3MCS Specialization Project Presentation

An Interactive Dashboard for Football Data Analysis

Akash Dasmondal (1847104)
Michael Susheel G (1847113)
Diniz Osten Anthony (1847146)

Under the guidance of Dr. Nizar Banu P K

A Brief Description of Modules

- The dashboard focuses on providing an analytical framework for decision-making processes in the realm of professional football
- Focuses on 3 modules:
- Module 1: An analysis of the player attributes and predicting optimal position
- Module 2: An analysis of BMI and its relation to player attributes and position
- Module 3: An analysis of player progression using flexible attributes
- All modules make use of data extracted from EA Sports FIFA player rating database
- The dashboard, built using the Flask framework, will be the primary method of displaying the reports and outputs

Module 1: Predicting Optimal Position

- Analysis of player performance patterns
- Determine relationships between player attributes
- Predictive analysis of player position based on statistics
- Predict player role using Decision Tree Classification
- Predict player position using Random Forest Classifier

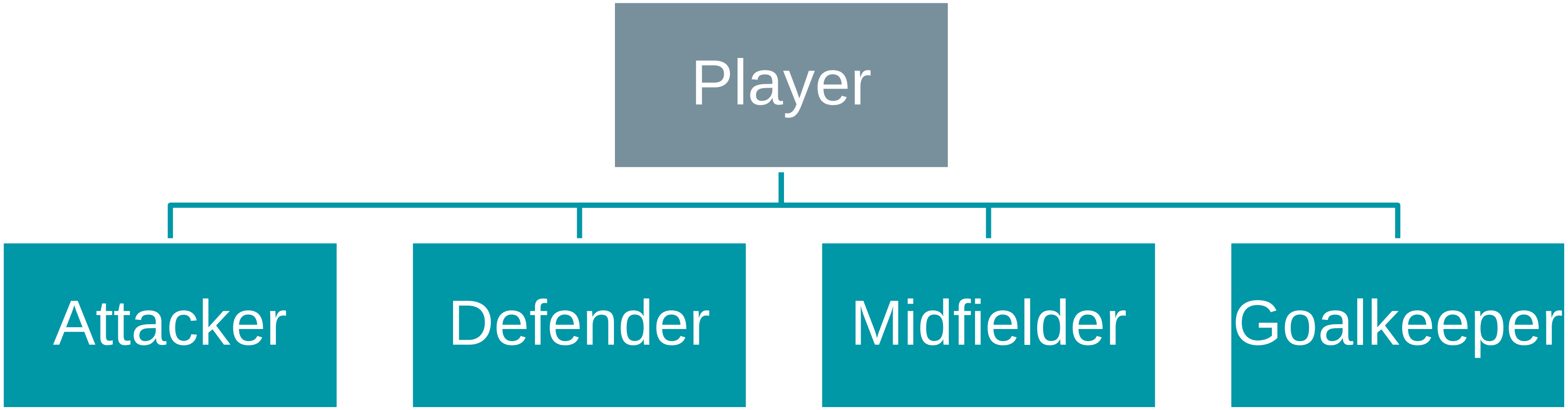


Fig 1.General Model for Roles

Fig 2.Primary Football Positions

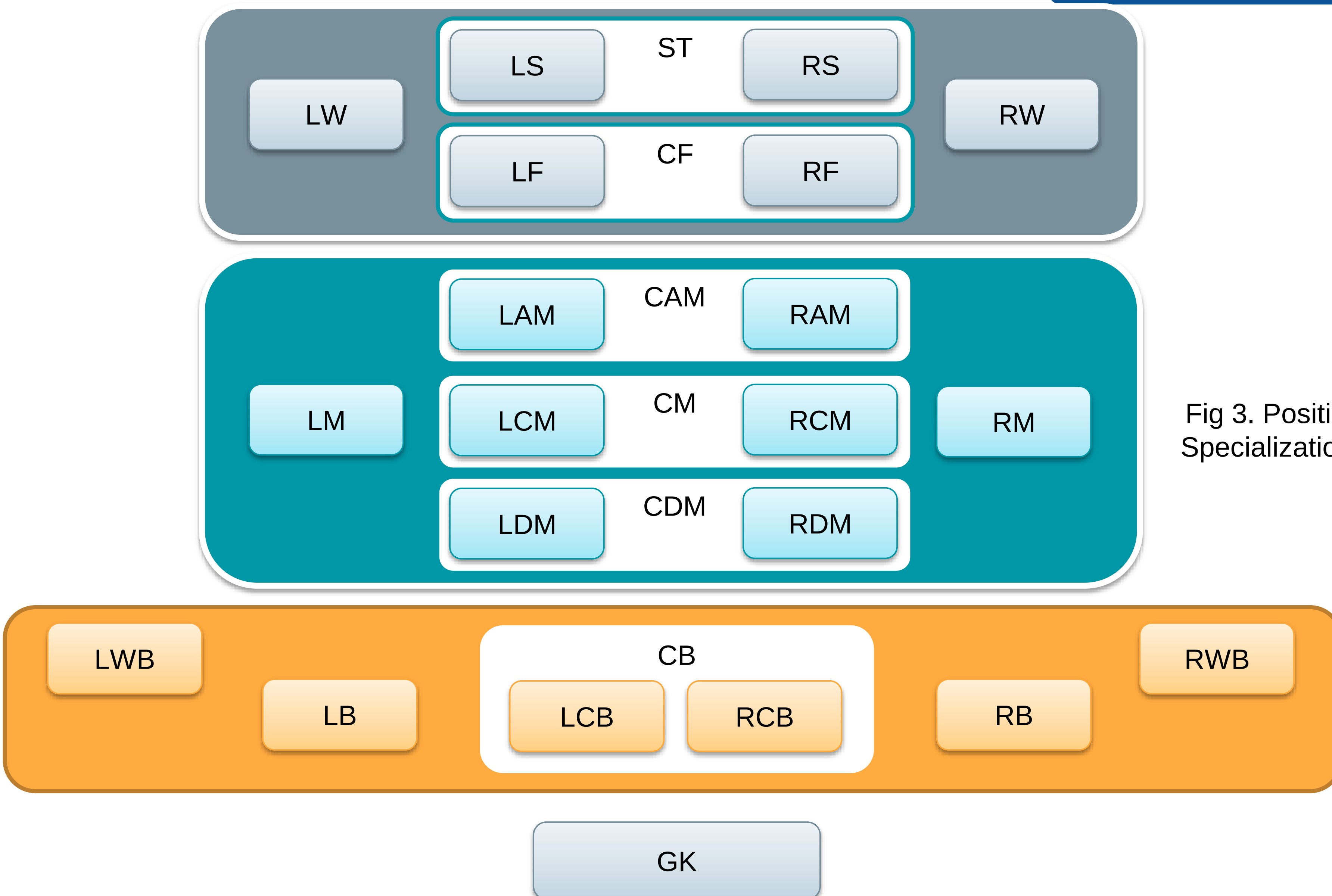


Fig 3. Position Specializations

Determining Prime Attributes Per Position

- Crossing – WB, FB, WM, CAM (0.78 – 0.88)
- Finishing – ST (0.9)
- Heading Accuracy – CB, ST (0.8 and 0.6 respectively)
- Short Passing – CF, CM (> 0.9)
- Volleys – ST, CF, CAM (0.74 – 0.83)
- Dribbling – CF, AW (> 0.9)
- Curve – CF (0.78)
- Free kick Accuracy – CF, CAM, CM (0.6 – 0.75)
- Long Passing – CM, CDM (> 0.8)
- Ball Control – ST, CF, AW, CAM, WM, CM (> 0.8)
- Acceleration – WM, AW (0.4 - 0.5)
- Sprint Speed – WM, AW (0.4 – 0.5)
- Agility – AW (0.72)

Determining Prime Attributes Per Position

- Reactions – All Positions (> 0.8)
- Balance – CF, AW (0.25 – 0.4)
- Shot Power – ST, CF, CAM (0.87 and around 0.75 for other two)
- Jumping – GK (0.42)
- Stamina – FB, WB (0.6 - 0.7)
- Strength – ST, CB (0.45 – 0.55)
- Long Shots – CAM, CF (> 0.8)
- Aggression – WB, FB, CB (0.6 – 0.75)
- Interceptions – CB (> 0.9)
- Positioning – ST, CF (> 0.9)
- Vision – CM, WM, CAM, AW, CF (> 0.8)
- Penalties – CF, CAM (0.6 – 0.7)
- Standing Tackle – CB (> 0.9)

Determining Prime Attributes Per Position

- Sliding Tackle – CB, FB, WB (> 0.8)
- GK Diving – GK (0.93)
- GK Handling – GK (0.91)
- GK Kicking – GK (0.77)
- GK Positioning – GK (0.93)
- GK Reflexes – GK (0.94)

Predicting Player Position

- Preprocessing
  - Retained only pertinent data for analysis
- Applied Decision Tree Classification on raw pre-processed data
- Test accuracy obtained : **40%**
  - Very unsatisfactory
- In order to increase accuracy, created generalizations:

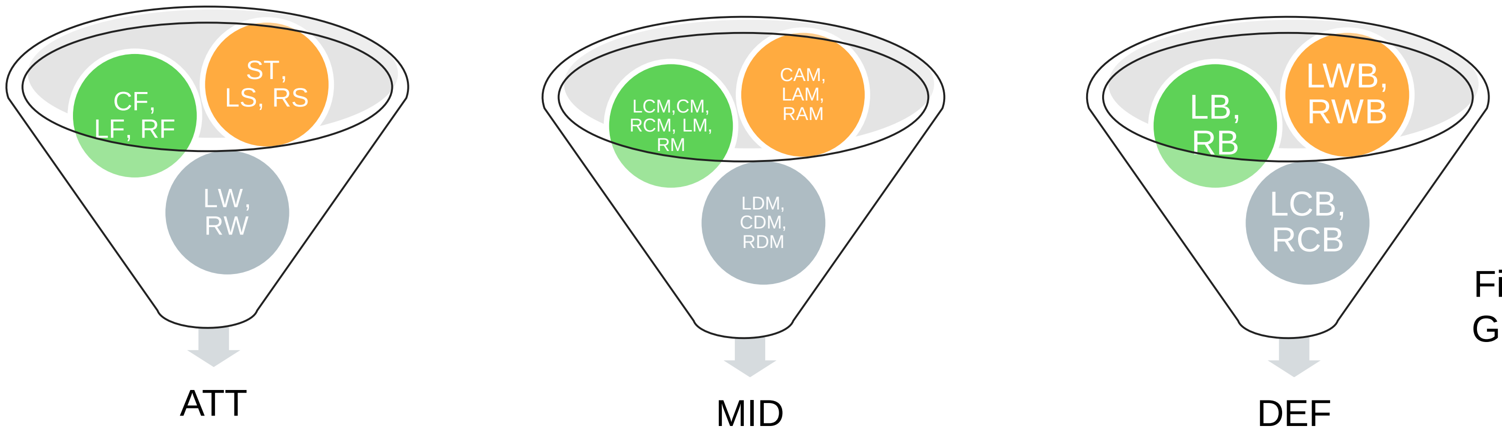


Fig 4. Position Generalization First Level

Predicting Player Position

- Applied Decision Tree Classifier on Generalized Data with the four position classes – ATT, MID, DEF, GK
- Test Accuracy obtained: **81%**
- To determine exact position, created two levels of classification
  - First check general position, then determine specialized position within the subgroup
  - Total number of Decision Tree Classification Models: **4**
    - One for the general classification
    - Three for the specialized classification
      - Attacker Classification Accuracy: **80%**
      - Midfielder Classification Accuracy: **55%**
      - Defender Classification Accuracy: **89%**

Requires probabilistic classification

Predicting Player Position

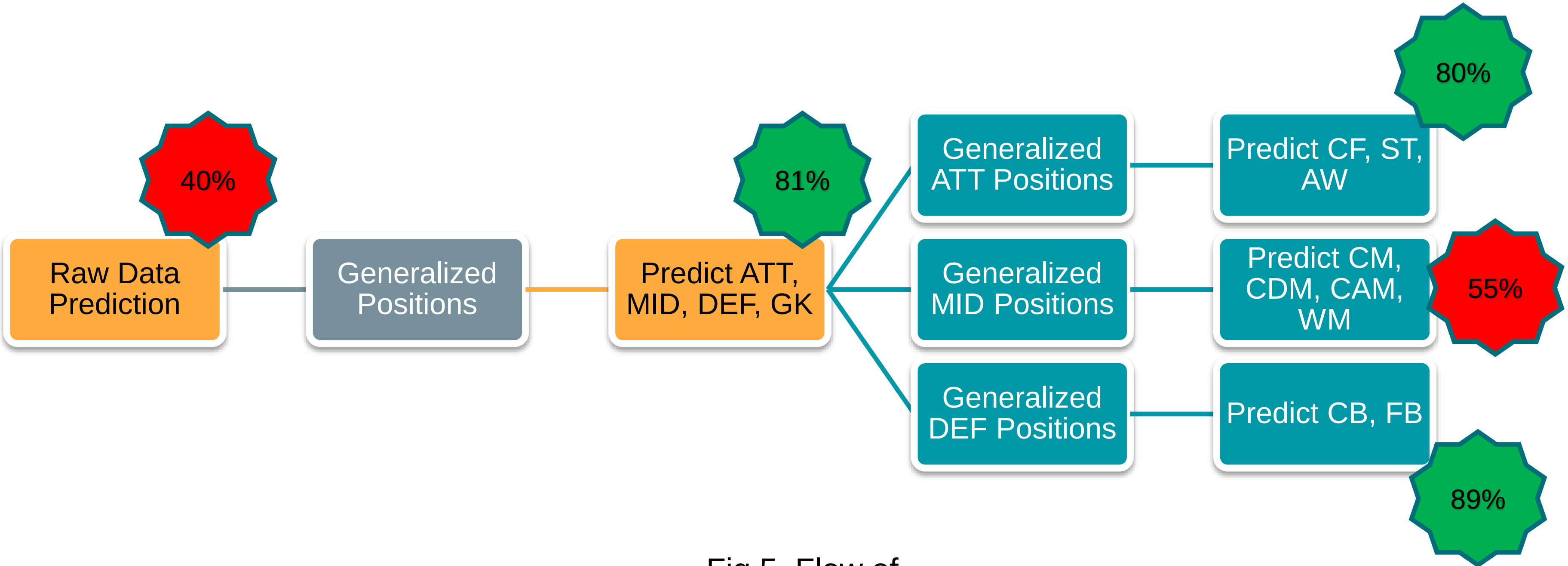


Fig 5. Flow of Prediction Operation

Predicting Player Position – Example Outputs

Messi=[[84.0, 95.0, 70.0, 90.0, 86.0, 97.0, 93.0, 94.0, 87.0, 96.0, 91.0, 86.0, 91.0, 95.0, 95.0, 85.0, 68.0, 72.0, 59.0, 94.0, 48.0, 22.0, 94.0, 94.0, 75.0, 96.0, 33.0, 28.0, 26.0, 6.0, 11.0, 15.0, 14.0, 8.0]]
Recommended Role: ATT
Recommended Position: CF

Neymar=[[79.0, 87.0, 62.0, 84.0, 84.0, 96.0, 88.0, 87.0, 78.0, 95.0, 94.0, 90.0, 96.0, 94.0, 84.0, 80.0, 61.0, 81.0, 49.0, 82.0, 56.0, 36.0, 89.0, 87.0, 81.0, 94.0, 27.0, 24.0, 33.0, 9.0, 9.0, 15.0, 15.0, 11.0]]
Recommended Role: ATT
Recommended Position: AW

Predicting Player Position – Probabilistic Classification

- In football, players can play in multiple positions, with some positions better suited to the player attributes.
- Initial generalization performed:
    - RCM, LCM -> CM
    - LDM, RDM -> CDM
    - LAM, RAM -> CAM
    - LM, RM -> WM
    - LS, RS -> ST
    - LF, RF -> CF
    - RCB, LCB -> CB
    - LB, RB, LWB, RWB -> WB

Predicting Player Position – Probabilistic Classification

- Implemented **Random Forest Classifier** – Accuracy **74%**
- Example:

Tenzin=[[80,70,50,85,65,60,60,70,70,75,70,70,75,70,65,75,65,60,60,65,70,75,70,70,85,70,70,60,60,0,0,0,0,0]

{'AW': 0.0,
 'CAM': 0.0,
 'CB': 0.0,
 'CDM': 0.6,
 'CF': 0.0,
 'CM': 0.4,
 'GK': 0.0,
 'ST': 0.0,
 'WB': 0.0,
 'WM': 0.0}

Predicting Player Position – Probabilistic Classification

- Implementing Random Forest Classification to main prediction model

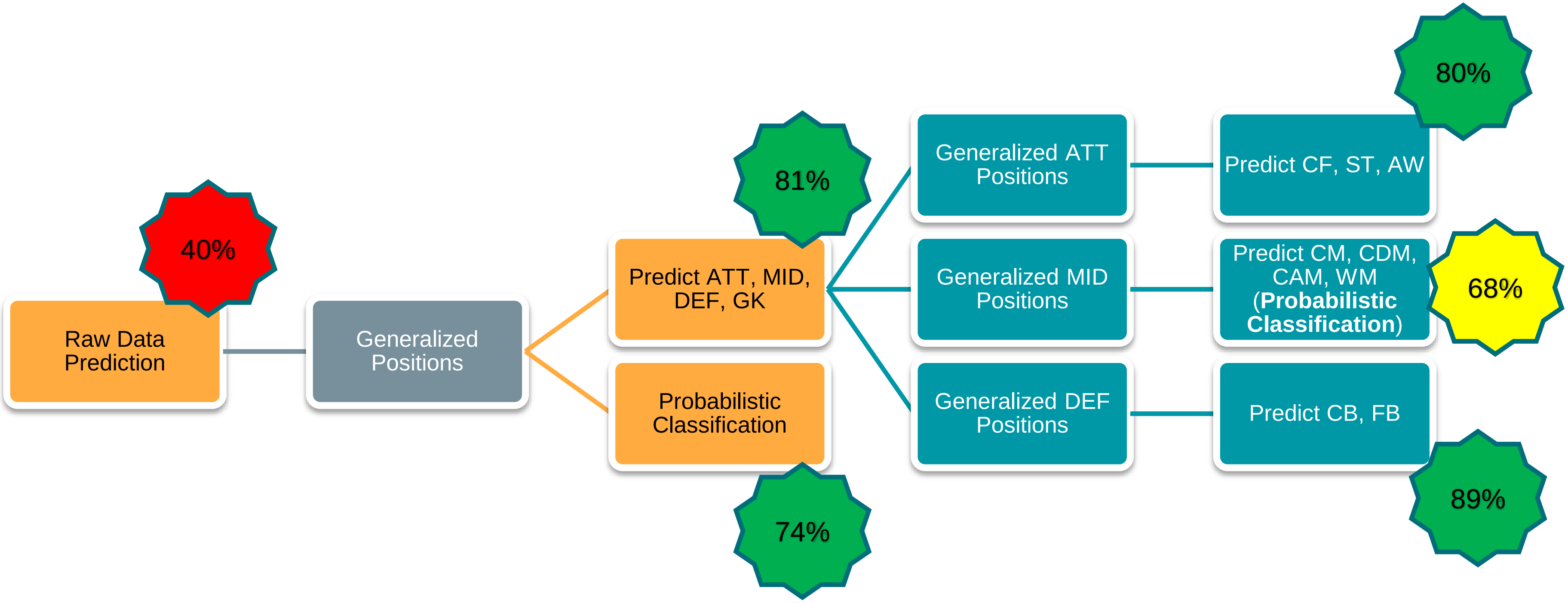


Fig 6. Flow of Prediction Operation – With Probabilistic Classification

Predicting Player Position – Probabilistic Classification

- Output:

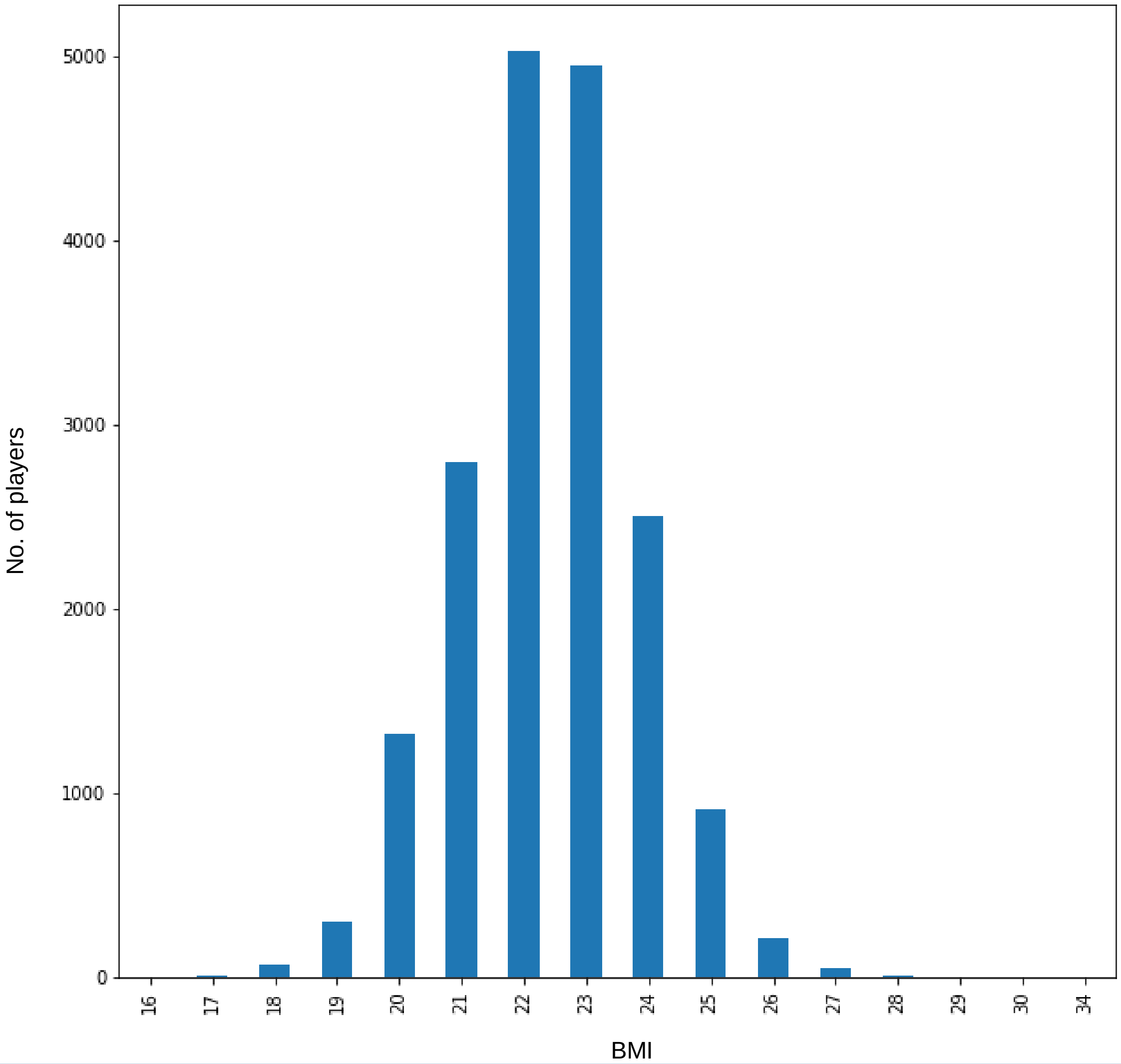
DeBruyne=[[93.0, 82.0, 55.0, 92.0, 82.0, 86.0, 85.0, 83.0, 91.0, 91.0, 78.0, 76.0, 79.0, 91.0, 77.0, 91.0, 63.0, 90.0, 75.0, 91.0, 76.0, 61.0, 87.0, 94.0, 79.0, 88.0, 68.0, 58.0, 51.0, 15.0, 13.0, 5.0, 10.0, 13.0]]

Recommended Role: MID
Recommended Position: {'CAM': 0.18, 'CDM': 0.08, 'CM': 0.66, 'WM': 0.08}

Module 2: Football Performance and BMI

- Body Mass Index is the measure of body fat based on the height and weight metrics of an individual
- Preprocessing involved:
  - Converting formats for height and weight
  - Calculating BMI
- BMI Categories:
  - <18.5 : Underweight
  - 18.5 to 25 : Normal
  - 25-30 : Overweight
  - >30 : Obese

BMI Overall Distribution



- Average BMI of all players : 22.39
- Mean BMI values Per Position found to vary minimally. Not a differentiating factor.
- Mean BMI values per player overall range found to vary minimally. Not a differentiating factor.
- Since Mean is not an effective parameter for analysis, the module adopts a histogram distribution per position

Fig 7. Overall BMI Distribution

BMI Distribution - GK

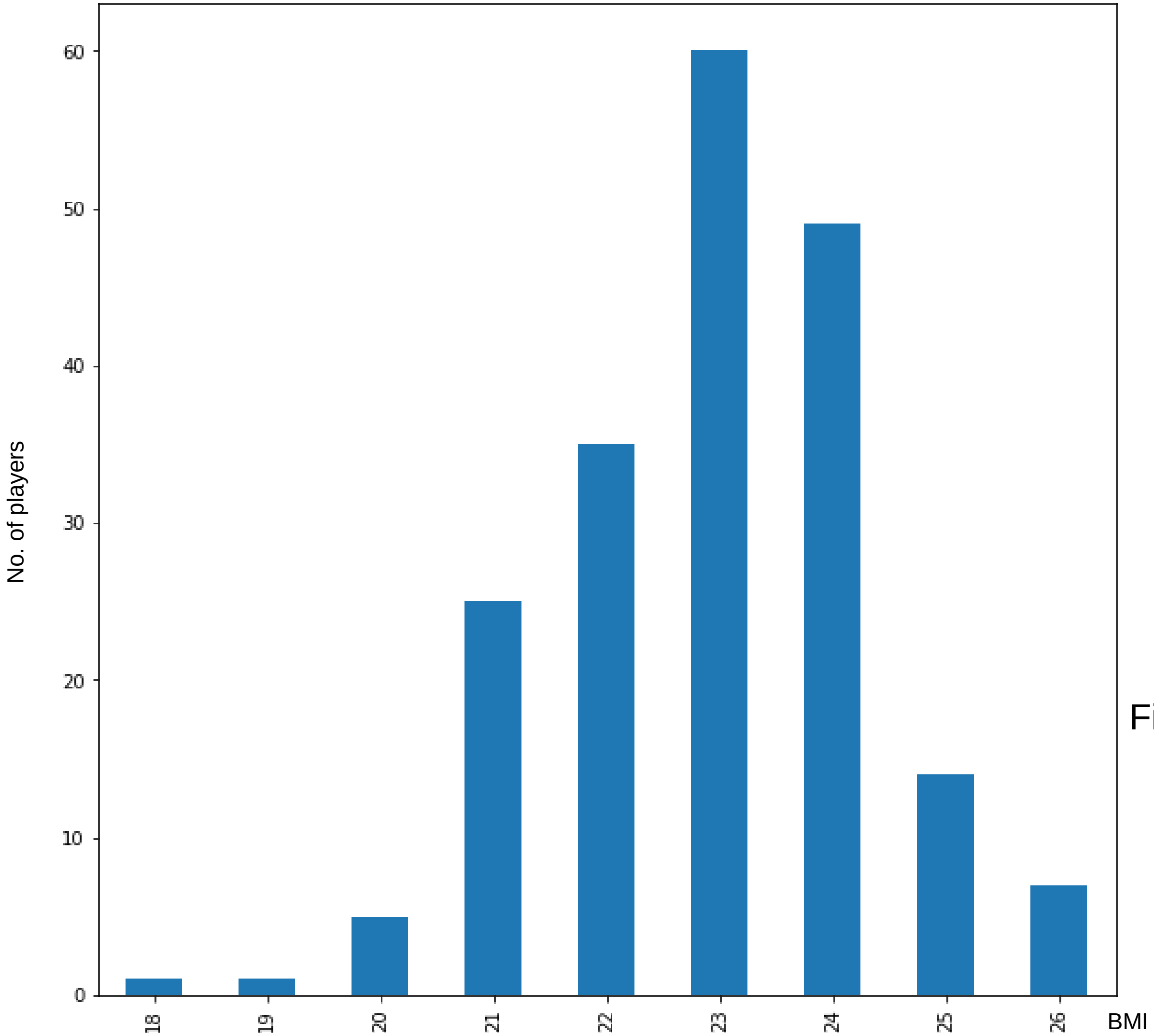


Fig 8. Goalkeepers BMI Distribution

BMI Distribution - CB

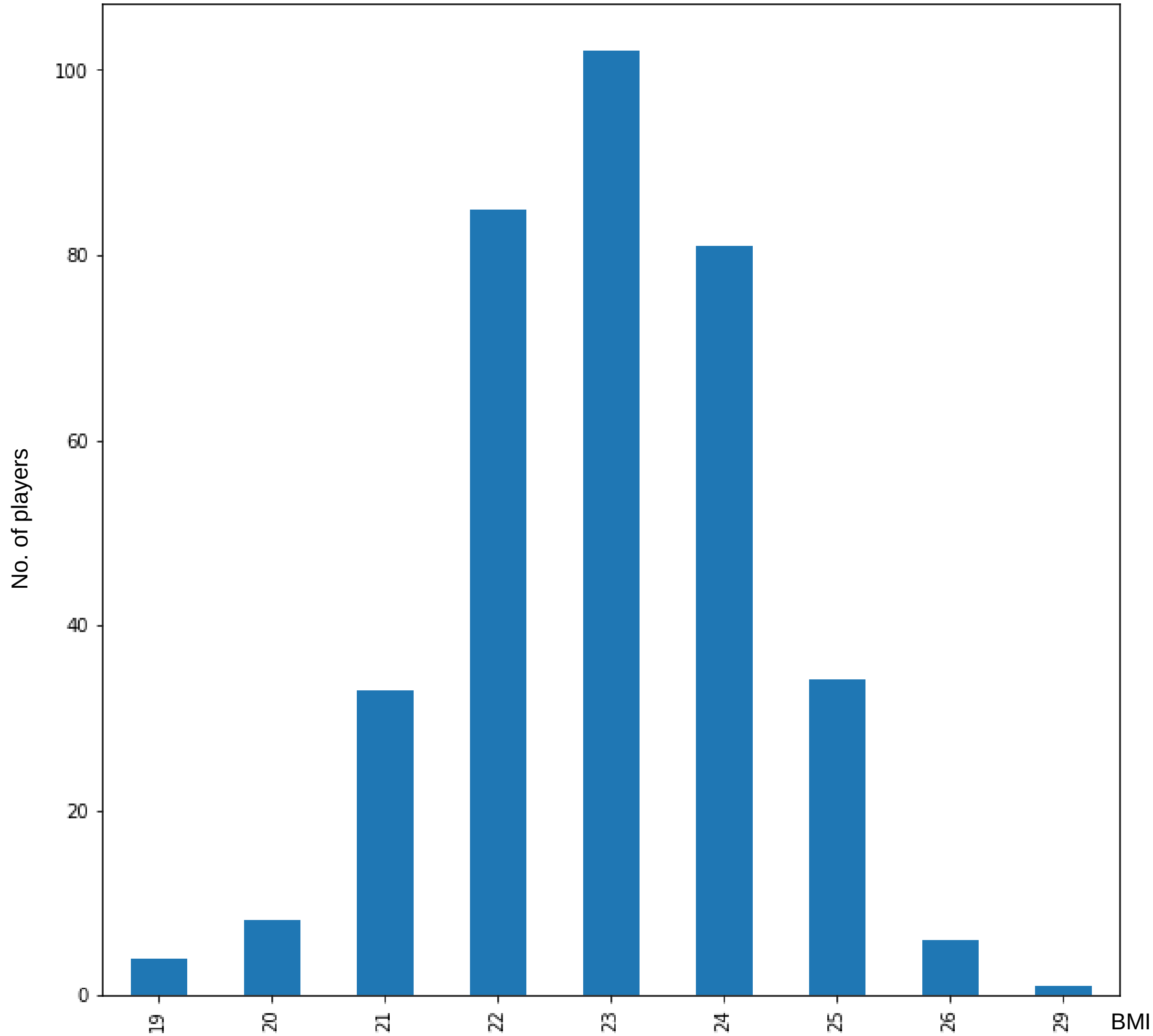


Fig 9. Center Backs BMI Distribution

BMI Distribution - WB

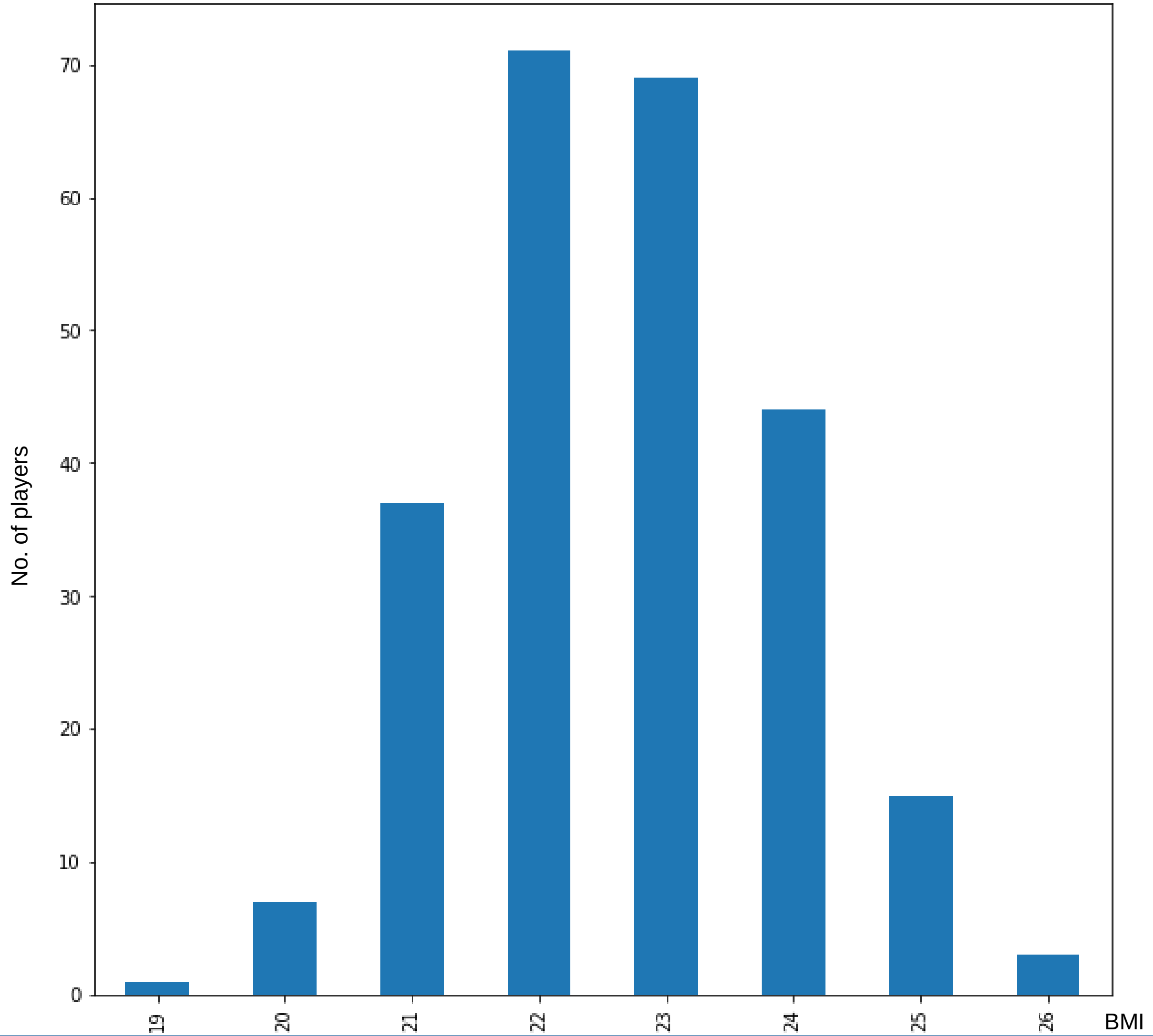


Fig 10. Wing Back BMI Distribution

BMI Distribution - CDM

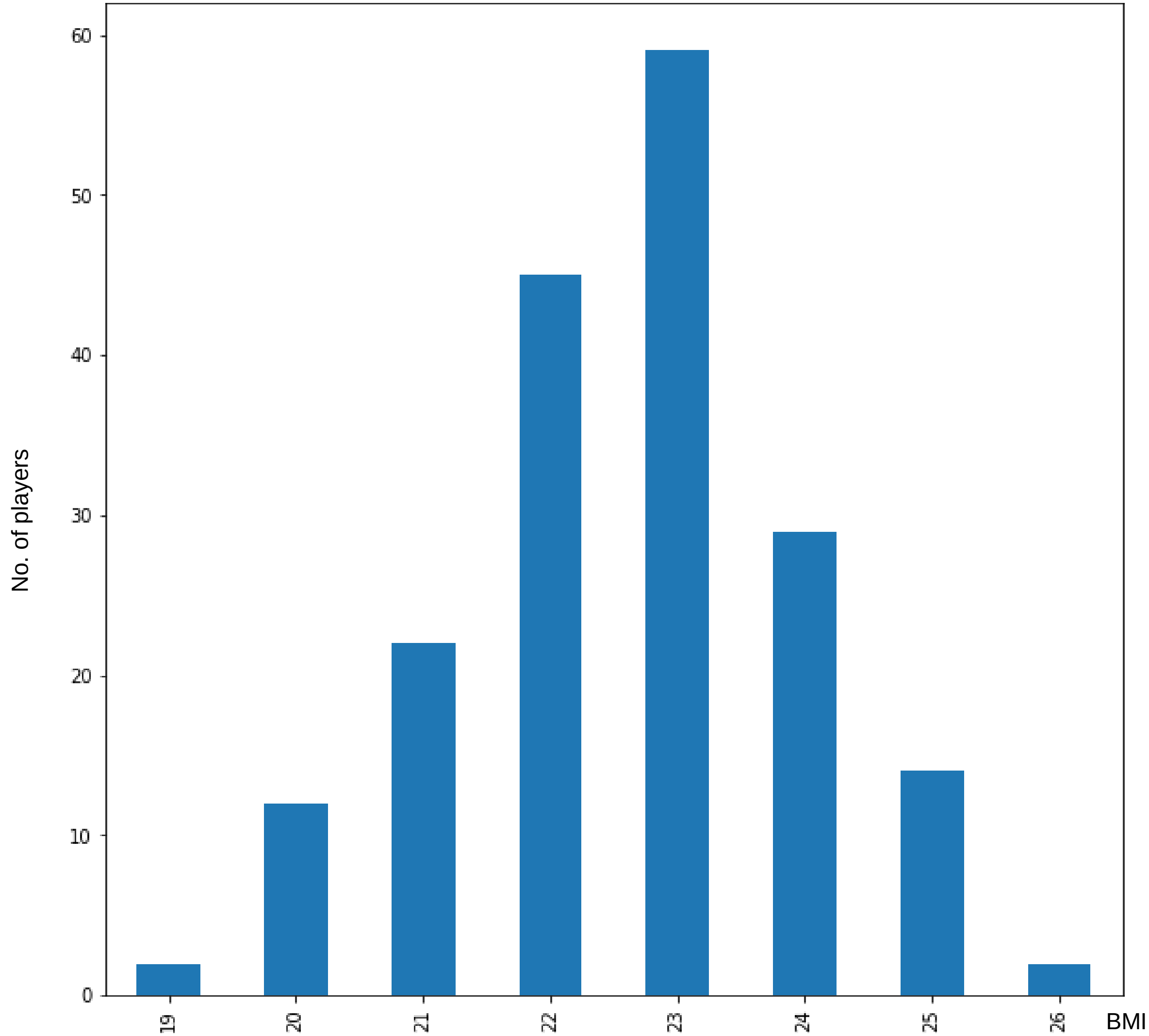


Fig 11. Center Defensive Midfielder BMI Distribution

BMI Distribution - CM

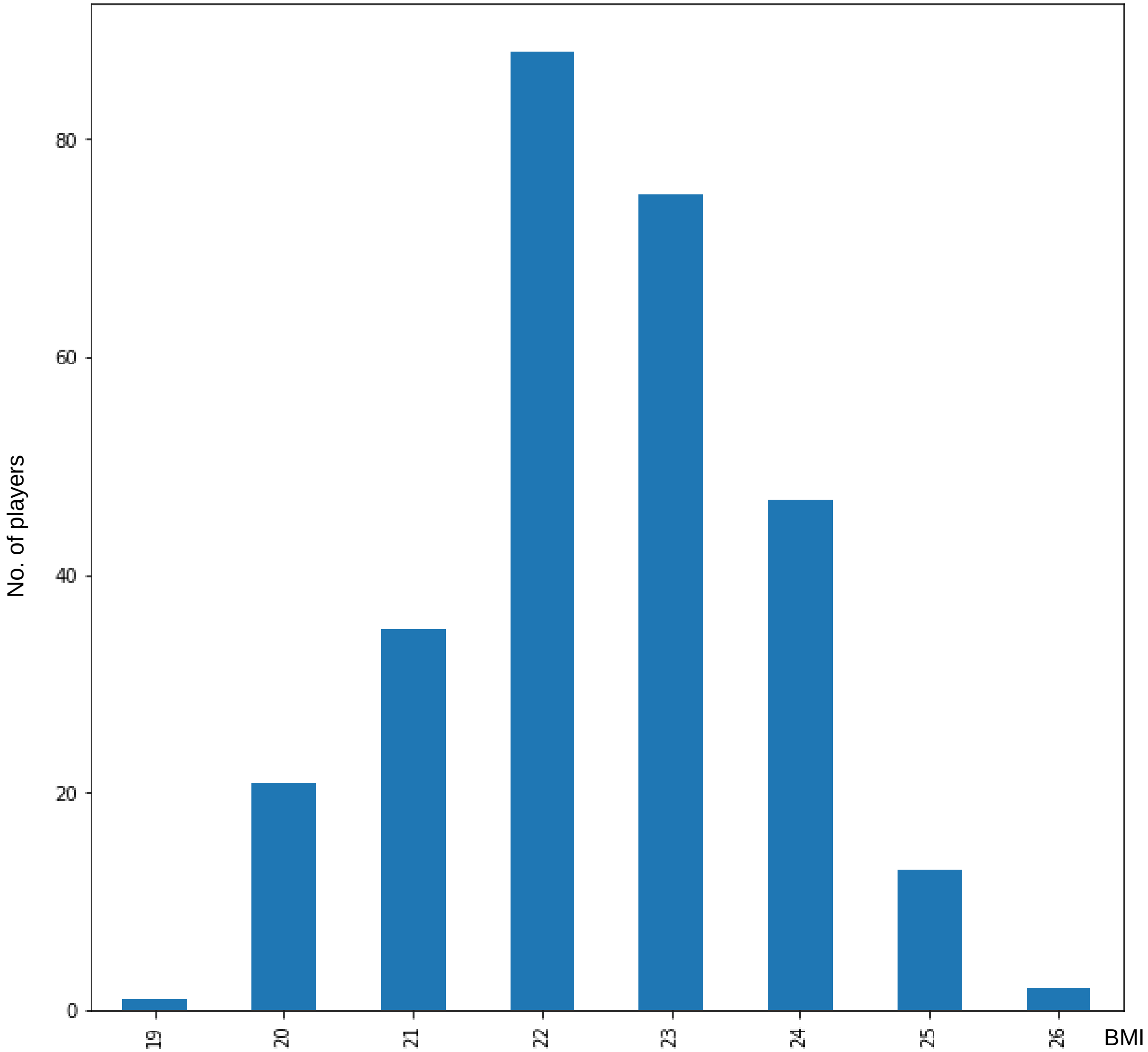


Fig 12. Center Midfielder BMI Distribution

BMI Distribution - CAM

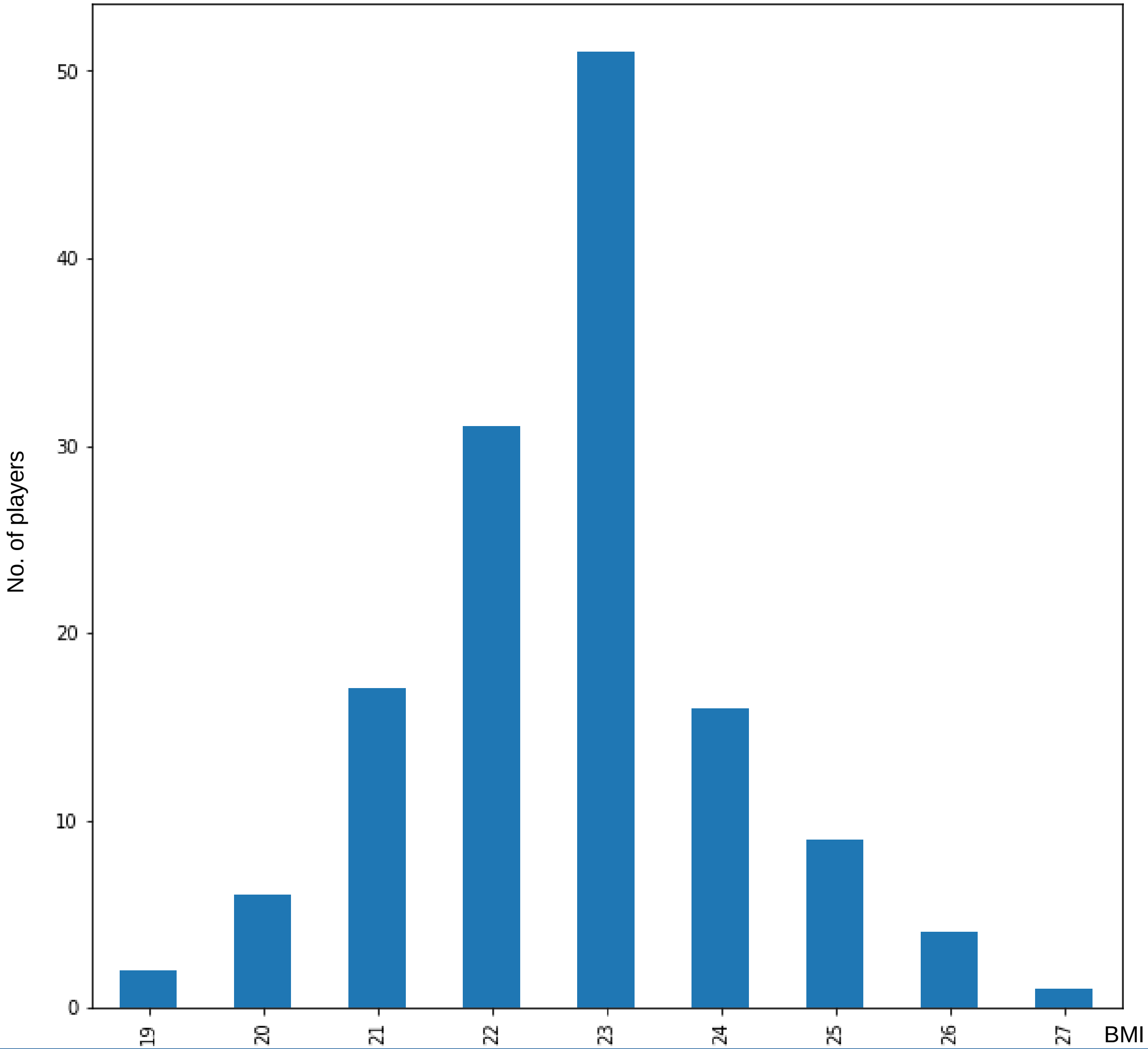


Fig 13. Center Attacking Midfielder BMI Distribution

BMI Distribution - WG

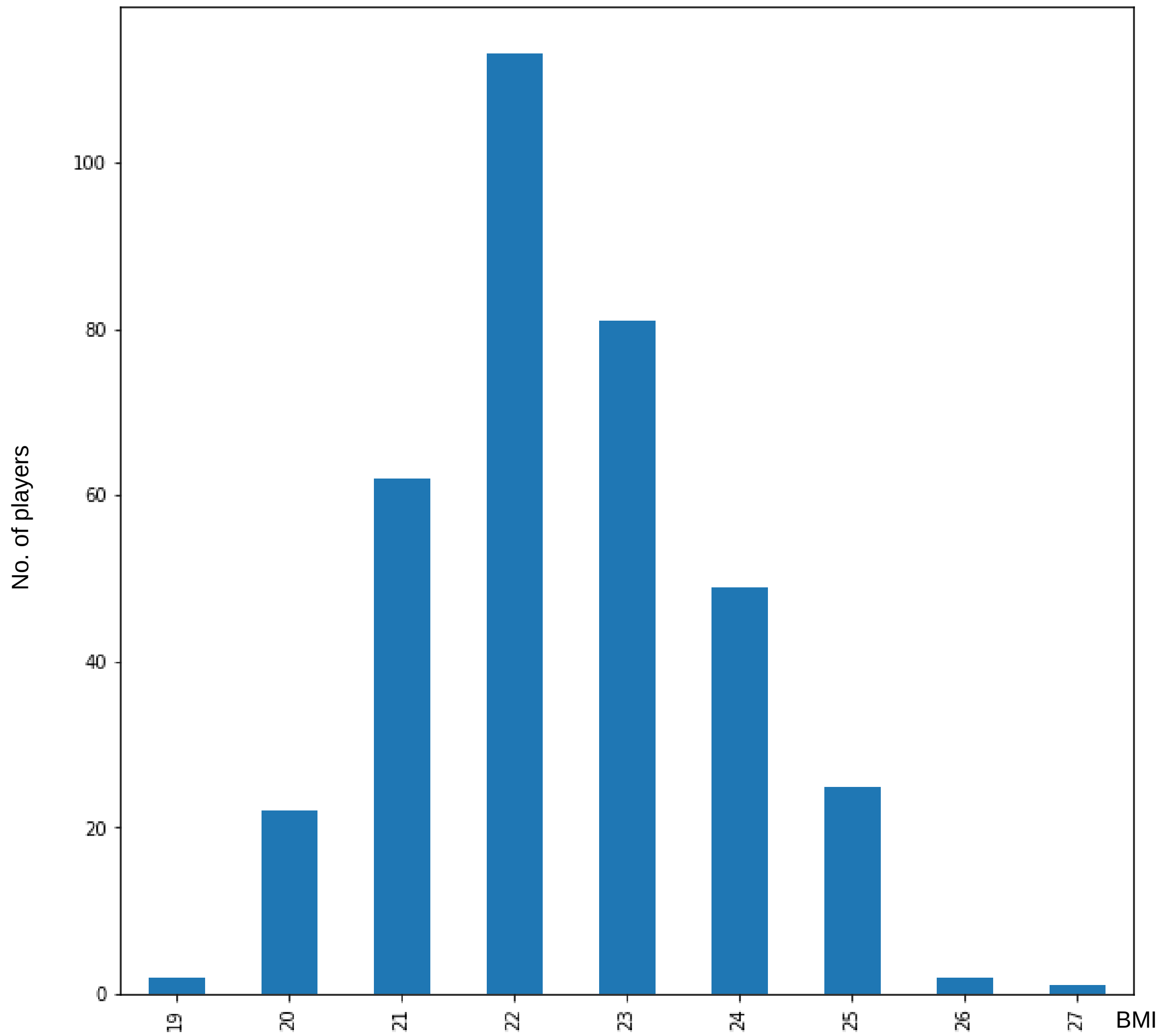


Fig 14. Winger BMI Distribution

BMI Distribution - ST

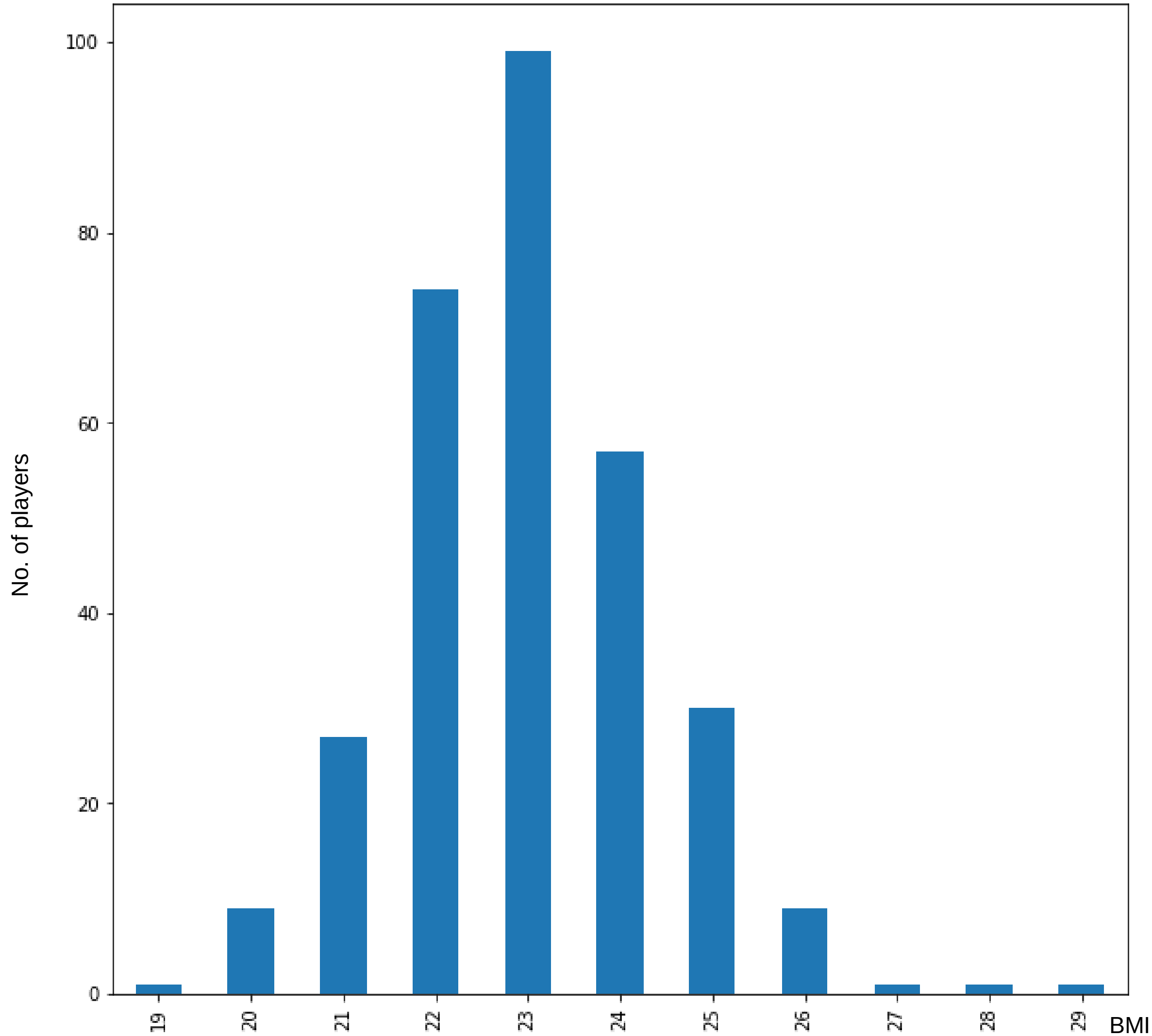


Fig 15. Striker BMI Distribution

Data Fields for Player Performance and BMI Analyser

- Player Name
- Team/League
- Player BMI
- Position
- Team Most Used Formation
- Player Rating
- Player Age
- Speed – Acceleration, Sprint Speed
- Dribbling – Ball Control, Agility
- Shooting – Finishing, Shot Power
- Physical – Stamina, Strength
- Defending – Tackle, Interception
- Passing – Short Passing, Long Passing
- Work Rate – H/M/L (Offensive/Defensive)

Some of these statistics will be used by the Progression Module (Module 3)

Module 3: Player Progression Analysis

- This module aims to create a framework that charts player progression within a team
- This progression framework highlights the following team properties
  - Player Turnover Score
  - Overall Team Progression Value
  - Position-Specific Team Progression Values
  - Contribution of Position-Specific Progression
  - Contribution of Position Flux and Youth Players to Progression Score
- Higher the Contribution of Position Flux indicates switching positions is ideal to improve player performance in this team.
  - Also indicates young players progression contributes much more than players in prime

Thank You