Applying supervised machine learning to predict optimal playing positions for rugby players

Rademeyer LC[1], Stander A[1], du Toit PJ[1], Nortje E[1], Kleynhans M[1]

1. Department of Physiology, Faculty of Health Sciences, University of Pretoria

Background:

Predicting sports positions can improve performance and reduce injuries of players

Aim:

To Predict correct playing positions for Rugby players by applying supervised machine learning techniques.

Materials & Methods:

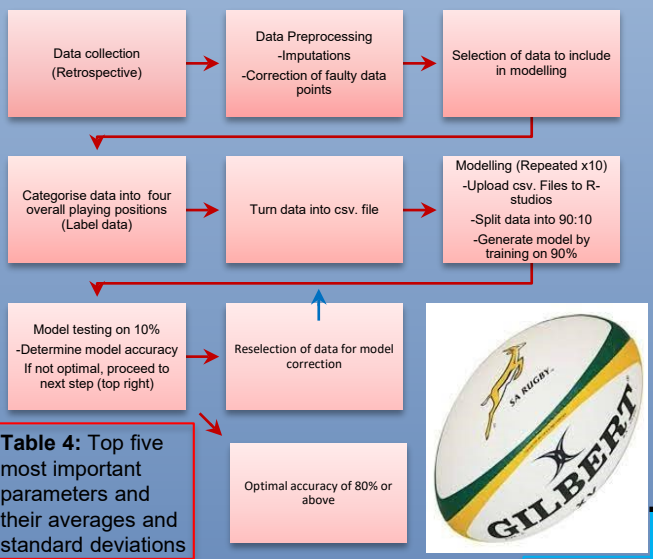

Table 4: Top five most important parameters and their averages and standard deviations

Parameters	Averages	Standard deviations
BMI	26.15	5.33
Weight (Kg)	81.81	18.56
Midaxillary circumference (cm)	14.28	7.81
Sub-cutaneous fat%	13.14	4.57
Abdominal circumference (cm)	20.1	10.31

Table 3: T-test results for top five most important parameters

T-test values \Parameters	Pos 1 vs Pos 2	Pos 1 vs Pos 3	Pos 1 vs Pos 4	Pos 2 vs Pos 3	Pos 2 vs Pos 4	Pos 3 vs Pos 4
BMI	0.85	0.79	0.84	0.93	1	0.92
Weight	0.88	0.74	0.77	0.85	0.88	0.97
Midaxcilliary circumfrence	0.66	0.37	0.43	0.39	0.5	0.75
Sub-cutaneous fat%	0.79	0.56	0.59	0.67	0.72	0.94
Abdominal circumference	0.75	0.48	0.49	0.5	0.52	0.96

Results:

1. Model accuracy came to 64.55%
2. There was a trend of heavier and larger forwards (position 1)
2. Position 3 and 4 parameters were relatively close each time.
3. Machine learning better predicted the positions than excel
4. The variation between the different positions appears to increase as the importance of the parameter decreases.

Table 1: Ten accuracies and the averages of all parameters and top twenty parameters.

Model nr	All Parameters	Top 20 Parameters
1	68.18%	72.73%
2	68.18%	72.73%
3	63.64%	59.09%
4	68.18%	77.27%
5	59.09%	63.64%
6	68.18%	63.64%
7	59.09%	63.64%
8	77.27%	68.18%
9	54.55%	54.55%
10	50%	50%
Average Accuacy	63.64%	64.55%

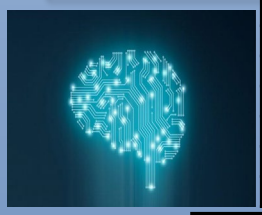

Table 2: BMI averages and standard deviations

Positions	BMI Averages (Kg)	BMI Standard deviations
Position 1 (players 1-3)	32.58	4.27
Position 2 (players 4-8)	25.84	3.45
Position 3 (player 9)	22.79	3.46
Position 4 (players 10-15)	23.89	5.26

Conclusion:

An optimal accuracy of 80% or above could not be achieved.
Output accuracy would likely improve with more data points.
Limitations: Assumption that players in the data are in the correct positions.
Future: Research output can be extended to other sporting codes.

Faculty of Health Sciences
Fakulteit Gesondheidswetenskappe
Lefapha la Disaense tša Maphelo