Data mining is used to take a wide variety of information and compile it into one large list. It can be used for various things, an example would be to increase a company's profits, decrease expenses, or do both. The technology that has been used for this extensive research is not necessarily a new finding, but the name for this research is a relatively new term. This type of technology analyzes what it is given, then finds the correlations between them and sorts them into categories. But can these correlations and compilations predict a future pattern or statistic based on the information that is given, especially in the cultural phenomenon of sports? The potential lists that can be compiled uses the broader topics of data, knowledge, and information.

The broad topics of data, knowledge, and information are very different. Something that are considered to be data can be any facts, numbers, or text that can be computed and processed. Information is considered to be any patterns, associations, or relationships from any bit of data given. Knowledge can be described as a wide variety of things outside data mining, but very specific when one is data mining. Knowledge is computed from the information that is given, predicting the future trends based on historical knowledge. These are all necessary to get the final list.

An example of a very broad, but extensive list was observed briefly in class. The piece observed was called "A Day in the Life of Americans." This DH research datamined the typical day for an average person in America, looking at where they were on any given day. Using the process of data mining through the data, information, and knowledge the system was presented, this list allows the reader to make a statistical analysis of a certain fact. Due to the information given, the author could make the

assumption that 40 percent of people aged 25-34 were at work around three PM on an average day.

When analyzing what someone was doing at a given time, this showed how there is a potential to map out the day in the life of an average American from the time they wake up to the time they go to sleep. If that is an ultimate possibility, could there be data mining to break down specifics in the world of sports?

There are various amounts of data and different statistical computations when considering the cultural phenomenon of sports. Every time sports fanatics turn on ESPN after a big game like the Super Bowl or March Madness, the viewer is almost expecting to see some sort of statistical output of the game. For example: during the 2016 NCAA Men's March Madness Tournament, Michigan State was picked in 22% of the brackets to win the whole tournament and the Kansas University was picked in 25% of the brackets to win the Tournament. Michigan State dropped the ball though, losing in the round of 64. Kansas made it much further, making it to the Elite Eight. Statistically, it was supposed to be a Michigan State and Kansas National Championship. But in reality, the championship was played by Villanova and North Carolina. Where Villanova beat UNC with a crazy ending in regulation. With that in mind, the usage of data mining could be used here to predict what happened to these two favored teams and statistically break down why they lost.

The usage of data mining in sports is beneficial to the team using it. It judges the overall performance of the team and all of their players. The teams that use these strategies are considered to be a more advanced team.

Where the first use of extensive statistics was applied is completely unclear in the sports world, one of the most notable primary uses of this came in 2002. The sport that it was applied to is one of this nation's biggest cultural phenomenons in America's Past-time. Billie Beane was (and still is) the general manager of the major league baseball team named the Oakland Athletics, and his team was struggling to string together a series of wins. The A's had the third lowest bankroll in the MLB at $44 million, competing with the New York Yankees (who's bankroll was $125 million). He went into Cleveland to talk to General Manager of the Indians, Mark Shapiro, to make a trade or two. A young man by the name of Peter Brand sat in on the meeting between the two GMs, advising Shapiro if he should accept the trade or not. When Beane left the office of the Cleveland Indians, frustrated as could be, he asked Mr. Brand what he was saying during the meeting. Mr. Brand then said that it was a compilation of statistics pertaining to on base average, slugging percent, fielding percentage and other important hitting and fielding statistics. Beane was enamored in what Mr. Brand told him, so he made Shapiro an offer to buy Mr. Brand from the Cleveland Indians. The offer was accepted, and Beane told him to pack his bags because he was coming to Oakland. By the end of the following season, they made the Divisional Series against the Boston Red Sox while still having one of the lowest bankrolls in the MLB. The A's unfortunately lost in the Divisional round, but this type of baseball earned a nickname. This term used to describe the team and the possible value of players based on statistics is called "Moneyball."

The science that was used by Brand was invented in 1980 by Bill James, it was deemed to be called sabermetrics. The original definition according to the Society For American Baseball Research as the search for objective knowledge about baseball. But Jim Albert of Bowling Green State University broke down the definition further as: the mathematical and statistical analysis of baseball records. Since the revolutionary discovery of sabermetrics, other professional baseball teams have hired sabermetric specialists to analyze their own players as well as the potential prospects the team is looking to recruit.

When looking at the data mining in the sports world, it doesn't just pertain to sabermetrics. Different scouts use a form of data mining to compile information on the prospects that they are looking at. Scouts look at different strengths and weaknesses that vary from the stats at hand, pattern discovery, to even how often they may get hurt (by comparing workouts and recovery). If sabermetrics was around when Billie Bene was drafted/scouted, he wouldn't have ever made it into the MLB. He was a teen prodigy that didn't even come close to making a name for himself in the big leagues. If other sports (such as football) used processes similar to sabermetrics and applying the probability they will be a bust or not, people like Josh Gordon or Johnny Manziel wouldn't have been drafted. To specify on Gordon and Manziel, their off field antics resulted in suspensions, then later cuts, from the Cleveland Browns (and the NFL, in Gordon's case).

Coming back to the MSU and KU scenarios, they needed to heavily rely on their 'clutch players' to make plays happen. Osama K. Solieman defined a clutch player as:

“The period in the game, specifically the fourth quarter and overtime, in which the game is in the balance and the teams are close in score is referred to as the clutch. The legends of the NBA often have cemented their legacies by performing best in these situations. For NBA organizations, players who can make plays in the clutch have additional value because more often than not effective performance in the final minutes of the game is the difference between winning and losing.” When people look at the statistics from the game, their ‘clutch’ or ‘key’ players weren’t performing at the level that the team needed them to. This is an example of a possible factor of when these players would be getting drafted in the NBA draft the following year, since nobody wants a ‘star’ to choke in a big game. For example, Denzel Valentine is considered to be MSU’s best player (or clutch player). He won Coach’s Player of the year among holding a couple records at the school. But in the big game of the NCAA tournament, with all the hype on him and his talent, he went 5 of 13 for field goals (shots that are worth two points), and 3 of 8 behind the three-point line. MSU went on to lose by nine points, losing 90-81.

With the strides that have been made using the research technique of data mining, this subject has become a major contribution to the field of digital humanities and sports. This technique allows DHers, companies, sports teams/organizations, or just average people to look at/compile long lists of things they may find beneficial for reasons that are personal to themselves. Teams and organizations might be able to predict the future investment of a player, people can make an estimate for what players can be good for fantasy sports teams, and DHers can use data mining for other information outside the world of sports.

Works Cited

Albert, Jim. “An Introduction to Sabermetrics.”

ESPN. “Middle Tennessee vs. Michigan State: Box Score.”

UCLA. “Data Mining: What is Data Mining?”

Society for American Baseball Research. “A Guide to Sabermetrics Research.”

Solieman, Osama K. “Data Mining in Sports: A Research Overview.”