MISSISSIPPI STATE UNIVERSITY

ATHLETE ENGINEERING, BAGLEY COLLEGE OF ENGINEERING

MODULE #4 ASSIGNMENT

IE-6990-553 SPECIAL TOPIC; DATA SCIENCE IN THE SPORT ECOSYSTEM: PERFORMANCE

DR. ADAM PETWAY

By

CHRISTOPHER IZENOUR

MANASSAS, VA

22 APRIL 2024

**Introduction**

Using historic NBA Draft Combine data for the 2000 – 2018 drafts, an analysis was completed to develop the draft board for the Top 5 prospects for a hybrid wing position. The team requires a player that is capable of guarding multiple positions on the floor. Key characteristics considered for this hybrid wing position include:

* Speed and agility to guard on the perimeter
* Size and strength to guard the blocks and rim
* Jumping ability to contest shots, protect the rim, and rebound

This analysis is important to the sport ecosystem because identifying college prospects for an upcoming draft can affect the trajectory of a team and influence the style of play for the team. Drafting a versatile hybrid wing capable of guarding multiple positions creates defensive flexibilities and forces the opposing team to alter offensive tactics to account for a defender that is not one dimensional.

**Dataset & Variables Considered**

The original dataset contained nearly 1,300 players. Because different years featured different tests and measurements, the dataset is an incomplete record. For this analysis, the dataset was reduced to 609 players, which complete data for each test and measurement considered. The tests and measurements included anthropometric measurements and strength & agility tests. Variables were converted to the metric system.

Hand width and hand length were not utilized due to the relatively low number of observations in the dataset (547). Including those measures, when combined with other measures that had missing data decreased the number of observations available for this analysis to about 230.

The analysis also utilized height w/o shoes because it afforded more observations than did height w/ shoes (1224 vs 921). The correlation between the two variables was 0.997.

**Anthropometric Measurements Utilized:**

* Weight
* Body Fat
* Height w/o Shoes
* Standing Reach
* Wingspan

**Strength & Agility Tests Utilized:**

* Max Bench Press
* Countermovement Jump
* Max Vertical Jump
* Three Quarter Sprint
* Lane Agility Time

Several derived measures were developed to ascertain anthropometric ratios and velocities, accelerations, forces and energies associated with strength and agility tests.

**Derived Anthropometric Measurements:**

* Standing Reach to Height w/o Shoes Ratio
* Wingspan to Height w/o Shoes Ratio

**Derived Strength & Agility Metrics:**

* Countermovement Jump Potential Energy
* Countermovement Jump Estimated Takeoff Velocity
* Max Vertical Jump Takeoff Velocity
* Max Vertical Jump Potential Energy
* Three Quarter Sprint Velocity
* Three Quarter Sprint Mean Acceleration
* Three Quarter Sprint Force

Analysis of player velocities, accelerations, forces and potential energies were calculated to aid in understanding how the athletes accomplished his time or height in the associated test. Comparing athletes’ derivative velocities, accelerations, forces, and energies provided greater insight into how well the athlete fit into the team’s hybrid wing requirements. Athletes with greater velocities, accelerations, forces and energies exhibited greater propensity for speed, agility and jumping, all of which are necessary for the hybrid wing.

**Methodology**

The anthropometric and strength & agility data for 609 athletes was analyzed using two machine learning algorithms (KMeans and Gaussian Mixture Model). The two algorithms created clusters of athletes that shared common characteristics. The Gaussian Mixture Model performed best. Each cluster’s anthropometric and strength & agility variables were averaged. These averages were then assessed to determine which cluster’s average characteristics best-fit the team’s requirements for the hybrid wing position. After selecting the best-fitting cluster, the top five prospects of that cluster were chosen based on the probability that the individual athlete’s data fit within that cluster. The probability is associated with the proximity of the athlete to the center of the cluster. The greater the probability (closer to the center of the cluster) the more representative the athlete is of the cluster.

**Findings and Draft Recommendation**

Of the five clusters found by the Gaussian Mixture Model, Cluster 2 exhibited characteristics that best-fit the requirements of the hybrid wing position. After rank ordering the 126 athletes of Cluster 2, the top five athletes are:

1. Justin Anderson
2. Alton Ford
3. Harrison Barnes
4. Doug Wrenn
5. Iman Shumpert

**Discussion**

**Cluster 2 Position Composition**

On average, Cluster 2’s characteristics best-fit the requirements of the hybrid-wing position. Most athletes are classified as either a type of forward or a combination of forward and shooting guard (Figure 1). These position classifications are best-suited for guarding multiple positions.

A table with numbers and letters

Description automatically generated

Figure 1

**Cluster 2 Anthropometric Characteristics**

On average, Cluster 2 did not exhibit the highest or lowest weight, body fat, height or wingspan. All five clusters showed nearly identical Standing Reach to Height w/o Shoes ratios, but Cluster 2 exhibited the greatest Wingspan to Height w/o Shoes Ratio (Figure 2).

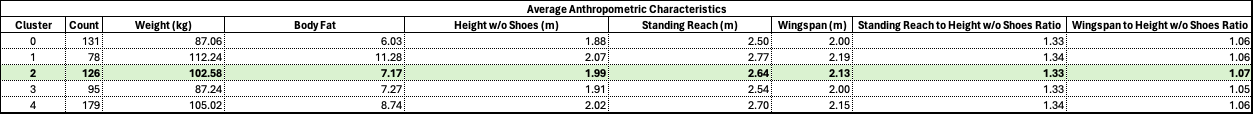


Figure 2

Using a boxplot to compare the Wingspan to Height w/o Shoes Ratio values by cluster (Figure 3), Cluster 2 exhibited a minimally higher median ratio relative to the other clusters. Cluster two’s green box is also higher than the other clusters, suggesting that the distribution of athlete heights is taller than other clusters.

A diagram of a box diagram

Description automatically generated with medium confidence

Figure 3

Using a radar chart, we can look at the anthropometric measures in normalized form. Values are scaled between 1 and 0. The cluster with the greatest value for a metric will be a 1 and the cluster with the lowest value will be a 0. Using normalized data scaled between 0 and 1 can assist with showing the relative differences between values all on the same scale.

From the radar plot (Figure 4), Cluster 2 (green) generally falls in the middle of the cluster, except with respect to the Wingspan to Height w/o Shoes ratio. As a cluster, Cluster 2 exhibits a length characteristic that is not present in other groups. This ratio indicates that the cluster has a length characteristic that is necessary to fulfill the hybrid wing requirement.

Relative to other athletes of similar height in other clusters, athletes from this cluster will have the length to contest passing lanes on both the perimeter and in the lane, contest shots, and cause turnovers.

A diagram of a person's body

Description automatically generated with medium confidence

Figure 4

**Cluster 2 Strength & Agility Characteristics**

Cluster 2’s strength and agility characteristics exhibited superior potential to fulfill the hybrid wing requirements relative to the other clusters (Figure 5). As a cluster, the athletes exhibited superior:

* Upper body strength as measured by bench press
* Jump height as measured by the max vertical jump and countermovement jump
* Potential energy at the peak of their jump tests (max vjump and cmj)
* Force production in the three quarter sprint

A screenshot of a computer

Description automatically generated

Figure 5

Athletes from this cluster exhibit the potential to handle the greater contact found in the lane/blocks when defending heavier forwards and centers backing down to the basket and to battle for rebounds.

Because of the ability to jump, athletes from this cluster will be better able to contest shots and protect the rim as well as contest shots on the perimeter.

Most important for the hybrid wing position, athletes from this cluster exhibit a superior ability to move change location based on the velocities and potential energies generated when jumping. Both measures indicate explosive power and skill when moving their centers of mass.

Cluster 2 had the second fastest three quarter sprint but generated the greatest amount of force (Figure 5 & 6).

A diagram of a box diagram

Description automatically generated

Figure 6

Cluster 2 also tied for second fastest average lane agility time (Figure 5), but this result is likely due to moving a greater body weight (102.58 kg) than the fastest (Cluster 0, 87.06 kg) and the cluster, Cluster 2 tied with (Cluster 3, 87.24kg) (Figures 5 & 7).

A diagram of a box diagram

Description automatically generated

Figure 7

Using a radar chart, we can look at the non-anthropometric measures in normalized form (Figure 8). It is evident that Cluster 2 (green), exhibits superior performances in multiple measures as described above in the tables and boxplots. As an example, the countermovement jump potential energy (cmj\_potential\_energy\_(J)) is much greater than the second best cluster, Cluster 4 (Figure 8).

A diagram of a star

Description automatically generated with medium confidence

Figure 8

**Top 5 Athletes of Cluster 2 – Top 5 Hybrid Wing Draft Prospects**

The Gaussian Mixture Model algorithm provided the probabilities that an observation fits within any of the clusters. For this analysis, the top five athletes from cluster 2 exhibited a greater than 99.0% probability of belonging to Cluster 2, which means there is a very small chance that the athlete is mis-labeled (Figure 9).

A table with numbers and text

Description automatically generated

Figure 9

**Top 5 Athletes of Cluster 2 – Anthropometric Data**

As was the case with Cluster 2 vs the other clusters, the top five athletes did not generally possess superior anthropometric characteristics relative to the average for Cluster 2. More measures were inferior (red filled cell) to the cluster average than were superior (green filled cell) (Figure 10). Weight and body fat were not measured against the cluster average as those values may either be inferior or superior relative to the cluster average, but having more mass would likely be beneficial when defending the lane/blocks.

Three of the five top 5 athletes exhibit the greater Cluster 2 trend of Wingspan to Height w/o Shoes Ratio. These athletes’ bodies meet the length requirement of the hybrid wing. Compared to their height, their wingspans are extremely long. This implies a capability to guard a multitude of positions through contesting passes, shots, rim protection and rebounding.

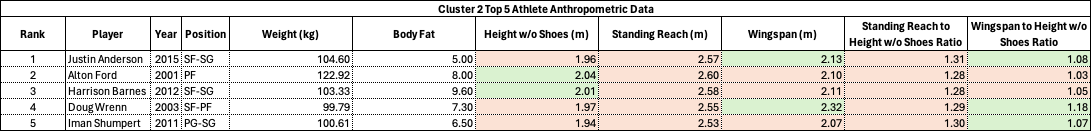


Figure 10

**Top 5 Athletes of Cluster 2 – Strength & Agility Data**

The Top 5 athletes of Cluster 2 exhibit superior strength and agility relative to the rest of the cluster. Nearly all measures are superior to the Cluster 2 average (green filled cells). Where an athlete’s measure is inferior relative to the Cluster 2 average (red filled cell), the athlete is marginally inferior (Figure 11).

A screenshot of a table

Description automatically generated

Figure 11

All but one athlete exhibits superior upper body strength as measured by the bench press implying a strength that would assist in defending against forwards and centers operating in the lane/blocks (Figure 11).

All five athletes exhibit:

* Tremendous potential energies in their jumping tests, which implies a high degree of muscle strength
* Greater countermovement jump heights than the cluster average

All but one athlete exhibited superior:

* Three quarter sprint times, velocities, and accelerations
* Max vertical jump heights and velocities

**Conclusion**

Using machine learning algorithms to cluster athlete anthropometric and strength & agility data provided an analysis free from analyst bias; focusing solely on objective measures and clustering athletes together based on similar characteristics. Taken together, the five athletes identified from Cluster 2 possess the requisite length and speed & agility capabilities to satisfy the hybrid wing requirements. These athletes’ anthropometric measures and strength & agility metrics indicate a capacity to guard multiple positions; on the perimeter where smaller, agile guards operate and, in the lane/blocks where larger, more powerful forwards and centers operate.

**References**

1. “Defining and Applying Load, Force, Velocity, Acceleration, and Power to VBT”, SIMPLIFASTER, <https://simplifaster.com/articles/new-applications-velocity-based-training/>.
2. “NBA Advanced Stats”, [*https://www.nba.com/stats/draft/combine-strength-agility*](https://www.nba.com/stats/draft/combine-strength-agility)