Exploring Correlations in Football Player Metrics and Providing Training Recommendations: A Machine Learning Approach

***Abstract:***

Football athlete training is a pivotal aspect of the sport, and crafting an effective training program poses a significant challenge due to the multitude of performance metrics and the variety of player positions. It is difficult to establish a unified training plan that caters to the diverse needs of all players. Our research employs factor analysis, unsupervised clustering, and machine learning algorithms to investigate the interrelationships among football player metrics. This approach enables us to provide tailored training focal points for football players of different positions and training levels. By understanding these correlations, we can offer more personalized and effective training recommendations, thereby enhancing player performance and overall team success.

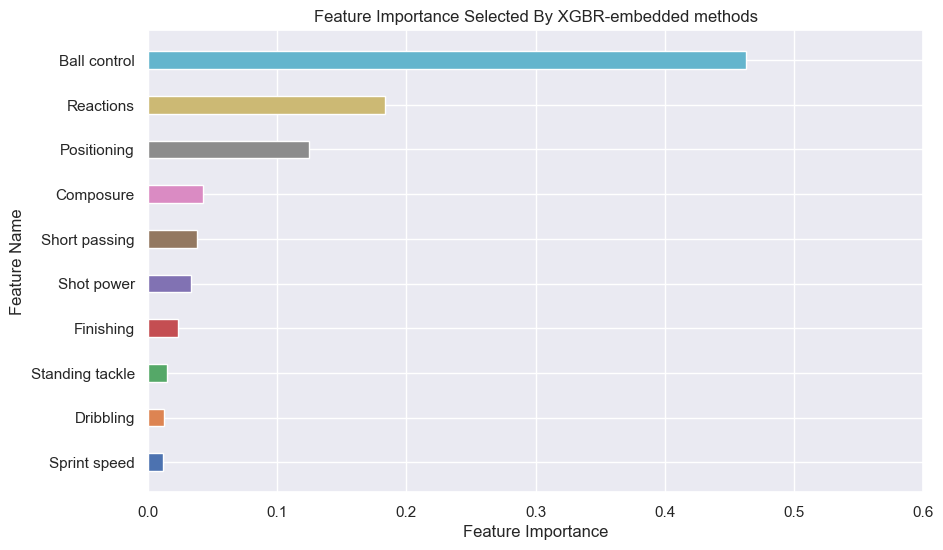
***Methedology:***

XGBR-embedded feature selection, Factor Analysis, Unsupervised Clustering(Agglomerative Clustering, Birch Clustering, GaussianMixture Clustering, Kmeans, MiniBatch Kmeans), Machine Learning Algorithms (XGBR, RandomForest, SVR, Lasso, Ridge, Elastic), Permutaion Importance, SHAP

***Result of Experiment:***

# Analysis of Forwards

Feature selection:

Bartlett's Test of Sphericity

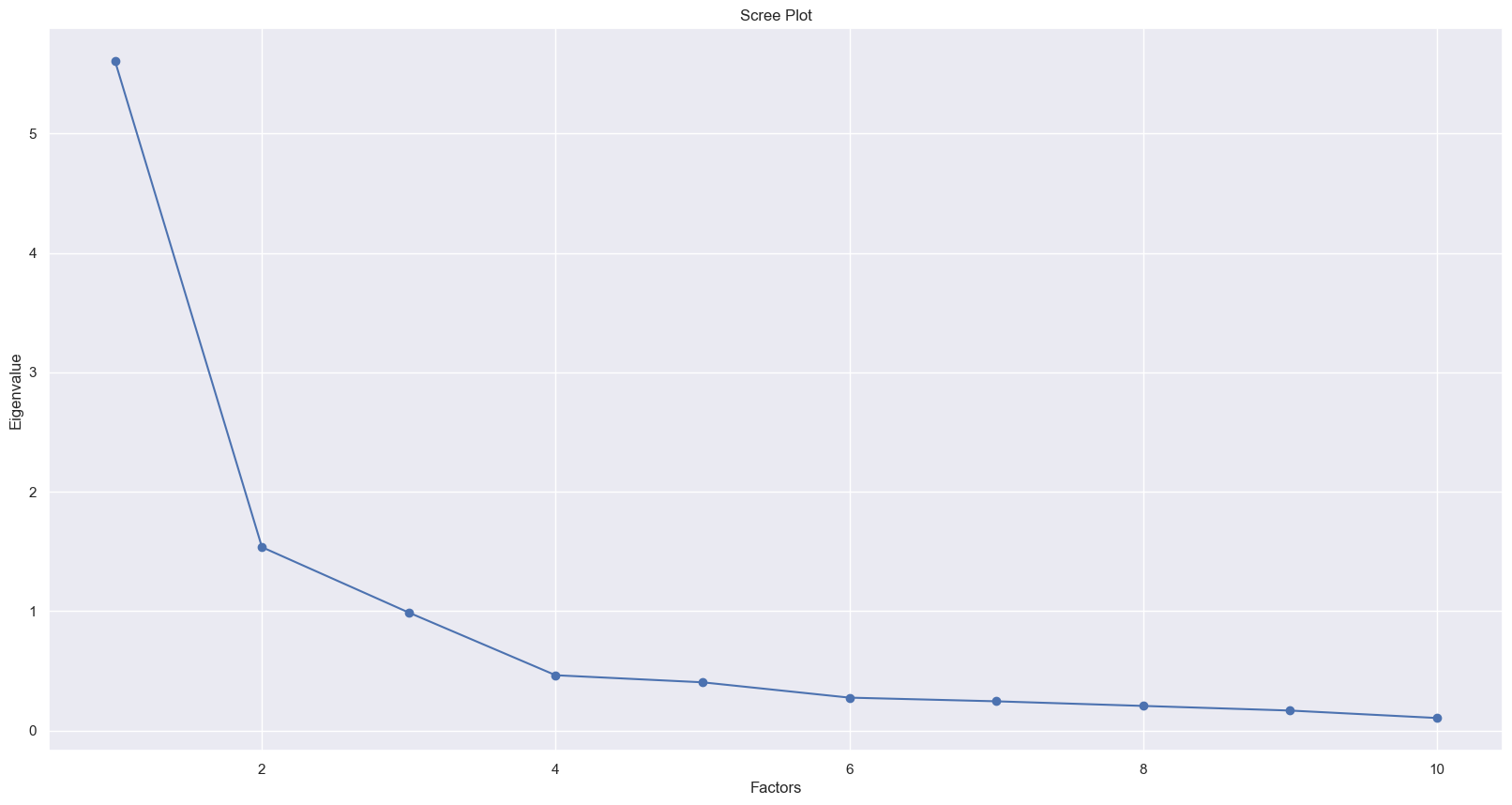
34994.841069697286, 0.0

Kaiser-Meyer-Olkin (KMO) Test

0.882802747898057

Scree Plot

We pick two features



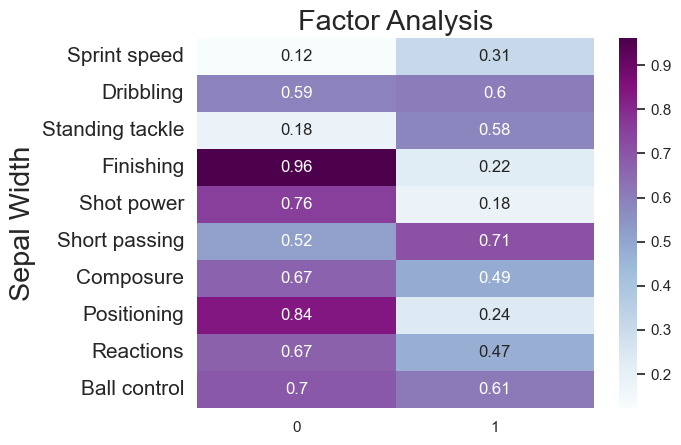
Calculate the explained variance of the factors

[0.56048257 0.71409772 0.81294401 0.85931838 0.89976907 0.92738236

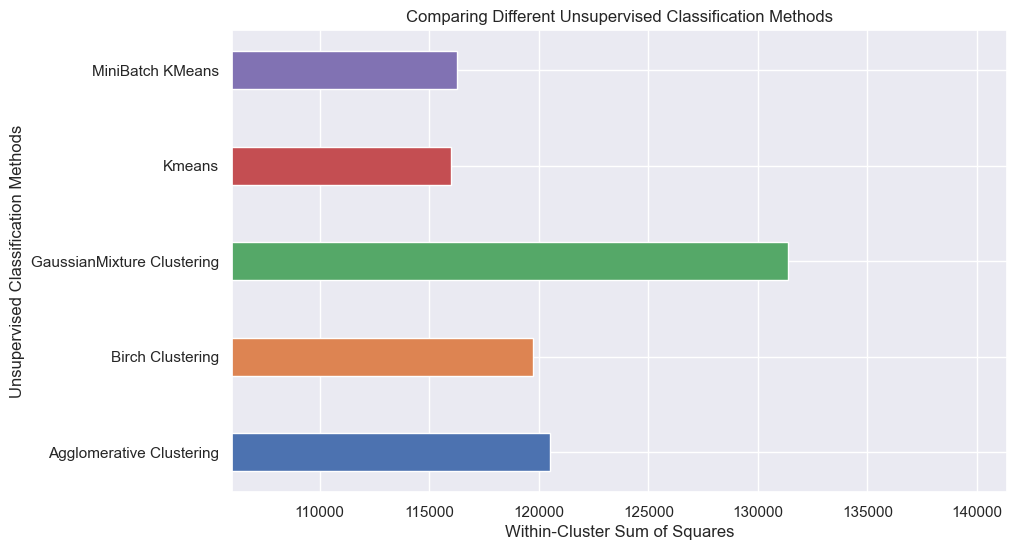
0.95192873 0.97261514 0.98942087 1. ]

累计方差解释： 0.7140977210018191

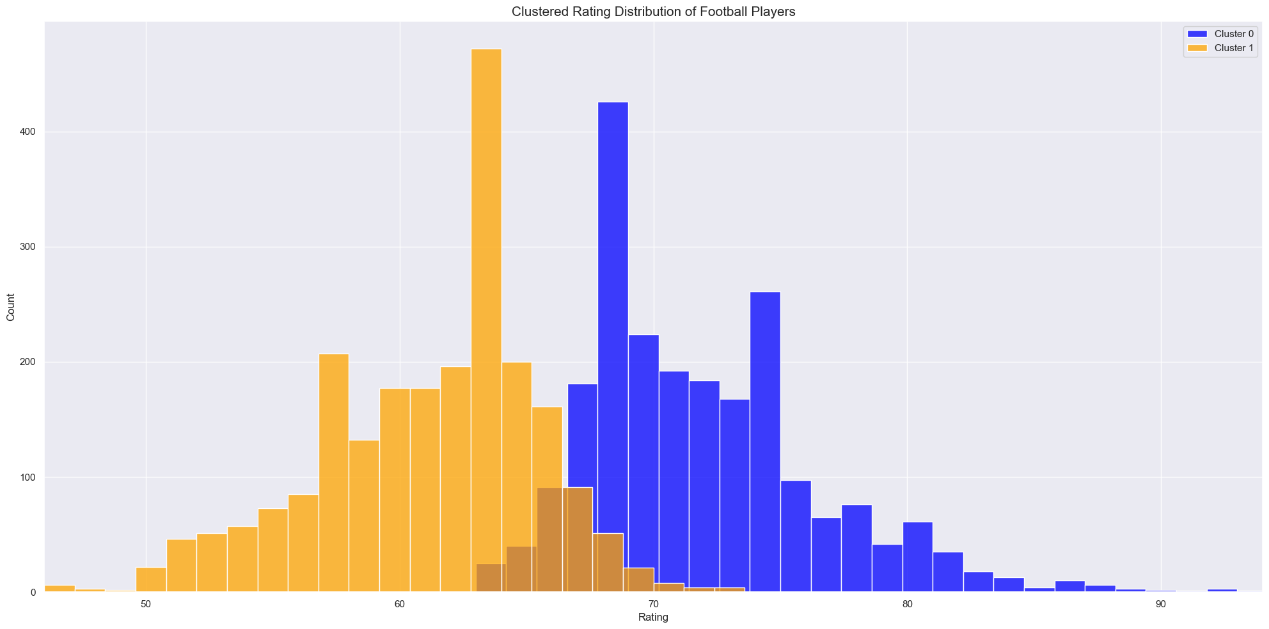
Result of Factor analysis



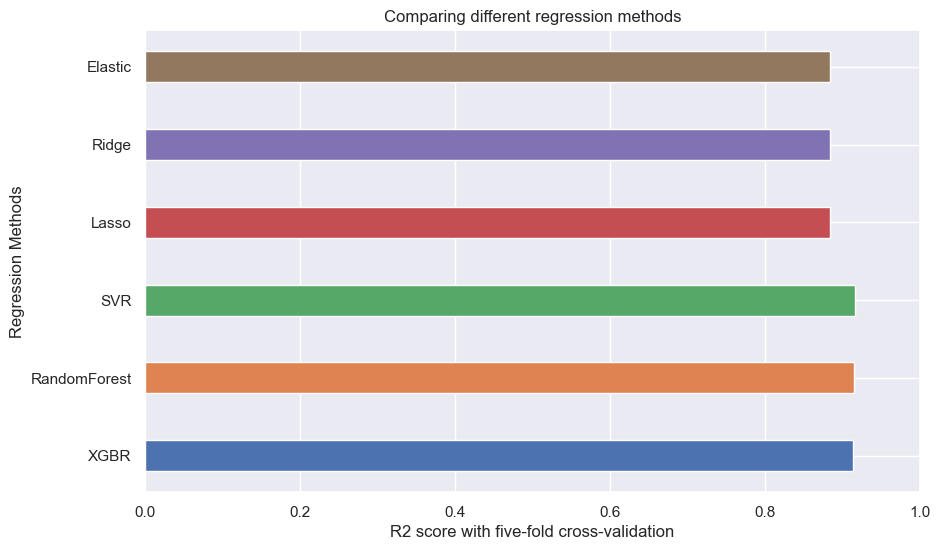
Find the best Unsupervised clustering

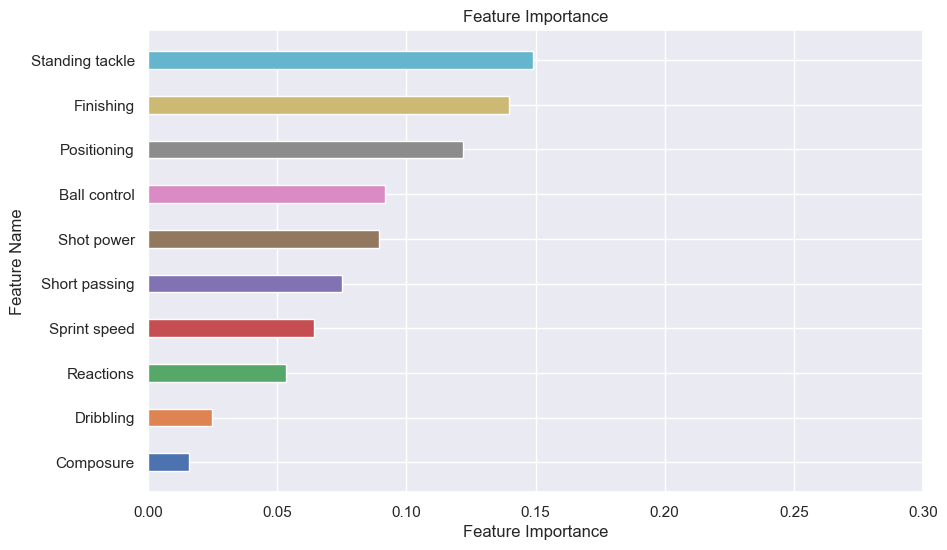


Result of Unsupervised Clustering:

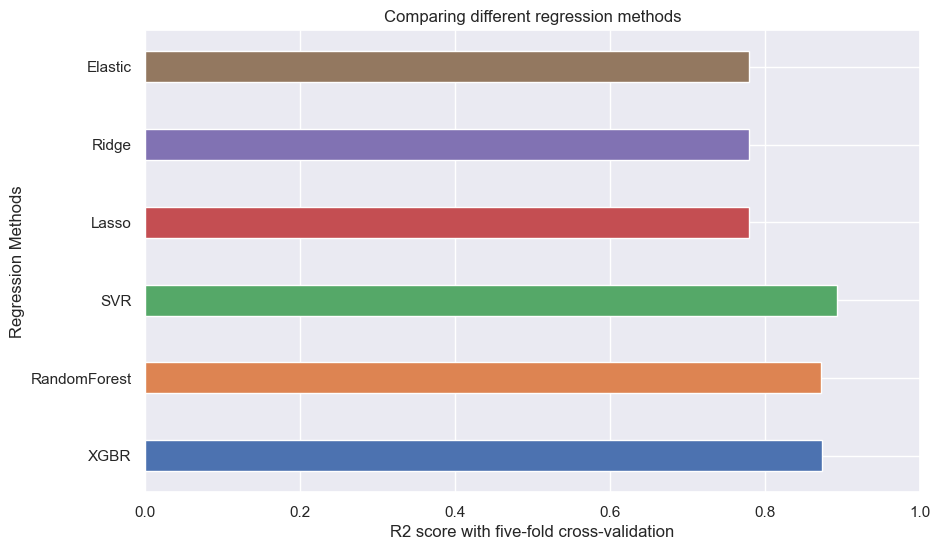


Find the best regression model of Cluster 0:

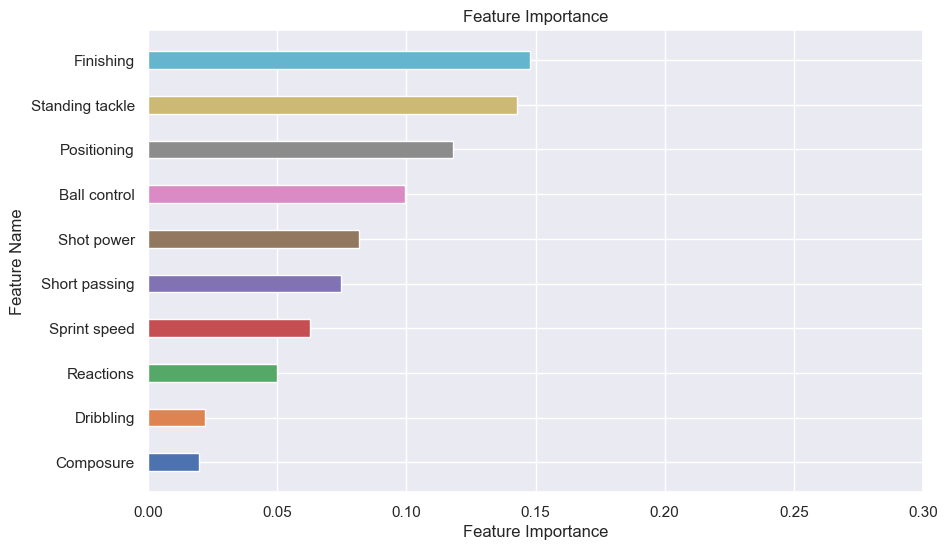


We use permutation importance to find the feature importance of SVR:

Find the best regression model of Cluster 1:



We use permutation importance to find the feature importance of SVR:



# Analysis of Midfielders

# Analysis of Defenders

# Analysis of Goalkeepers