ABSTRACT

Each year, thousands of players from all around the world submit their name to the National Basketball Association (NBA), in hopes that their talent warrants an invitation to the upcoming draft. The draft is a process that enables teams to select the best talent to add to their roster. One of the major obstacles for teams in this process is trying to weed out the bad players, so to speak, from the good. Traditional selection methods have no real empirical backing and currently possess no full proof measure of ensuring that the players chosen will perform well in the league. The purpose of this research paper is to develop models that can assist scouts in selecting the optimal players and prevent the selection of players who are more prone to "bust" or exhibit poor performance. It illustrate how pre-draft characteristics such as an individual's height, hand size, weight, wingspan, and other attributes serve as a predictor for their performance.

INTRODUCTION

Determining a potential player's worth or benefit to the team is a pivotal decision across all sports that can truly determine an organisation's success or failure for years to come. Currently, in the National Basketball Association, when trying to determine which players are better than others, teams use a rather subjective process entailing scouts sent out to watch individuals perform, and then based on certain metrics predetermined by the team gauge how much added benefit the player can give to their firm. While players are still subject to participating in a draft, where height, weight, hand size and other physical ability tests are recorded, there is no real methodology in place where data driven insights can aide this process. Instead, these measurements are just tertiary data to aid the eyeball test from the team scouts. The data was collected from https://data.world/achou/nba-draft-combine-measurements. The data consists of information regarding their

Measurements for NBA draft combine participants from DraftExpress.com:

player, year ,draft pick, body fat, height with no shoes, height with shoes, wingspan, standing reach, vertical max, vertical max reach, vertical no step, vertical no step reach, hand length, hand width, weight, agility , bench and sprint . Each year, the NBA hosts a draft where the best basketball talent from all around the world has the opportunity to submit their name into the metaphorical hat in hopes that one of the thirty teams will value them high enough for selection. The way the draft works is teams are awarded a pick 1-30 based upon their prior year's performance, then these teams each take turns selecting the best player (or fit) left on the board to be the newest member of their squad. Once each of the thirty picks have been made, the second round commences, after all sixty picks have been performed, then the draft ceases until the following year. Due to the fact that each team has only two picks, just one error allows a team to be left with essentially a multi-million dollar deficit which can haunt them for years to follow, thus the need for good data driven insight in this selection method is imminent.

The NBA also has gone into analytics in a big way. Almost every team now has an NBA analytics department in the front office. Data is collected using cameras that record every movement of both the ball and all 10 players 25 times per second.

At an analytics conference this year, NBA commissioner Adam Silver said teams now have players wear monitor not only during games but during practice to measure, in part, performance and fatigue. They even have saliva sampled as it contains indicators of fatigue. Teams track and quantify a player's diet.

All of this works toward having a better, healthier player and a better basketball team.

While NBA teams have some flexibility in filling key roles with free agents, much of a team is built through the draft. A wrong choice in the first round can set a team back years. Teams look for whatever advantage they can when evaluating picks, and analytics play a key role.

Nothing shows fans what analytics have brought to basketball more than the explosion in attempts for three-point shots. In 2012, teams averaged about 18 three-point attempts per game. In 2017, that number reached 27.

Lebron James is known as one of the fastest players in the NBA, but he also worked very hard to train that skill. Agility is **the ability to change direction quickly while maintaining body control**. Kyrie Irving is a good example of a player with high-level agility when he changes directions while ball handling.

Running speed is very important for basketball players, particularly running up and down the court. The test involves running a maximum sprint over 3/4 of a basketball court (**75 feet or 22.86 meters**). This test is used in the Basketball SPARQ testing and at the NBA combine.

Speed and agility training is crucial for basketball players to **improve footwork skills** as well as improve cardio-respiratory stamina. Speed and agility training is also key in decreasing injury for basketball players.

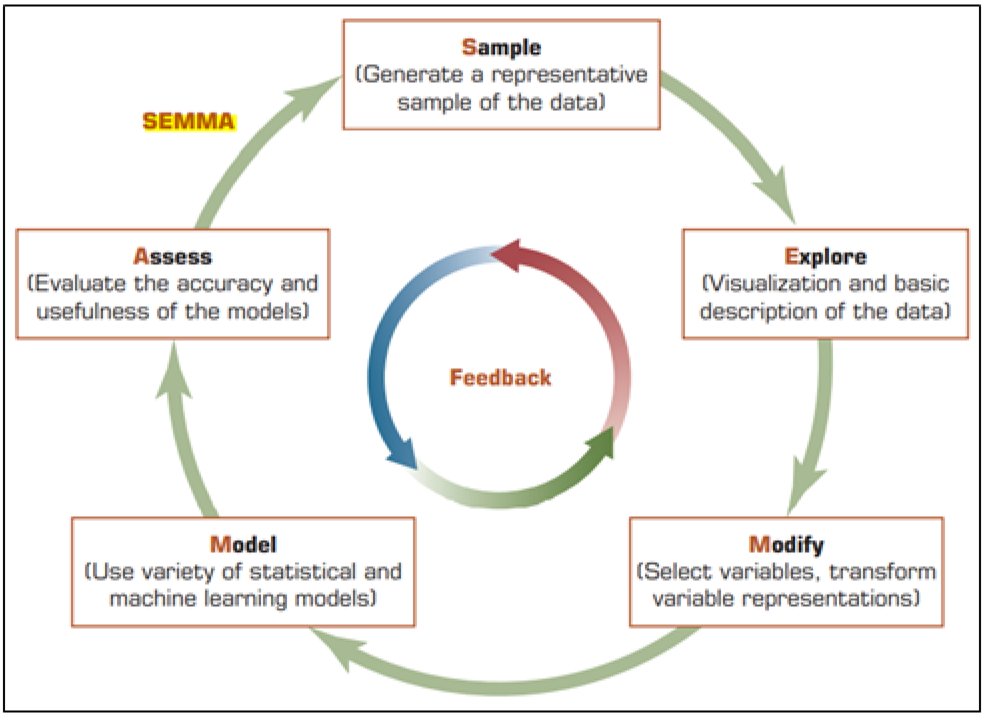
METHODOLOGY

The NBA Combine dataset consisted of eighteen input variables, and one(player)which served as the target variable. Information in this dataset regards personal characteristics which were measured during their Combine day. These attributes are all taken by professionals, so the accuracy of their reflection upon each individual is extremely high. For a complete record of information concerning this dataset, reference table below:-

Variable	Variable Description
players	Name
year	Year
Draft pick	North American The right of a sports team to select a player during the annual selection process.
Hand Length	The length of players hand from bottom of palm to middle finger in inches
Sprint	The time in seconds recorded during the sprint drill
Body Fat	The level of body fat recorded during players medical
Hand Width	The width in inches of players hand
Height with shoes	Individuals height with their basketball shoes on
Height noShoes	Individuals height in their socks
Weight (lbs.)	The amount each individual weighs
Agility	The time in seconds recorded during the lane agility drill
Vertical max	The height a player jumped from a standing position
Vertical max reach	The height recorded allowing the player to jump in any fashion they see fit
Bench	The number of times each player benched 135lbs
Wingspan	The length in inches measured from a players fingertip on one hand to the fingertip on the other
Standing reach	The distance in inches a player jumped from a standing position
Vertical no step	The height recorded-allowing the player to jump

Vertical no step reach	The height a player jumped from a standing position

Using this dataset we cannot-create any model.



I utilised several exploratory tools such as box plots, histograms, scatter plots and odds ratio to clearly illustrate the relationship between the target and my input variables. Initially, I found wingspan and a few other variables to have a relatively high association with the target variable, bringing forth a multicollinearity.I have created one column for BMI of the player in the data and also done year wise analysis of the player. Created a new data frame in which I have added a column for stamina, reach and mobility .I have also created profile report of data. I have also shown top ten players .In power bi I have also created visuals. I have analysed year-wise comparison
And key metrics and factors and show the meaningful relationships between attributes.

CONCLUSION

The results clearly illustrate that these models can accurately predict a performance based upon several factors and interactions amongst those factors from the NBA Combine. Several unique predictive models were comprised for this prediction and the Champion Model was determined according to validation mean squared error. Once the specifications from both Champions were run on the entire dataset, according to average squared error, the model that performed the best is the decision tree. The performance of these models were not as high as I would have liked, so that leaves room for fine tuning/training in the future. Nonetheless, these models are still more reliable compared to the baseline and can still aid executives in making data driven decisions during the draft. Both the regression and decision tree provide a baseline metric of a player's probable first year performance, which can be used as a direct comparison between multiple individuals. I suggest using the performance metric as a valuation factor in addition to the information gathered by the teams' scouts to determine a players' overall impact during their first season. This study is-comprised as a target variable and Combine performance/characteristics as inputs for the NBA. That being said, this research, in its current state does present a few shortcomings. A player's contribution cannot be measured by this metric alone, as certain teams require players to take on more specific roles, and there are other factors such as defensive presence which certainly have an influence but are not recorded by this formula. The future scope of this research includes fine-tuning these models to predict more specific characteristics of player.

REFERENCES

NBA Combine Statistics, Available at stats.nba.com/draft/combine.
https://data.world/achou/nba-draft-combine-measurements