

# AN6003

# Analytics

# Strategy

Group Members

DELWIN RAMSES VAN STRIEN

JO BUM JOON

NGUYEN NGOC THIEN HUONG

ZHANG JIANYU

ZHANG NING

ZHAO HENGJIE

01 - Introduction

02 - Descriptive Statistics

03 - Predictive Models

04 - Conclusions



# 01 - Introduction

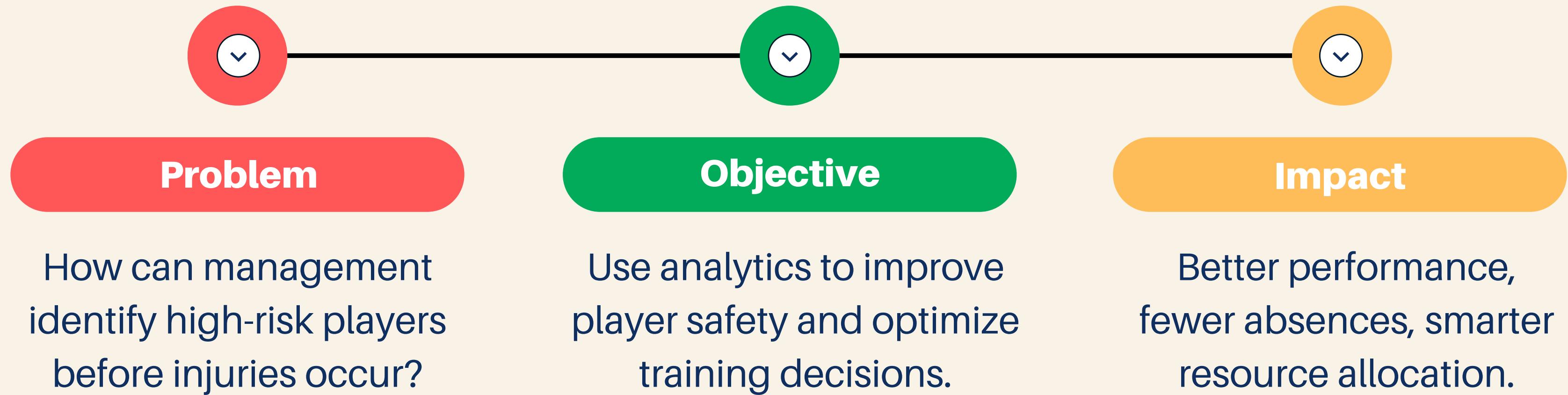


# Injury Prevention in University Football

- Football injuries remain a key challenge at the university level.
- They affect player health, team performance, and resource planning.
- Data analytics enables **proactive** prevention instead of **reactive** response.



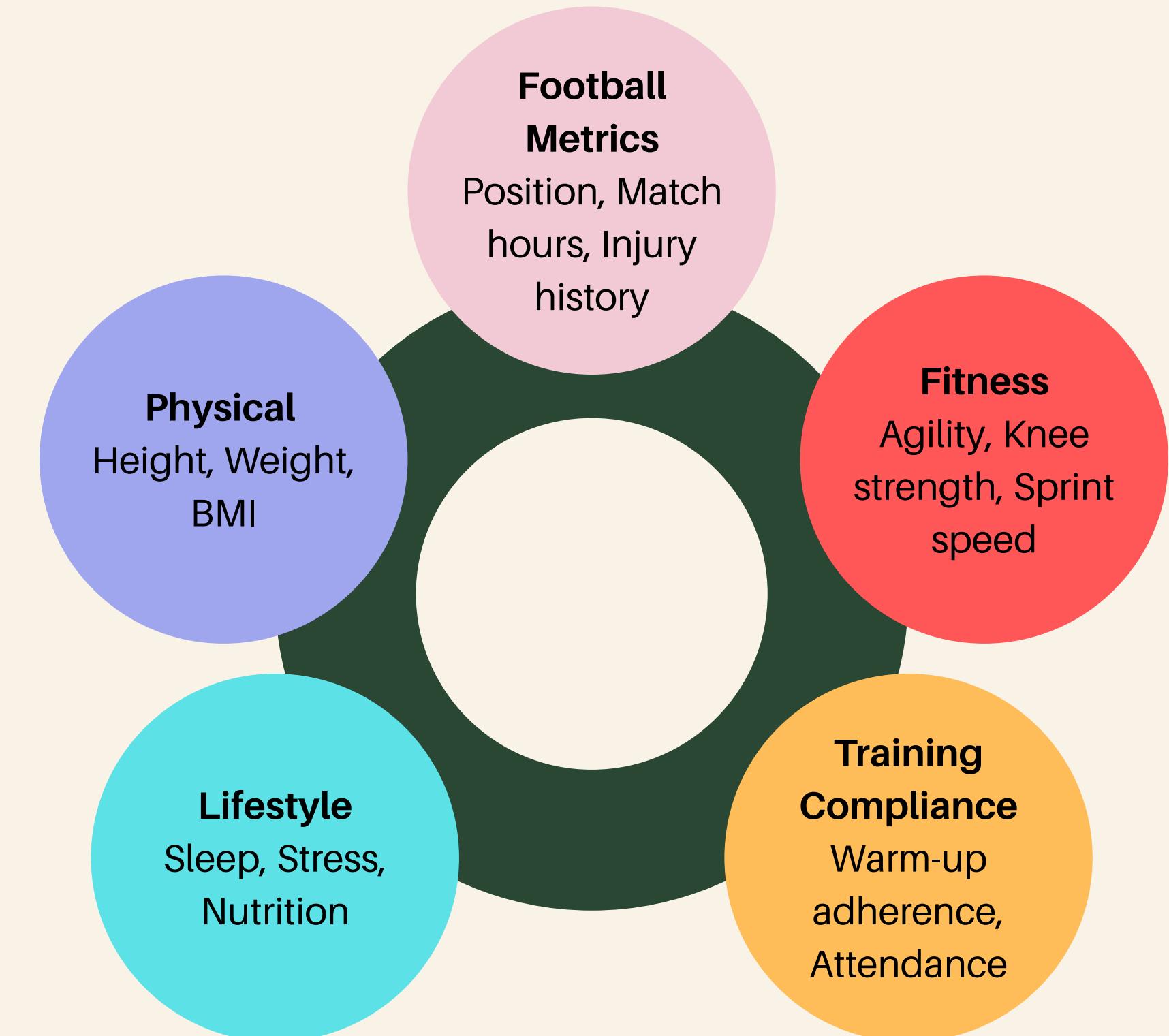
# From Challenge to Goal



# University Football Injury Dataset

## Dataset Summary

- 800 players aged 18–24 from collegiate & provincial leagues
- 18 input features grouped into 5 categories
- Target: Injury Next Season ( $\geq 7$  days absence)



# Analytical Framework

## Two-Stage Analytical Approach

### Stage 1 – Overall Actions

- Logistic Regression & Random Forest
- Predict injury likelihood and identify key factors

### Stage 2 – Detailed Threshold

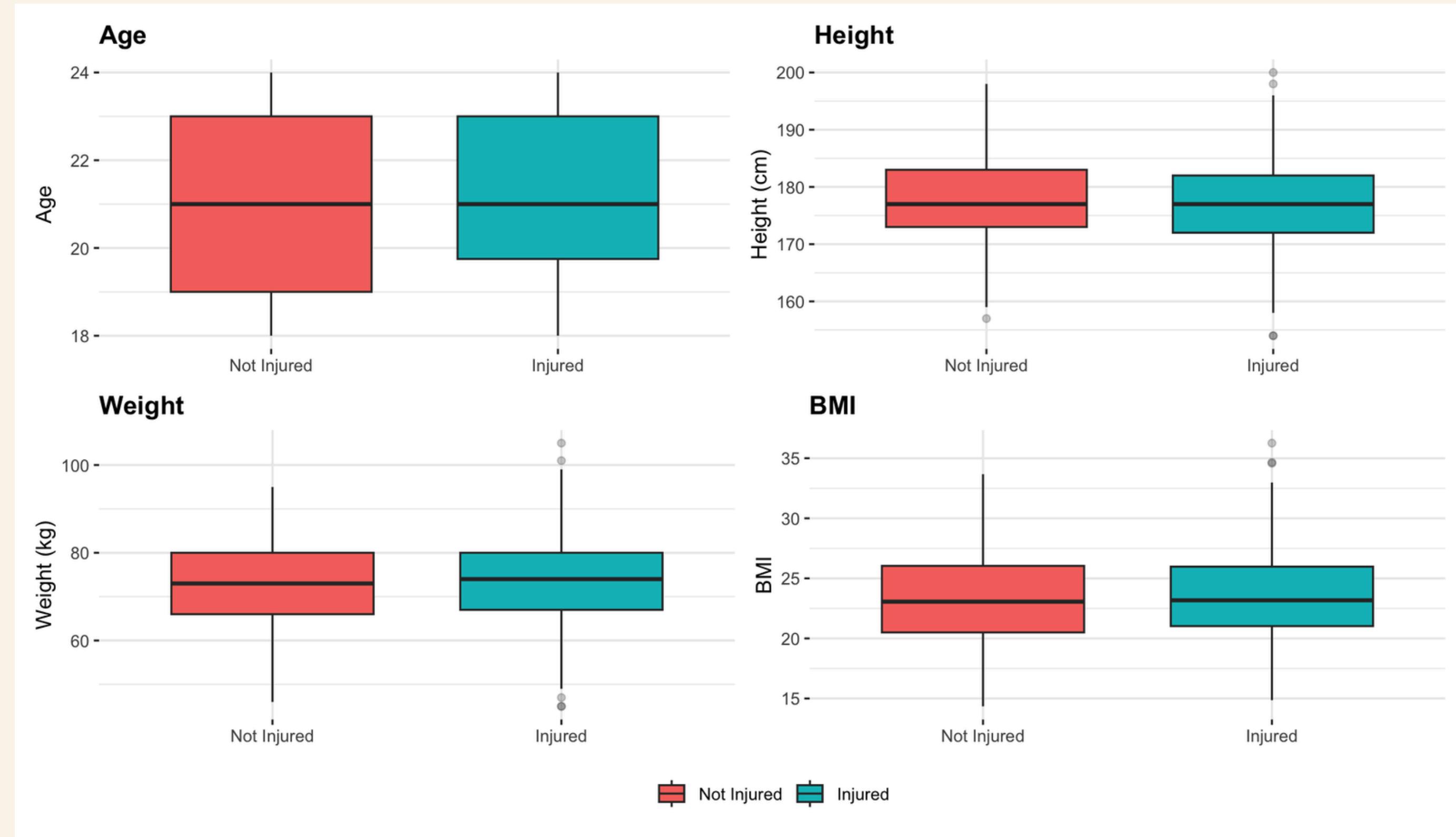
- Clustering & CART
- Segment players and determine actionable risk rules



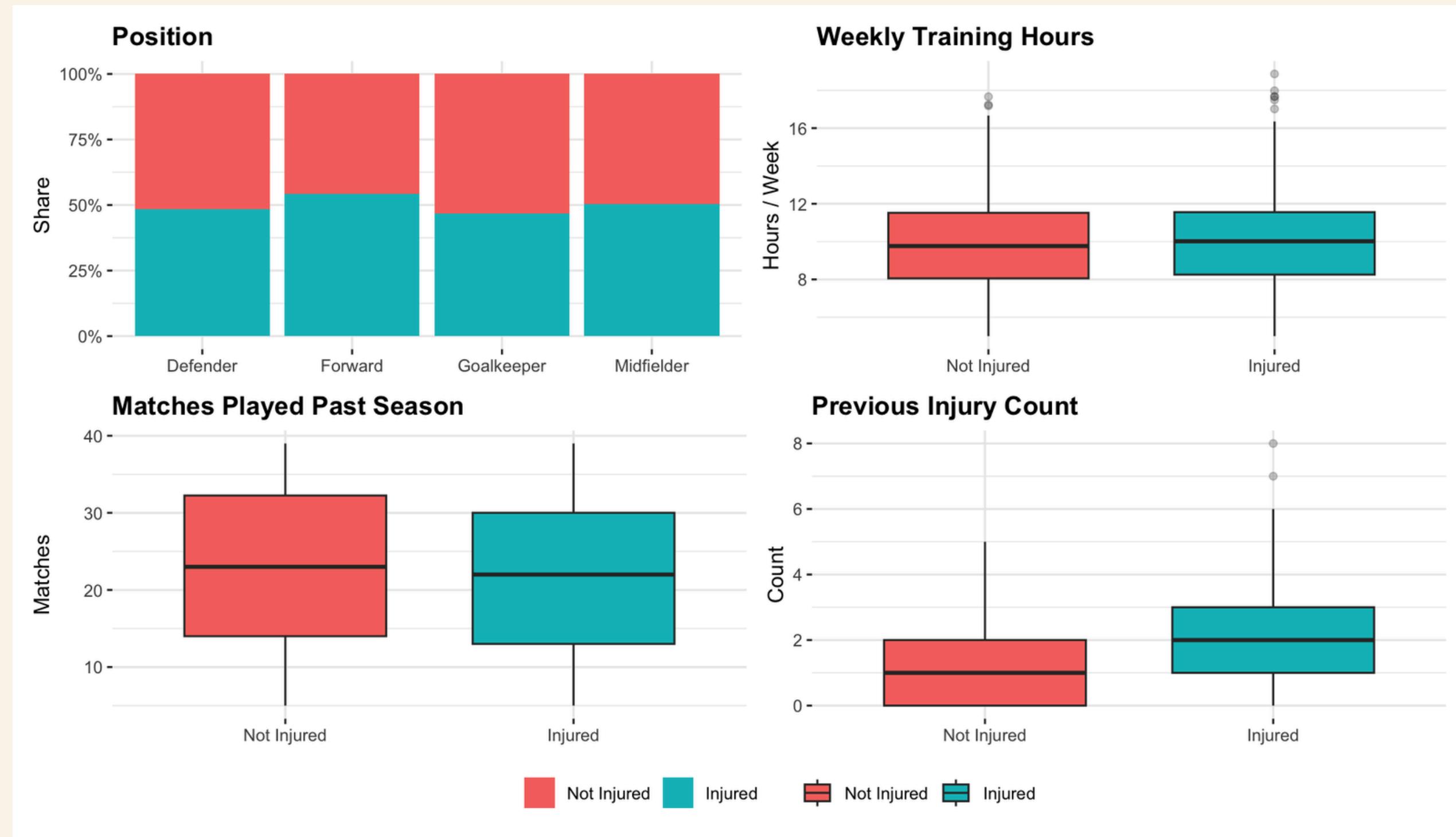
## 02 - Descriptive Statistics



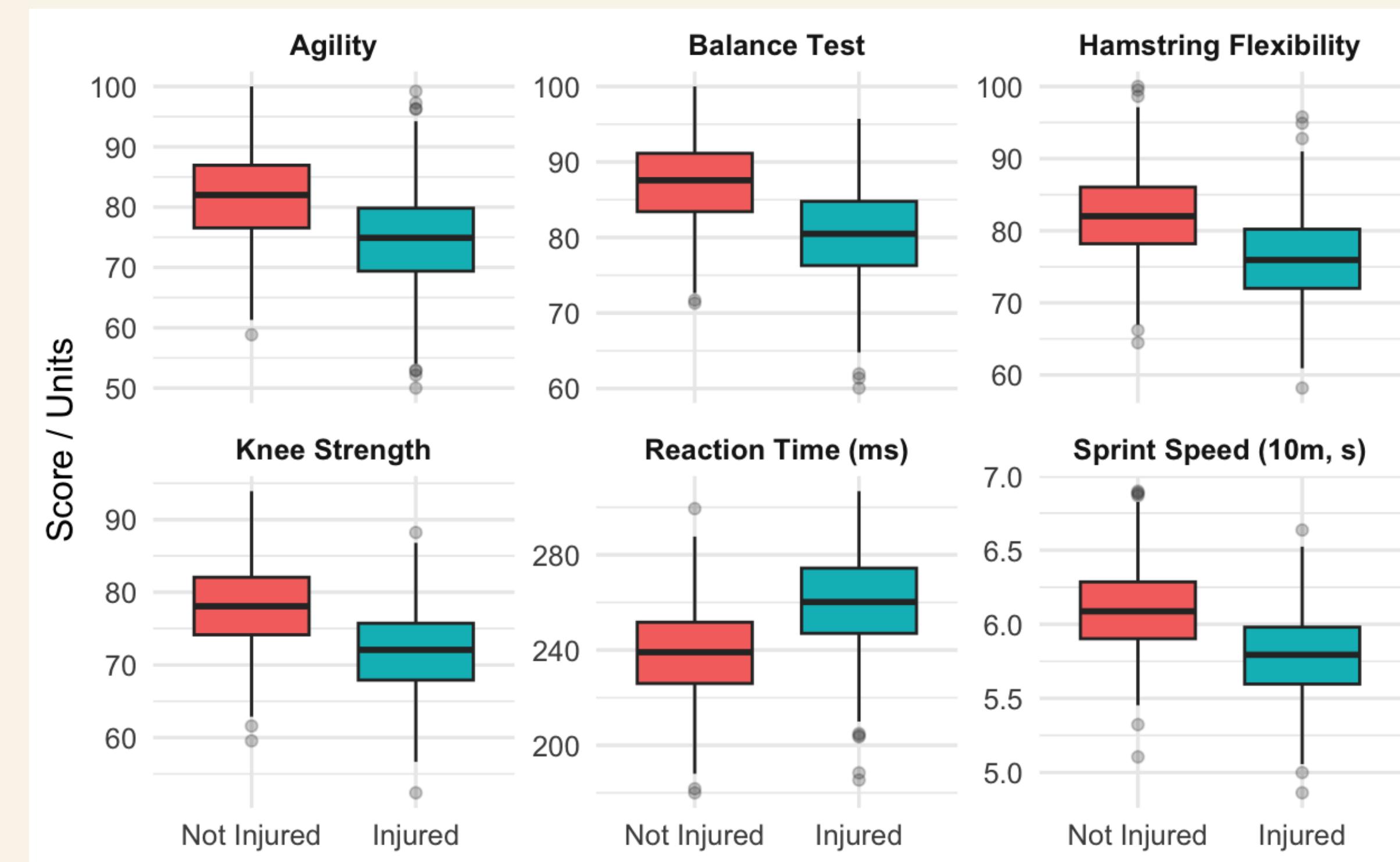
# No significant differences in physical attributes



# Except for previous injuries, unclear differences observed other football-specific attributes



# Not injured players score better across physical assessment factors



# Not injured players also have more sleep, higher nutrition quality and lower stress level



## 03 - Predictive Models

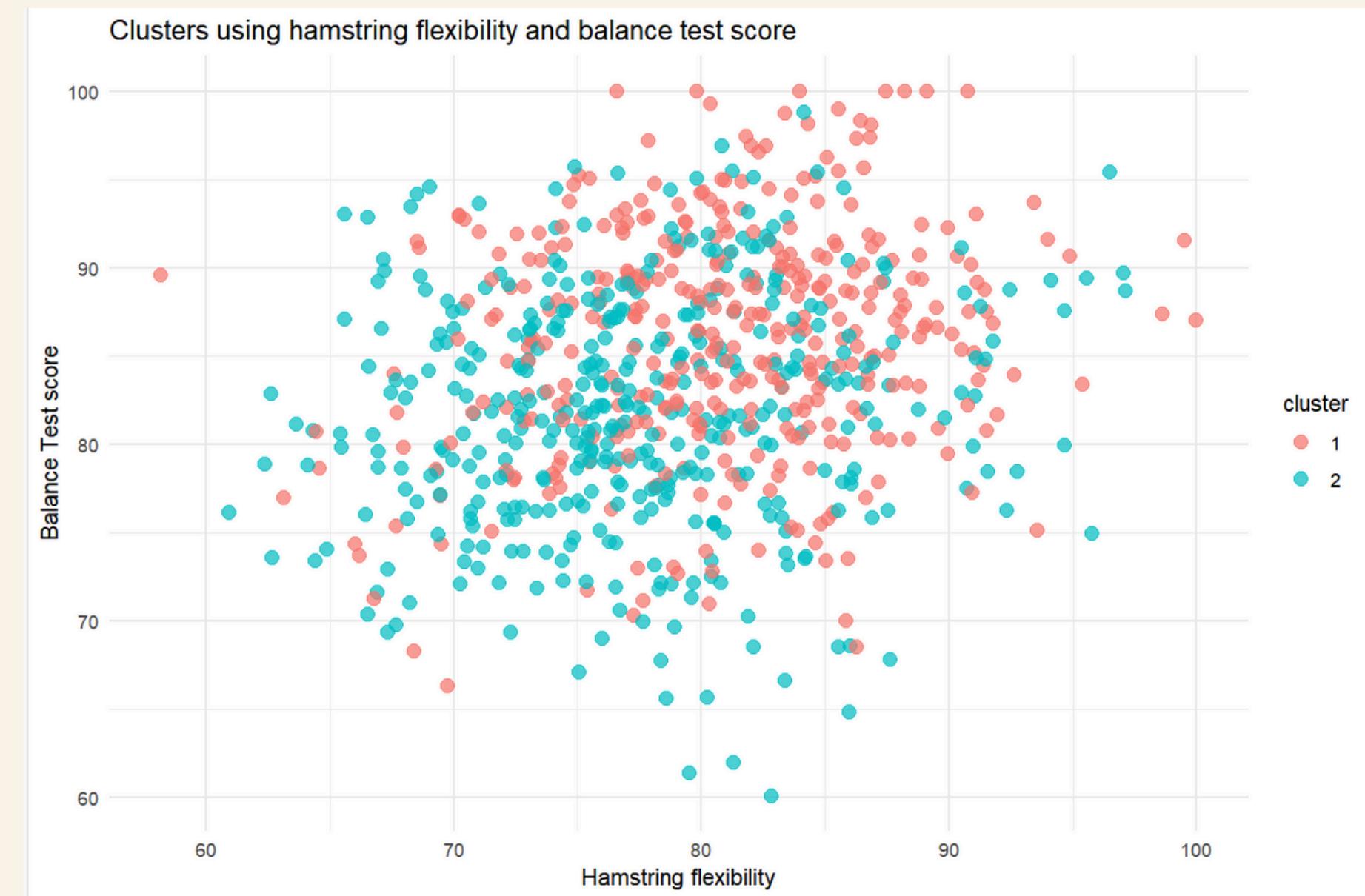
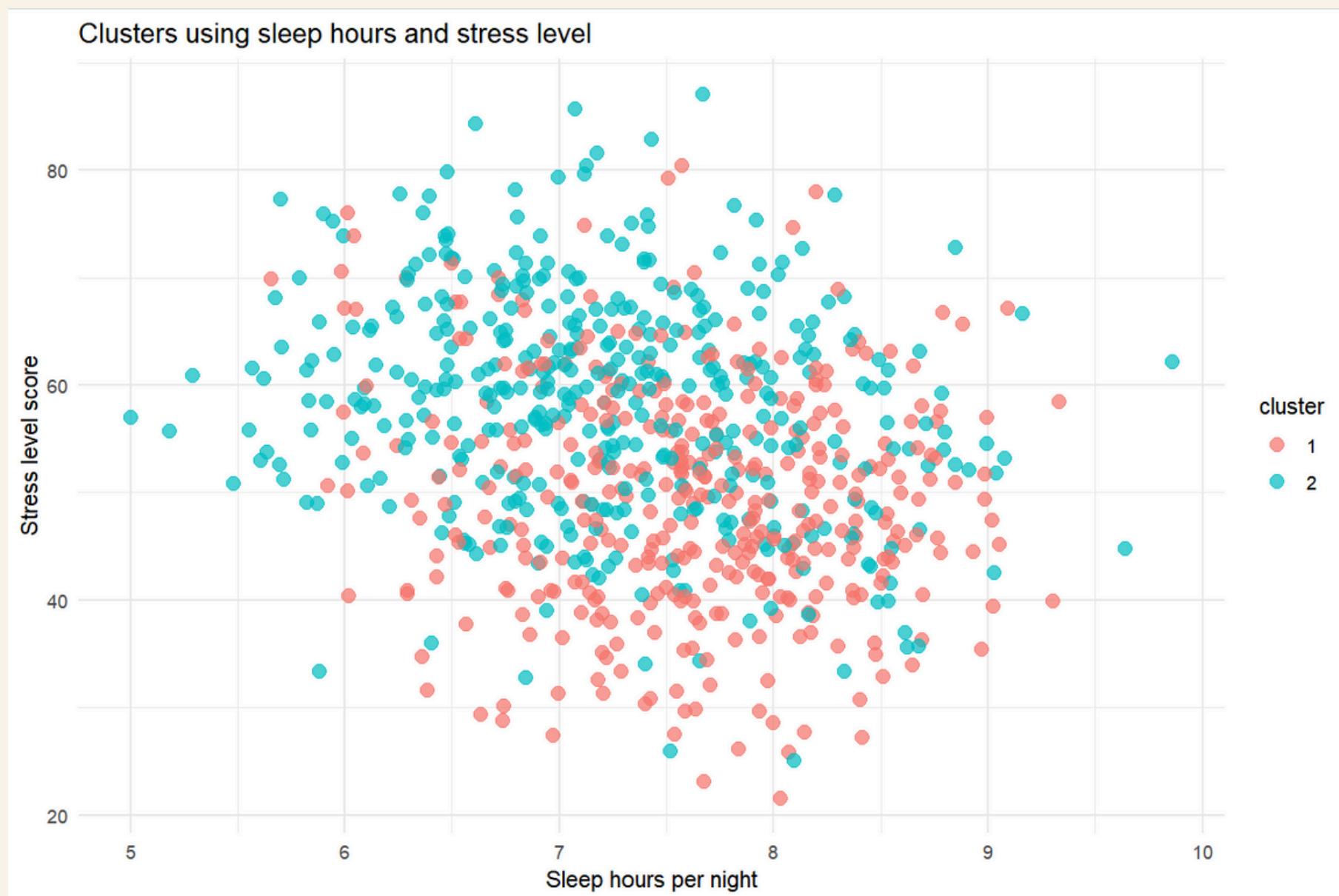


# Clustering

## Injury Rate

Cluster 1: 21.4%

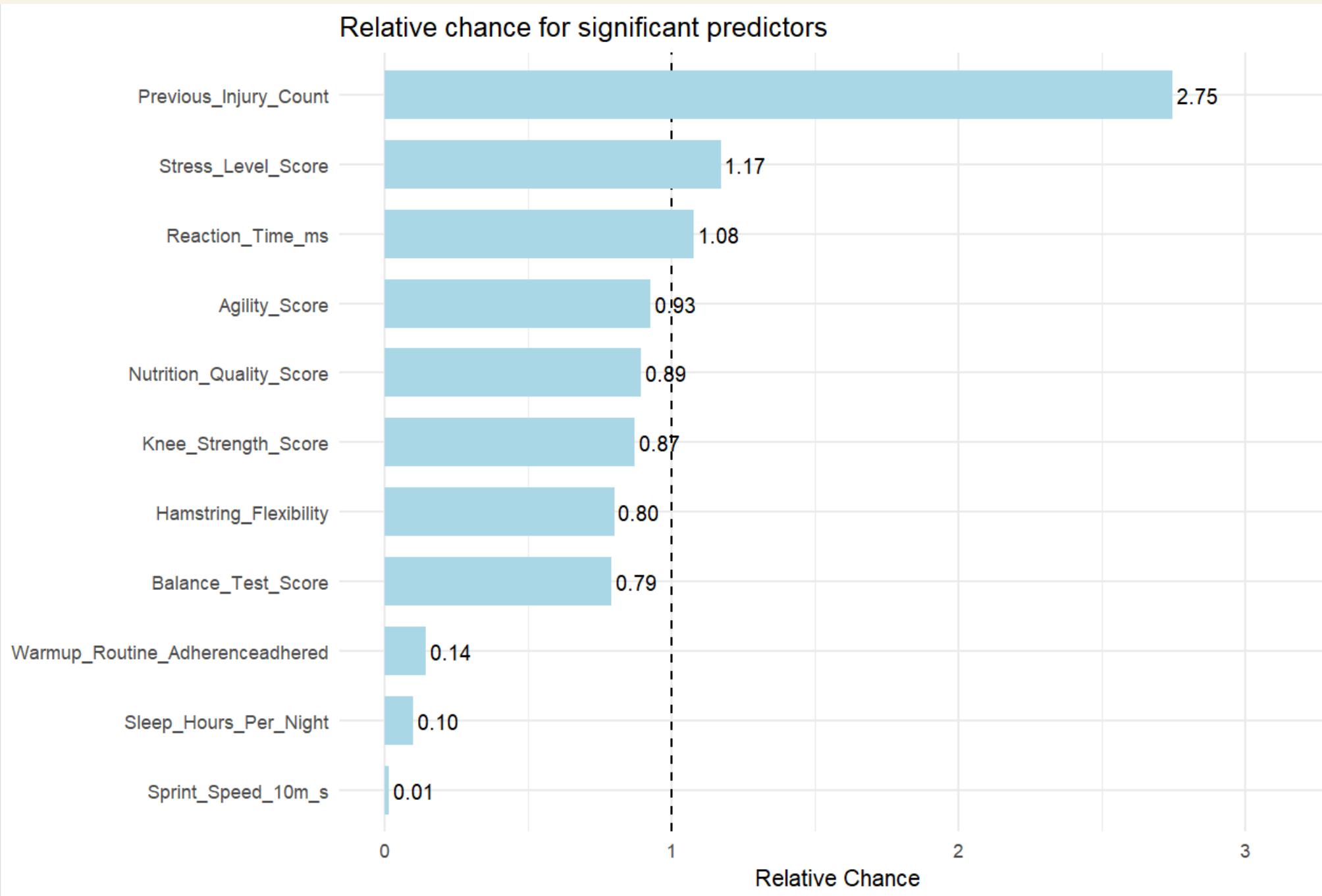
Cluster 2: 76.8%



**Noticeable gaps** between injury-prone and injury-resistant groups → baseline for further guidelines

# Logistic regression

## Important findings



### Football-Specific Metrics:

One more previous injury, twice more likely to get injured.

### Fitness Assessments:

1 unit increase in balance test score, 21% less likely to get injured.

### Lifestyle Factors:

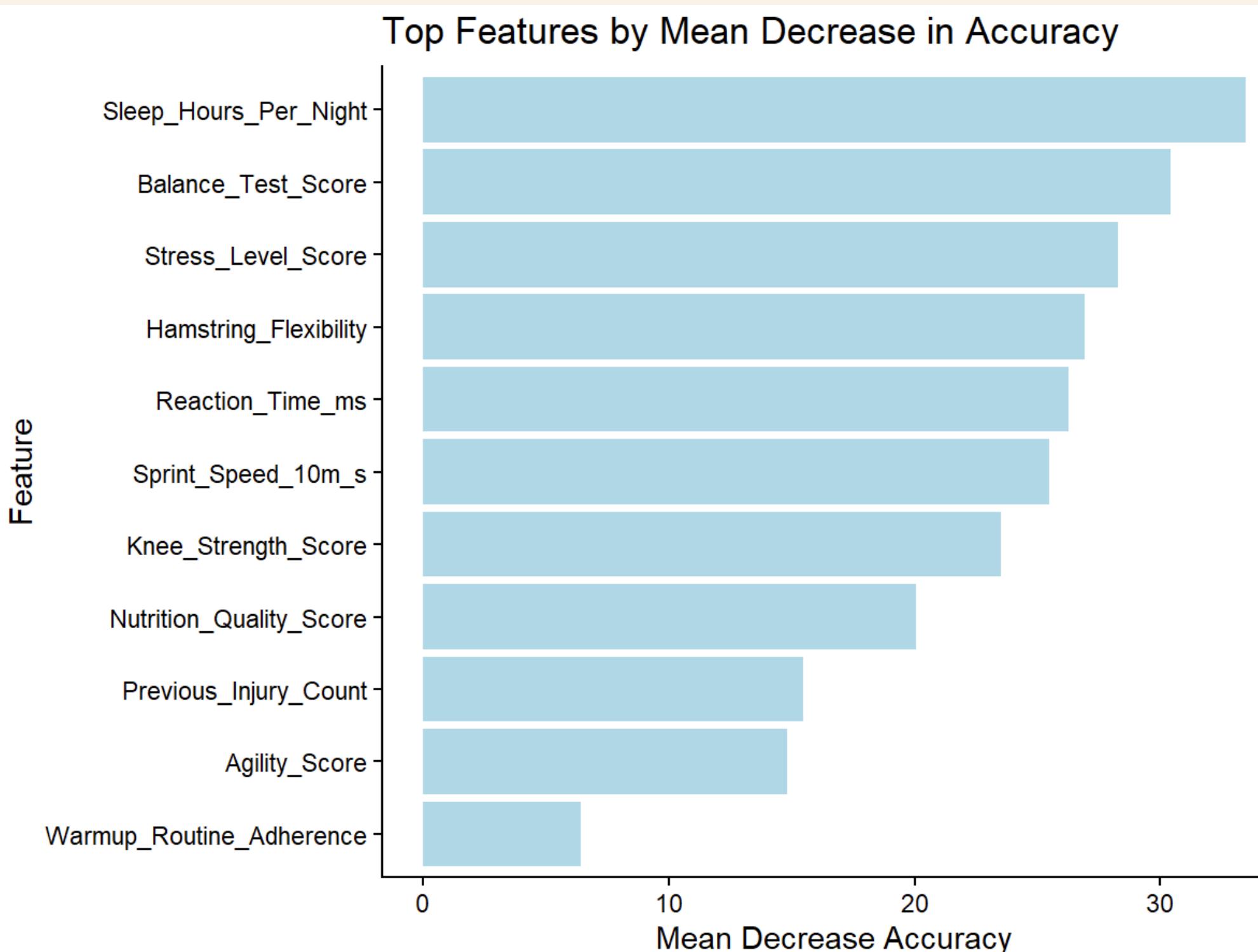
1 more hours of sleep, 90% less likely to get injured.

### Training Compliance:

If the player adheres to the warm-up routine, he/ she is 86% less likely to get injured.

# Random Forest Model

## *Top Influencing Factors*



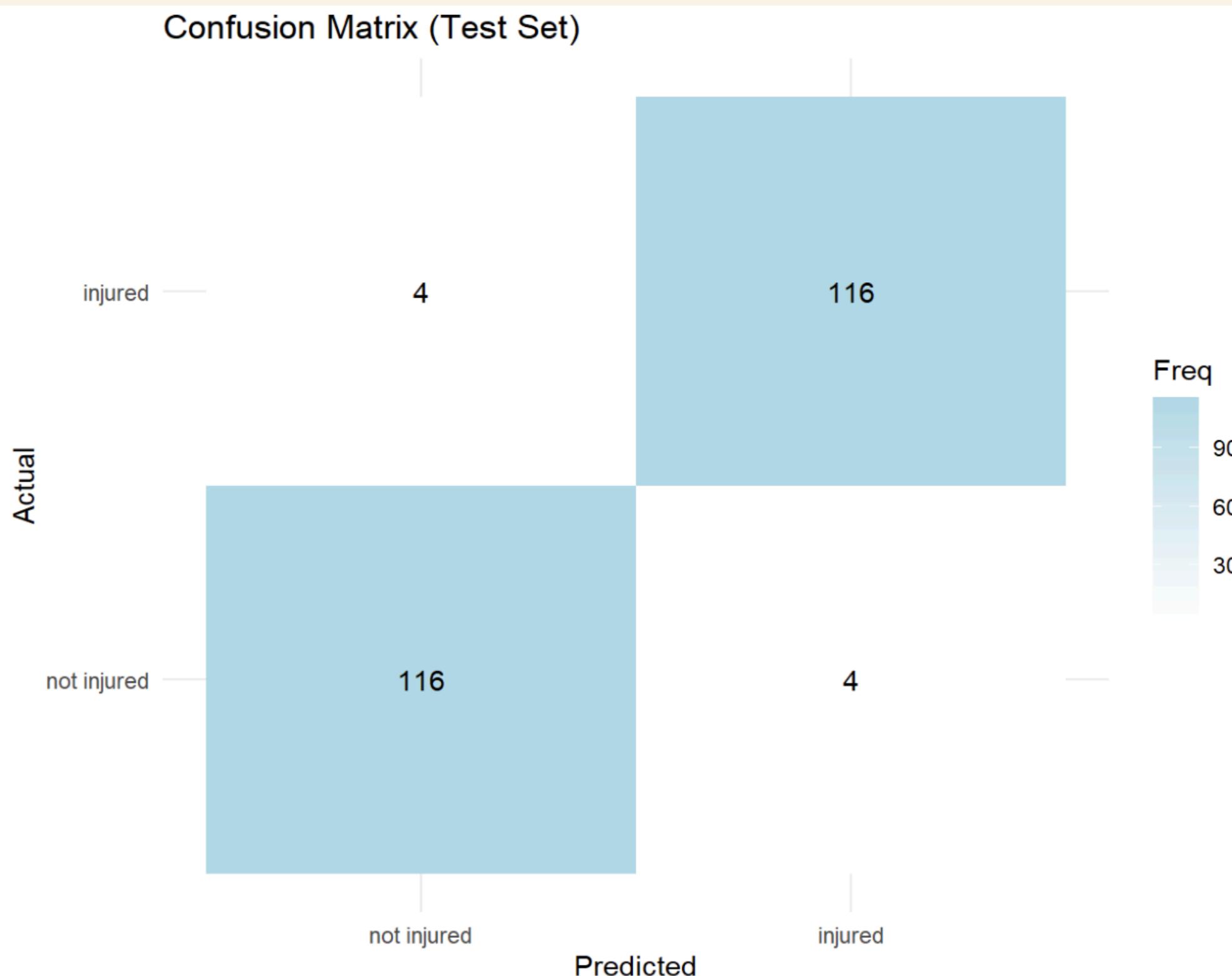
**Lifestyle factors:**  
Sleep\_Hours\_Per\_Night,  
Stress\_Level\_Score,  
Nutrition\_Quality\_Score

**Physical Fitness Assessment factors:**  
Balance\_Test\_Score,  
Hamstring\_Flexibility,  
Reaction\_Time\_ms

**Football-specific factor:**  
Previous\_Injury\_Count

**Training Compliance factor:**  
Warmup\_Routine\_Adherence

# Random Forest Model



Accuracy: 96.67%

**Top Right:** successfully predicted as “unlikely to be injured”



**Low-Risk Group**

**Bottom Left:** successfully predicted as “likely to be injured”



**High-Risk Group**  
Need more attention!!!

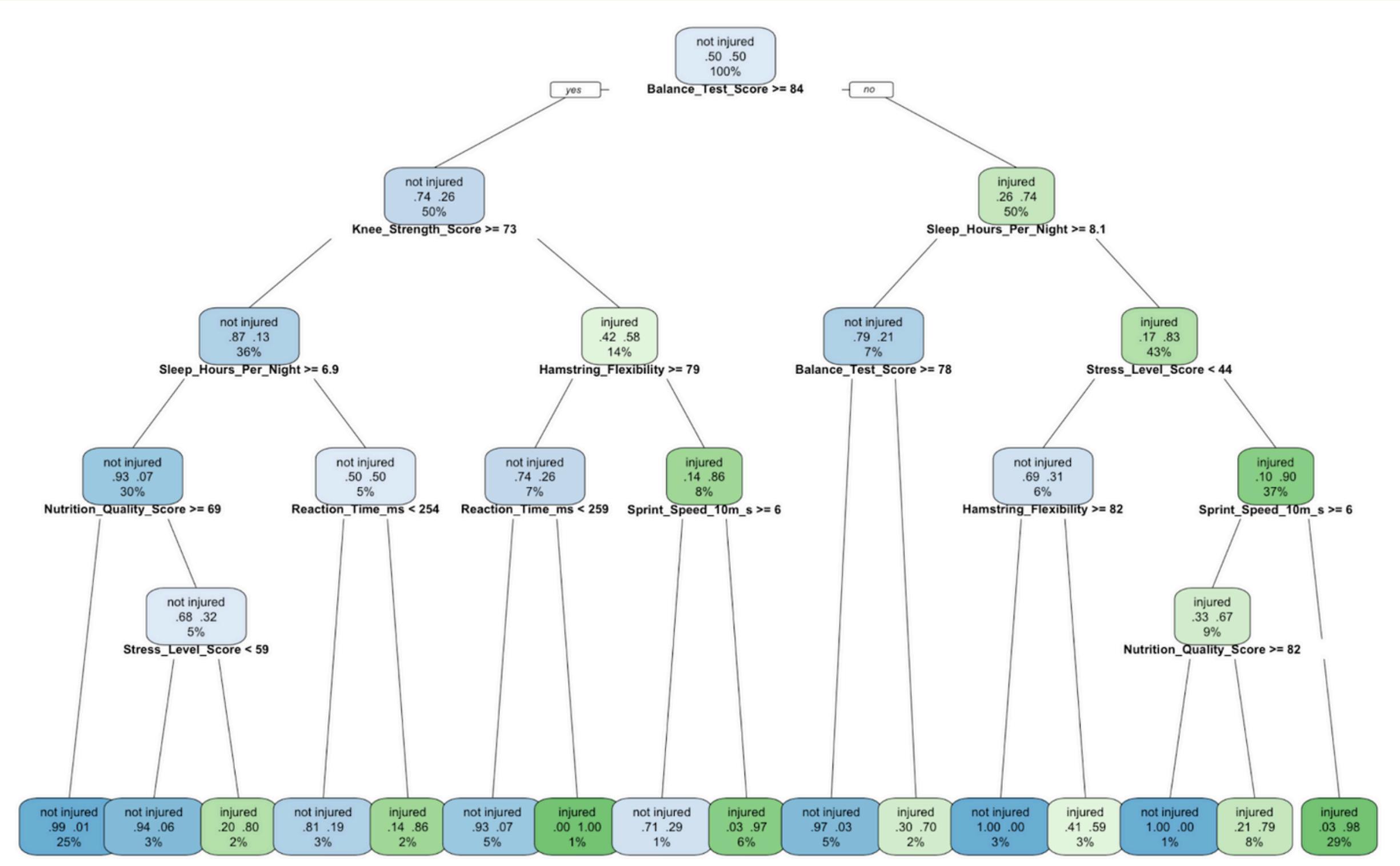
# Classification and Regression Tree

Pruned Decision Tree

Predicts whether a player will be injured or not next season.

Each split shows a threshold where injury risk changes noticeably.

Helps coaches and staff see which factor combinations are linked to higher or lower injury risk.



## 04 - Conclusions



# Action Plan

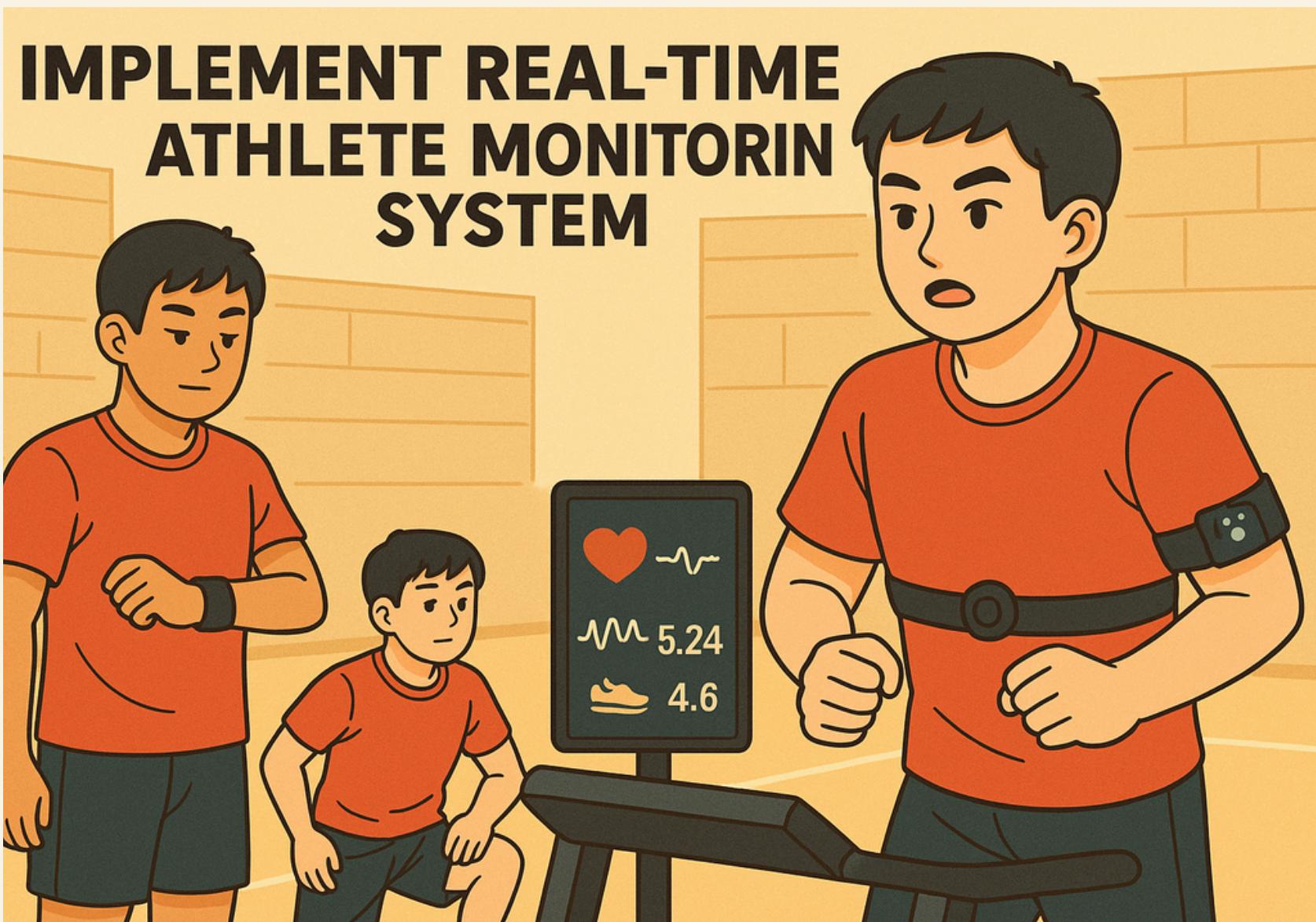
Factor Group	Factor	Target Threshold	Recommended Action
Physical Factors	Balance Test Score	$\geq 84$	Perform balance and stability exercises to reach or maintain a high balance score.
	Hamstring Flexibility	$\geq 82$	Perform regular hamstring stretching and mobility exercises.
	Reaction time ms	$< 254$	Engage in reaction and agility drills to improve response times.
	Sprint Speed m/s	$\geq 6$	Include sprint and speed drills to maintain optimal quickness and reduce injury risk.
	Knee Strength Score	$\geq 73$	Incorporate strength training focused on knees and lower limbs.
Lifestyle Factors	Sleep Hours Per Night	$\geq 8.1$	Aim for at least 8 hours of sleep each night to support recovery.
	Stress Level Score	$< 44$	Practice stress management techniques such as mindfulness or relaxation exercises.
	Nutrition Quality Score	$\geq 82$	Follow a balanced diet and consult nutrition professionals to optimize performance.

# Conclusions

## Predictive Analytics in University Sports Management

- Demonstrates **how data-driven models enhance** injury prevention and athlete well-being
- Combines Logistic Regression, Random Forest, Clustering, and CART **to achieve high predictive accuracy**
- Bridges statistical modeling with practical, actionable **insights for coaching and medical teams**

*Implement Real Time Athlete Monitoring System*



# Recommendations

*Early Warning Detection,  
Proactive Recovery Plan,  
Adaptive Training*

## Practice & Lifestyle Guidelines

Establish a compilation of practice and lifestyle guidelines to avoid injuries

## Early Preventive Actions

Identify injury-prone athletes for early adjustments and prevention

## Athlete Monitoring Systems

Implement a real-time monitoring system to help manage injury risk proactively



Thanks

# Appendix

# Logistic Regression Model

```
Call:  
glm(formula = Injury_Next_Season ~ ., family = binomial, data = trainset)  
  
Coefficients:  
  
Estimate Std. Error z value Pr(>|z|)  
(Intercept) 96.3002373 49.0555078 1.963 0.049636 *  
Age 0.0008502 0.1349633 0.006 0.994974  
Height_cm -0.1016948 0.2712490 -0.375 0.707725  
Weight_kg -0.0475005 0.3362939 -0.141 0.887675  
PositionForward 0.2961067 0.7495167 0.395 0.692796  
PositionGoalkeeper -0.5683980 0.8036702 -0.707 0.479409  
PositionMidfielder -0.3808158 0.7678977 -0.496 0.619951  
Training_Hours_Per_Week 0.1745738 0.1058684 1.649 0.099154 .  
Matches_Played_Past_Season -0.0325012 0.0274731 -1.183 0.236801  
Previous_Injury_Count 1.0107700 0.2656066 3.806 0.000142 ***  
Knee_Strength_Score -0.1402995 0.0437340 -3.208 0.001337 **  
Hamstring_Flexibility -0.2246293 0.0502965 -4.466 7.97e-06 ***  
Reaction_Time_ms 0.0737060 0.0146572 5.029 4.94e-07 ***  
Balance_Test_Score -0.2377597 0.0529052 -4.494 6.99e-06 ***  
Sprint_Speed_10m_s -4.2236190 0.9847299 -4.289 1.79e-05 ***  
Agility_Score -0.0764375 0.0332713 -2.297 0.021596 *  
Sleep_Hours_Per_Night -2.3349888 0.5016963 -4.654 3.25e-06 ***  
Stress_Level_Score 0.1571514 0.0321660 4.886 1.03e-06 ***  
Nutrition_Quality_Score -0.1148693 0.0340838 -3.370 0.000751 ***  
Warmup_Routine_Adherenceadhered -1.9525519 0.5944029 -3.285 0.001020 **  
BMI 0.1052554 1.0360149 0.102 0.919077  
---  
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## Significant variables

**Football-specific metrics:**  
previous injury count

**fitness assessments:**  
knee strength, hamstring flexibility, reaction time, balance score, sprint speed, agility score

**lifestyle factors:**  
sleep hours, stress level, nutrition quality

**Training compliance:**  
warm-up routine adherence

# Classification and Regression Tree

*CART Model Predictions vs Actual Injury Rates*

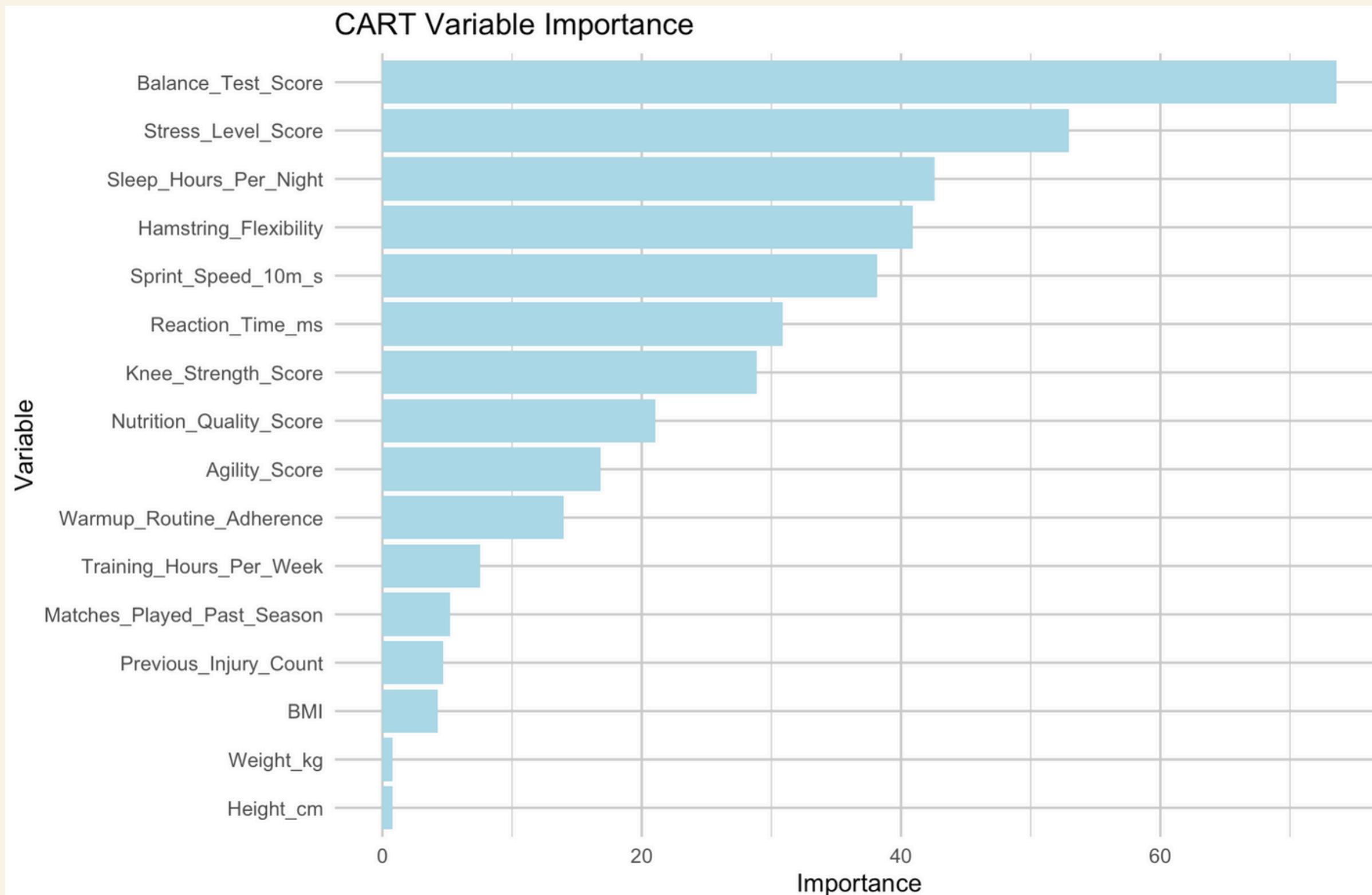
Confusion Matrix (Test Set)



**Accuracy: 82.1%**

# CART Most Important Variables

## Top Influencing Factors



### Physical Assessment Factors

Balance Test Score  
Hamstring Flexibility  
Sprint Speed m/s

### Lifestyle Factors

Stress Level Score  
Sleep Hours Per Night  
Nutrition Quality Score

### Training Compliance Factor

Warmup Routine Adherence