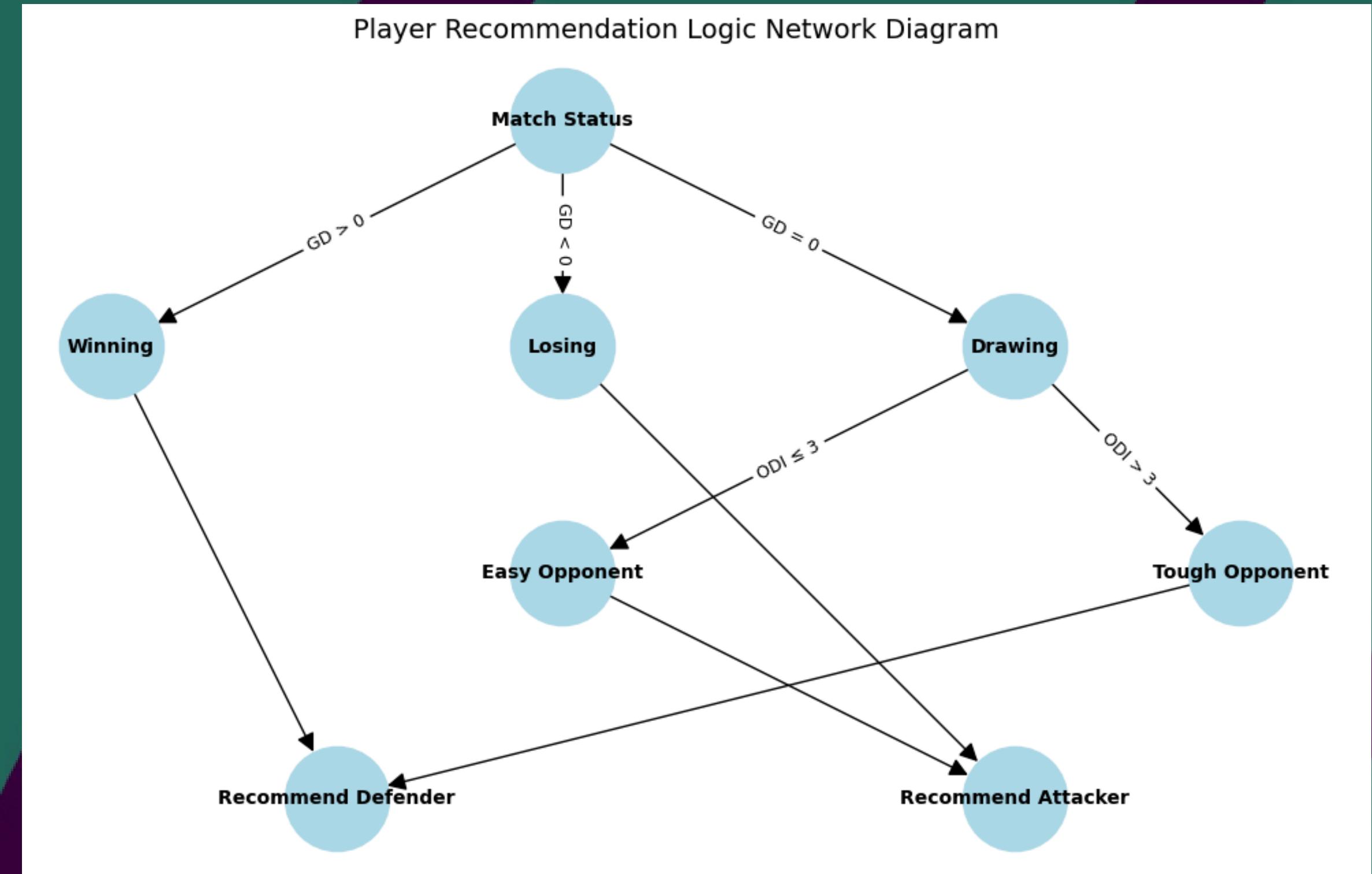
# In-Game Decision-Support

**Step 1: Identify and Prescribe Candidate Substitutions based on OPT.**

**Step 2: Determine Candidates on the Bench based on Individual Clustering.**

**Step 3: Identify Best Player(s) Within-Cluster.**

**Step 4: Sanity Check - the Coach's Call!**



# Manchester United v/s Fulham

Step 1: OPT Prescribes **Class 3 - Offensive In, Offensive Out.**

Step 2: Four Candidates (**Cluster 2**) on the Bench!

Step 3: Substitute In - Alejandro **Garnacho** and Joshua **Zirkzee\***. (Highest xG)

Step 4: Sanity Check: Drawing Against a Weak Team - Sub in Attacker.

## Double change for the Reds

Nearly everyone inside Old Trafford is on their feet as Erik ten Hag goes to the bench. Amad and Mason Mount make way, as Joshua Zirkzee and Alejandro Garnacho are introduced.

**Zirkzee the hero as Man Utd beat Fulham in 2024/25 opener**

16 Aug 2024



\*Joshua Zirkzee Scored an 87th-Minute Winning Goal!

15.095: Machine Learning Under a Modern Optimization Lens

# Thank you! Questions?

Malcolm Gaynor, Atharva Navaratne  
*MIT Sloan School of Management*

