MAE & RMSE

Mean absolute error (MAE) and Root mean square error (RMSE) are two of the most common metrics used to measure accuracy for continuous variables.

Mean absolute error (MAE) measures the average absolute difference between two continuous variables.

Root mean square error (RMSE) measures the sample standard deviation of the difference between predicted values and the actual observations.

In my example, I will try to predict the value (eur_value) of a player by using the following variables:

- 'shot_power',
- 'heading_accuracy',
- 'finishing',
- 'ball_control',
- 'age',
- 'height_cm',
- 'positioning',
- 'potential',
- 'composure'

The MAE and the RMSE of the model are:

```
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
print('MAE', mean_absolute_error(y_test, y_pred))
print('RMSE', np.sqrt(mean_squared_error(y_test, y_pred)))
```

```
MAE 178007.5020839122
RMSE 199217.24009841928
```

However by changing the ‘height_cm’ to the ‘overall’ column the MAE and RMSE have improved

```
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
print('MAE', mean_absolute_error(y_test, y_pred))
print('RMSE', np.sqrt(mean_squared_error(y_test, y_pred)))
```

```
MAE 138064.87913309253
RMSE 154165.78093260023
```

This improvement is attributed to the ‘overall’ column being more significant than the ‘height_cm’ column. Logically, the overall skill of the player is more likely to influence the value of a player rather than their height.

The variable importance are shown below:

- Height column

	0	1
7	potential	0.667787
4	age	0.142825
3	ball_control	0.077919
8	composure	0.059917
6	positioning	0.022850
0	shot_power	0.009341
2	finishing	0.009209
1	heading_accuracy	0.005976
5	height_cm	0.004176

- Overall column

	0	1
5	overall	0.830614
7	potential	0.099880
4	age	0.030214
6	positioning	0.012251
2	finishing	0.008220
3	ball_control	0.007193
0	shot_power	0.004886
8	composure	0.004043
1	heading_accuracy	0.002699