Basic decision tree classifying primary position of a Premier League player

Target classes: **Goalkeeper**, **Defender**, **Midfielder**, **Forward**

More intense colour reflects greater confidence of predictor

Branches left always indicate "True", branches right indicate "False"

Hyperparameter "maximum depth" set to 4

All metrics are "per 90 minutes"

The "value" array shows how many GK, DF, MF and FW (in that order) fall into this leaf from the train data.

The leaf is very light green colour to indicate a weak confidence in classifying a given player as a Defender at this level of available information.

At this point, the decision tree has not made a single assessment on the features, so the classification is trivial as it leans towards Defender purely because it is the most populous class in the train data.

Progressive Passing Distance ≤ 143.5
gini = 0.698
samples = 229
value = [15, 83, 74, 57]
class = Defender

True →

Passes Into Final Third ≤ 1.805
gini = 0.432
samples = 76
value = [0, 0, 24, 52]
class = Forward

False →

Touches Final Third ≤ 0.265
gini = 0.588
samples = 153
value = [15, 83, 50, 5]
class = Defender

Progressive Passing Distance ≤ 93.6
gini = 0.256
samples = 53
value = [0, 0, 8, 45]
class = Forward

Touches Final Third ≤ 26.7
gini = 0.423
samples = 23
value = [0, 0, 16, 7]
class = Midfielder

gini = 0.0
samples = 15
value = [15, 0, 0, 0]
class = Goalkeeper

Touches Final Third ≤ 6.16
gini = 0.506
samples = 138
value = [0, 83, 50, 5]
class = Defender

Progressive Carry Distance ≤ 85.85
gini = 0.136
samples = 41
value = [0, 0, 3, 38]
class = Forward

Touches Final Third ≤ 24.75
gini = 0.486
samples = 12
value = [0, 0, 5, 7]
class = Forward

Pass Completion Rate ≤ 78.5
gini = 0.266
samples = 19
value = [0, 0, 16, 3]
class = Midfielder

gini = 0.0
samples = 4
value = [0, 0, 0, 4]
class = Forward

gini = 0.0
samples = 39
value = [0, 39, 0, 0]
class = Defender

Crosses Into Box ≤ 0.18
gini = 0.545
samples = 99
value = [0, 44, 50, 5]
class = Midfielder

gini = 0.0
samples = 27
value = [0, 0, 0, 27]
class = Forward

gini = 0.337
samples = 14
value = [0, 0, 3, 11]
class = Forward

gini = 0.278
samples = 6
value = [0, 0, 5, 1]
class = Midfielder

gini = 0.0
samples = 6
value = [0, 0, 0, 6]
class = Forward

gini = 0.49
samples = 7
value = [0, 0, 4, 3]
class = Midfielder

gini = 0.0
samples = 12
value = [0, 0, 12, 0]
class = Midfielder

gini = 0.305
samples = 44
value = [0, 7, 36, 1]
class = Midfielder

gini = 0.477
samples = 55
value = [0, 37, 14, 4]
class = Defender

Created by Justin Lyons / @lyonjust
Data provided by FBRef as at 15 December 2021