

Could AI help decide Baseball Hall of Fame inductees?

Chicago Tribune · 20 Jan 2023 · By Sheldon Jacobson Sheldon Jacobson is a professor of computer science at the University of Illinois at Urbana-Champaign. He employs his expertise in data-driven, risk-based decision-making to evaluate and inform public policy.

On Tuesday the new members of the Baseball Hall of Fame will be announced, and the debate will ensue as to who does or does not belong, who has been unjustifiably snubbed and how banned substances should be factored into the decision.



Around 400 sports writers determine who will be enshrined among baseball's all-time greatest players in Cooperstown, New York, on July 23. The big question is: Should we allow these decisions to be left solely to sports writers and their subjective process?

Further, has getting voted into the Hall become as much about acceptable off-field behavior and judgment as exceptional on-field performance?

Clearly, on-field performance is necessary. However, based on the list of players with outstanding statistics who are not in line to get voted in anytime soon, this is no longer sufficient.

The use of performance-enhancing drugs clouded the statistics. Eclipsing 500 home runs is no longer a guarantee of getting into the Hall, with 28 players in this elite club. The same can be said with surpassing 3,000 hits, with just 33 players in this group.

But what about poor behavior and judgment that do not affect statistics?

There are few saints among us. Baseball players are human. Some of their off-field behavior may tarnish their reputations, their character and their image. The question is whether such behavior should be sufficient to overshadow what may have been an outstanding professional career.

This oft-discussed issue is most pertinent when it comes to Pete Rose, the all-time leader with 4,256 hits, a record that will likely never be broken. His gambling escapades clearly crossed many lines that led to his lifetime ban from baseball in 1989. Yet placing bets, even on baseball and the team he played for and managed, did not alter his statistics in the field. Would he have had fewer hits if he had not gambled?

He claimed that when he gambled on his own team, he always gambled for them to win, though there is some evidence that suggests otherwise.

Given the proliferation of sports gambling today, with the ability to gamble on every play during every game, Rose's gambling by today's standard would be considered rather pedestrian. It is hard to believe that all professional athletes are not gambling on sports, including their own sport. Whether they are doing it to the extent that Rose did is unclear.

If induction into the Hall requires exceptional on-field performance and immaculate off-field behavior and judgment, then the number of new inductees may slither down to a trickle.

These points are not meant to condone bad behavior and certainly not to reward it. They are designed to highlight that the criteria for induction into the Hall is far more nuanced than most would like to acknowledge.

Baseball statistics are the best indicator of on-field performance. They are objective, indisputable and unforgiving. A hit is a hit, a strikeout is a strikeout.

There is a certain amount of luck that occurs on any given at-bat or in any given game. However, over a full season, or more importantly a full career, such luck evens out. No one can be a career .320 hitter by chance, nor can a pitcher win 250 games with luck alone.

Behavior and judgment are far more subtle and subject to interpretation. Would Babe Ruth, known for his off-field antics and raucous carousing, be voted into the Baseball Hall of Fame today?

At the same time, when induction into an institution such as the Hall is determined by people — namely, baseball sports writers — and their perspectives, not solely on the statistics that support exceptional performance on the field, then a more specific set of criteria delineating all relevant factors is needed.

A group of baseball pundits could assemble a set of performance criteria that would define exactly what must be achieved by every type of player to gain entry in the Hall. Then every player's set of statistics could be weighed against such criteria. If off-field behavior and judgment are to be included, this could also affect a player's chances. With baseball now a sport driven by data analytics, perhaps the shift from sports writers to artificial intelligence is the direction needed to make selections. This would take into account both on-field and off-field factors so that a decision could be made that captures the spirit of what a Hall of Fame-worthy career should be.

As a compromise, AI algorithms could produce a more refined set of candidates, with baseball writers giving the final approval. Would such an approach lead to different selections than what baseball writers would arrive at on their own? Would Barry Bonds or Roger Clemens get a more favorable opinion from an AI selection algorithm than baseball writers?

Every Major League Baseball player hopes that one day he will be inducted into the Hall. If analytics drive how the game is played today, perhaps analytics should also be used to determine if players are worthy of such a distinction.