A Prediction of Youth Football Player's Future Success using Machine Learning

Grant Isdale
MSc in Information Technology

Background

- Historical data of academy youth football player's physical performance and biometric tests were obtained from a client, along with an indication of whether that player signed a full-time senior professional contract ('success') at some point following their time in the academy

- The client wanted to answer 3 main questions:
  1. *Can you predict success from the performance in these tests?*
  2. *Which tests are most indicative of success?*
  3. *Does time spent within the academy affect the likelihood of success?*

Data

Variables	Data Type
Player ID	Numeric
Date of Birth	Numeric
Date of Test	Numeric
Age (years)	Numeric
Height (cm)	Numeric
Weight (kg)	Numeric
5m Sprint (s)	Numeric
10m Sprint (s)	Numeric
20m Sprint (s)	Numeric
CMJ (cm)	Numeric
Yo-Yo (level)	Numeric
Success (Signed Professional)	Category (Y/N)

Figure 1: Table of data given by client

- Data was normalised within age groups using a z-score

Prediction of Success

- *Random Forest:*

Failure	Success	←Classified as
440	3	Failure
89	1	Success

Test Accuracy = 82.7%

- *Multilayer Perceptron:*

Failure	Success	←Classified as
443	0	Failure
91	0	Success

Test Accuracy = 82.9%

- *Naive Bayes:*

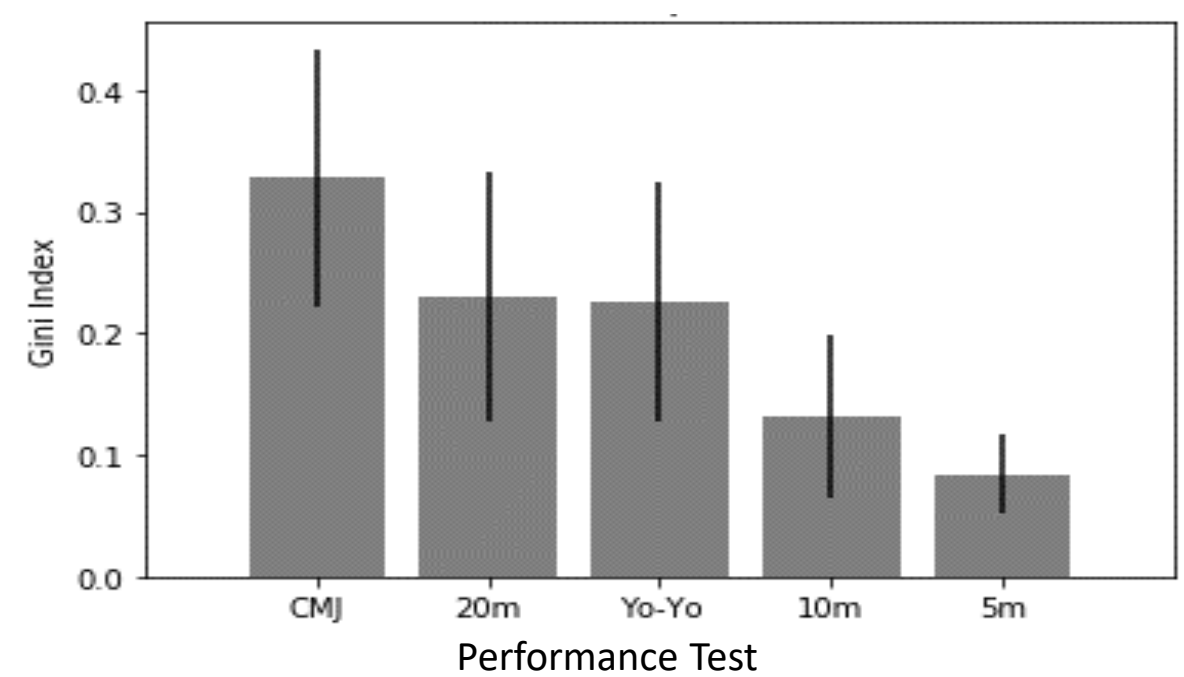| Failure	Success	←Classified as
443	0	Failure
91	0	Success

Most Indicative Tests

Figure 2: Bar chart of importance of each test on success

Time Spent In Academy vs. Success

- Correlation: r = 0.303, n = 497, p = 0.01
- Chi Squared: $X^2(11) = 62.2$, $p < 0.001$

Conclusion

- As indicated in the confusion matrices, the machine learning algorithms were unable to classify or make any predictions from the data
- Given the correlations in the data, the algorithms may make classifications if more data was availible
- The most indicative test seemed to be the CMJ (counter movement jump), a test of leg power, which supports previous findings from client that leg power/agility test scores are most important
- There is a small positive correlation and association between years spent in the academy and success