Predicting Team Advancement in Major League Baseball Postseason Using Borda Count

CHIH-CHENG CHEN[1], TIAN-SHAING KUO[2], KUANG-TSAN HUNG[3], JUNG-CHANG WANG[4], MING-YAO CHEN[5], CHUNG-YU TSAI[1]

[1]Department of Sport Management
Aletheia University
32, Zhen Li St., Tamsui Dist., New Taipei City
TAIWAN

[2]Department of Regimen and Leisure Management
Tainan University of Technology
529, Zhongzheng Rd., Yongkang Dist., Tainan City
TAIWAN

[3]Department of Event Management
Hungkuo DeLin University of Technology
1, Ln. 380, Qigyun Rd., Tucheng Dist, New Taipei City
TAIWAN

[4]Department of Tourism
Aletheia University
32, Zhen Li St., Tamsui Dist., New Taipei City
TAIWAN

[5]Department of Sports Information & Communication
Aletheia University
32, Zhen Li St., Tamsui Dist., New Taipei City
TAIWAN

Abstract: - The prediction of sports competition outcomes has long been a topic of interest in academia and among sports enthusiasts. This study focuses on Major League Baseball (MLB) as its research subject, encompassing the years 2020, 2021, and 2022. By employing a set of established evaluation criteria, comprising five pitching and five hitting indicators from previous literature, the regular-season performance of the 30 MLB teams across both leagues (National League and American League) over the three-year period was compiled. Subsequently, a data normalization technique combined with the Borda count concept was proposed to develop a model for forecasting team advancement in the postseason. The predictive accuracy of the model presented in this study for determining MLB postseason qualifiers from 2020 to 2022 fell within the range of 55.6% to 66.7%, akin to models utilizing extensive datasets. Notably, the proposed model is more comprehensible and user-friendly, offering ease of understanding and application for sports enthusiasts and facilitating its potential utilization and dissemination in the sporting community.

Key-Words: - short-term sporting events, predicting models, normalization

Received: April 15, 2024. Revised: September 11, 2024. Accepted: October 13, 2024. Published: November 5, 2024.

1 Introduction

Predicting the outcomes of sporting events is a highly sought-after topic among sports enthusiasts. However, typical sports fans can only access team or player performance information from official websites of major sporting events. They rely on intuition or win-loss records to assess the outcome of the next match. Research related to predicting sports

event outcomes can be broadly categorized into two main types. One category focuses on predicting individual performance of professional sports teams (Bailey, Loeppky, & Swartz, 2020), while the other category involves predicting team performance (Huang & Li, 2021; Jia, Wong, & Zeng, 2013). However, most studies in this field rely on extensive data for prediction. For example, Bailey, Loeppky, and Swartz (2020) utilized logistic regression analysis to predict the batting average of hitters in the Major League Baseball (MLB), using data from over 30 teams and more than 162 games per season (over 2,400 games), which generated over 30 batting indicators per game. On the other hand, Jia et al. (2013) collected over 2 million records of MLB pitching and hitting data from 2007 to 2011 using artificial intelligence (AI) and machine learning techniques, aiming to predict game results for the 2012 season. Huang and Li (2021) also employed deep learning techniques, a form of AI, to predict MLB game outcomes based on information gathered from a total of 4,858 games in the 2019 season. However, many sporting events do not have as many matches or accumulate a large volume of game data, making it less convenient to analyze them using big data methods. For instance, the MLB postseason, which garners significant attention from baseball enthusiasts worldwide, lasts approximately one month and consists of 28 to 45 matches, which is significantly fewer compared to more then 2,000 matches played in a single regular season of MLB.
To address the limitation of limited data, this study will employ Multi-Criteria Decision Making (MCDM) methods, commonly utilized for evaluation, ranking, and selection across various domains, to predict the outcomes of short-term sporting events. MCDM methods have been frequently applied in research related to sports (Chen, Lee, & Tsai, 2014; Pradhan, 2018), including prediction studies (Romero, Lozano-Murcia, Lopez-Gomez, Angulo Sanchez-Herrera, & Sanchez-Lopez, 2021; Saqlain, Jafar, Hamid, & Shahzad, 2019). The Borda count (a kind of MCDM method) is dvoting method employed to rank or rate options. In the context of predicting sports event outcomes, it can be applied in various domains such as ranking teams or athletes and forecasting match rankings (Dalton-Barron et al., 2022; Kaiser, 2019; Nutting, 2016).

Drawing upon the aforementioned information, the fundamental aim of this research is to utilize the Borda count methodology. Through a comprehensive analysis of MLB teams' performance throughout the regular season, the study endeavors to predict the likelihood of these teams progressing to the postseason. Moreover, it strives to evaluate their prospects for triumph in the playoffs by scrutinizing their matchups against their respective opponents.

2 Methods

2.1 Data

The study to collect the MLB 2020 to 2022 statistics data of regular season by the MLB official website (mlb.com) which provides game stats. In the data applied, the technical variables correspond to values of the teams, not to the individual values of the players. All the ten criteria (technical variables) were used (Chen et al., 2014; Jia et al., 2013) as criteria including earned run average (ERA), walks and hits per inning pitched (WHIP), batting average against (AVG), strkeouts per 9 IP (SO/9), walks per 9 IP (BB/9), runs batted in (RBI), batting average (HAVG), on-base percentage (OBP), slugging percentage (SLG), at bats per home run (AB/HR). Before using Borda Score, we used regression analysis in which Wins were the dependent variable, and ERA, WHIP, AVG, SO/9, BB/9, RBI, HAVG, OBP, SLG and AB/HR were independent variables. The analysis results showed that the R-squared values were 0.974. It means that the technical variables cited in this study have good explanatory power for Wins.

2.2 Board count

The Borda count is a voting method employed to rank or rate options. In the context of predicting sports event outcomes, it can be applied in various domains such as ranking teams or athletes and forecasting match rankings (Dalton-Barron et al., 2022; Kaiser, 2019; Nutting, 2016). In order to analyse the data, we conducted normalization proceduce and Borda count to determine each MLB team’s performance. There are multistep procedure that provides a comprehensive ranking system for sets of data. Steps for Board count as follows:
Step 1. criteria normalization
There are three different methods of normalizing the sequences including benefit type, defect type, and target type methods (Chen et al., 2022). In this study, the benefit and defect method are used.
The benefit type (SO/9, RBI, HAVG, OBP and SLG) method indicates that the larger target value is better. The calculation is as follows:

$$x_{ij}^{*} = \frac{x_{ij} - min_i x_{ij}}{max_i x_{ij} - min_i x_{ij}} \tag{1}$$

Where $max_i x_{ij}$ is the largest value in criteria j and $min_i x_{ij}$ is the smallest value in criteria j.

The defect type (ERA, WHIP, AVG, BB/9, AB/HR) method indicates that the smaller target value is better. The calculation is as follows:

$$x_{ij}^* = \frac{max_i x_{ij} - x_{ij}}{max_i x_{ij} - min_i x_{ij}} \quad (2)$$

Where $max_i x_{ij}$ is the largest value in criteria j and $min_i x_{ij}$ is the smallest value in criteria j.

Step 2 Borda count calculation

By applying normalization techniques, the performance data of each team across ten criteria is transformed into a range of 0 to 1. This normalized data is then multiplied by 30 (as there are 30 teams in MLB). Finally, the sum of the performance scores for each criterion is calculated for each team. The higher Borda count is better. The calculation is as follows:

$$Borda\ count\ of\ Team_{ij} = \sum_{j=1}^{10} x_{ij}^* \times 30 \quad (3)$$

3 Results

3.1 Performance of MLB Teams for Advancement in the Postseason

Since 2012, the MLB postseason consists of eight teams, comprising three division champions from each league and two wild card teams. The MLB has organized the postseason into three rounds of competition: the Division Series, League Championship Series, and World Series. The Division Series employs a "best of 5" format, while the subsequent rounds utilize a "best of 7" format. During the Division Series, the team with the best record in a league faces off against the wild card team. In cases where both teams belong to the same division within the league, the division champion with the second-best record competes against the wild card team. Notably, the team with the superior regular-season record enjoys the home-field advantage during the series, while the wild card team is not granted such a benefit, except during the World Series.

In 2022, MLB increased the postseason excitement by adding three wild card teams to each league, making a total of 12 teams in the playoffs. This transformed the postseason format from three rounds to four, including the new Wild Card Series. In this "best of 3" format, three division champions with weaker regular-season records and the wild card team compete for a spot in the Division Series. The winner then faces the top two divisional teams. Division champions have the home-field advantage, and the winner advances to the League Championship Series. In this study, we used the data from MLB regular seasons 2020 to 2022. By employing equations (1) to (3), the Borda count for each team in the current year is calculated. A higher Borda count value signifies a stronger team with a higher probability of winning during matches.

3.1.1 2020 MLB postseason

In 2020, due to the impact of Covid-19, the regular season games were reduced to 60. Additionally, the qualifying rules for the postseason were modified. In the 2020 playoffs, each league had 8 teams advancing to the postseason, comprising the top two teams from each of the three divisions (6 teams in total) and 2 wild card teams. This resulted in a total of 16 teams across both leagues, competing in a four-round format.

Among the teams that qualified for the playoffs in 2020, the top three performers in the National League (NL) were the Los Angeles Dodgers, San Diego Padres, and Atlanta Braves, as illustrated in Table 1. Meanwhile, in the American League (AL), the top three teams were the New York Yankees, Chicago White Sox, and Minnesota Twins.

Table 1.
Borda count and ranking of MLB teams in the 2020 postseason

TEAM	League	Borda count	Ranking
New York Yankees (X)	AL	207	4
Chicago White Sox (W)	AL	200	5
Minnesota Twins (X)	AL	189	6
Tampa Bay Rays (X^)	AL	188	7
Cleveland Indians (Y)	AL	168	8
Oakland Athletics (X)	AL	158	10
Toronto Blue Jays (W)	AL	147	13
Houston Astros (Y)	AL	139	14
Los Angeles Dodgers (X^)	NL	268	1
San Diego Padres (Y)	NL	226	2
Atlanta Braves (X)	NL	208	3
Cincinnati Reds (W)	NL	165	9
Milwaukee Brewers (W)	NL	154	11
Chicago Cubs (X)	NL	148	12
St. Louis Cardinals (Y)	NL	130	15
Miami Marlins (Y)	NL	92	16

Note: X is Division Champion, Y is Division second place, W is Wild Card, ^ is most wins of league, AL is American League, NL is National League

3.1.2 2021 MLB postseason

In 2021, MLB reverted to its original playoff format, where apart from the divisional champions, the two best-performing teams in each league also qualified for the postseason as wild card teams. As a result, a total of five teams from each league secured playoff berths. According to the data presented in Table 2, the two top-performing teams in the National League were the Los Angeles Dodgers and San Francisco Giants. In the American League, the best-performing teams were the Houston Astros and Chicago White Sox.

Table 2.
Borda count and ranking of MLB teams in the 2021 postseason

TEAM	League	Borda count	Ranking
Houston Astros (X)	AL	227	3
Chicago White Sox (X)	AL	221	4
Tampa Bay Rays (X^)	AL	212	5
Boston Red Sox (W)	AL	187	7
New York Yankees (W)	AL	183	8
Los Angeles Dodgers (W)	NL	246	1
San Francisco Giants (X^)	NL	229	2
Atlanta Braves (X)	NL	188	6
Milwaukee Brewers (X)	NL	181	9
St. Louis Cardinals (W)	NL	127	10

Note: X is Division Champion, W is Wild Card, ^ is most wins of league, AL is American League, NL is National League

3.1.3 2022 MLB postseason

Starting from 2022, MLB made further modifications to the playoff advancement regulations. The number of wild card teams was increased from two to three in each league, resulting in an additional playoff berth for each league. The playoff format was also altered, shifting from the divisional champions directly advancing to the division series to the divisional champion with the worst record beginning their postseason journey from the wild card stage. Based on the data provided in Table 3, the two top-performing teams were the Los Angeles Dodgers in the National League and the Houston Astros in the American League. The third and fourth positions were secured by the New York Yankees (American League) and the Atlanta Braves (National League), respectively.

Table 3.
Borda count and ranking of MLB teams in the 2022 postseason

TEAM	League	Borda count	Ranking
Houston Astros (X^)	AL	243	2
New York Yankees (X)	AL	238	3
Toronto Blue Jays (W)	AL	217	6
Cleveland Guardians (X)	AL	178	8
Seattle Mariners (W)	AL	175	10
Tampa Bay Rays (W)	AL	172	12
Los Angeles Dodgers (X^)	NL	279	1
Atlanta Braves (X)	NL	233	4
New York Mets (W)	NL	229	5
Philadelphia Phillies (W)	NL	193	7
St. Louis Cardinals (X)	NL	175	9
San Diego Padres (W)	NL	174	11

Note: X is Division Champion, W is Wild Card, ^ is most wins of league, AL is American League, NL is National League

3.2Predicting Accuracy in the MLB Postseason

In this study, we utilize the Borda count of teams that entered the playoffs from 2020 to 2022 as a criterion for assessing their advancement during the respective years. The results are presented in the following section.

3.2.1 Predicting Accuracy of 2020 MLB postseason

According to the data presented in Table 4, the Borda count proposed in this study correctly predicted the winning team in five out of eight matchups during the 2020 MLB Wild Card stage. Moving on to the Divisional stage, it accurately predicted two out of four matchups. Subsequently, it correctly predicted the winning team in both the League Championship and World Series matchups. The team ranked first in the Borda count for that year, the Los Angeles Dodgers, went on to win the World Series championship. Overall, the Borda count model proposed in this research achieved a correct prediction rate of 66.7% across all 15 matchups analyzed.

Table 4.
2020 MLB Postseason Predictions

Series		Wild Card			Disision			League			World		
	BC	P	A	J	P	A	J	P	A	J	P	A	J
TB	188												
TOR	147	TB	TB	1									
CLE	168	NYY	NYY	1									
NYY	207				NYY	TB	0						
MIN	189				OAK	HOU	0						

HOU	139	MIN	HOU	0										
OAK	158	CWS	OAK	0										
CWS	200							TB	TB	1	LAD	LAD	1
LAD	268							LAD	LAD	1			
MIL	154	LAD	LAD	1									
SD	226	SD	SD	1									
STL	130				LAD	LAD	1						
CHC	148				ATL	ATL	1						
MIA	92	CHC	MIA	0									
ATL	208	ATL	ATL	1									
CIN	165												

Correct ratio is 66.7%

Note: BC is Borda count, P is predict, A is actual, J is Judge, 1 is correct, 0 is incorrect, TB is Tampa Bay Rays, TOR is Toronto Blue Jays, CLE is Cleveland Indians, NYY is New York Yankees, MIN is Minnesota Twins, HOU is Houston Astros, OAK is Oakland Athletics, CWS is Chicago White Sox, LAD is Los Angeles Dodgers, MIL is Milwaukee Brewers, SD is San Diego Padres, STL is St. Louis Cardinals, CHC is Chicago Cubs, MIA is Miami Marlins, ATL is Atlanta Braves, CIN is Cincinnati Reds.

3.2.2 Predicting Accuracy of 2021 MLB postseason

Table 5 presents the data on the Borda count of each team and the outcomes of the matchups during the 2021 MLB postseason. Utilizing the Borda count proposed in this study, we achieved correct predictions for the winning teams in both of the 2021 MLB Wild Card stage matchups. Moving on to the Divisional stage, we accurately predicted three out of four matchups. Subsequently, in the League Championship, we correctly predicted the winning team in one out of two matchups. Finally, in the World Series matchup, we made an accurate prediction for the winning team. The team with the highest Borda count in 2021, the Atlanta Braves, emerged as the World Series champions. Overall, the Borda count model proposed in this research achieved a correct prediction rate of 55.6% across all nine matchups analyzed in the 2021 MLB postseason.

Table 5.
2021 MLB Postseason Predictions

Series		Wild Card			Disision			League			World		
	BC	P	A	J	P	A	J	P	A	J	P	A	J
BOS	187												
NNY	183	BOS	BOS	1									
TB	212				TB	BOS	0						
HOU	227				HOU	HOU	1	HOU	HOU	1			
CWS	221												
LAD	246										HOU	ATL	0
STL	127	LAD	LAD	1									
SF	229				LAD	LAD	1	LAD	ATL	0			
MIL	181				ATL	ATL	1						
ATL	188												

Correct ratio is 55.6%

Note: BC is Borda count, P is predict, A is actual, J is Judge, 1 is correct, 0 is incorrect, BOS is Boston Red Sox, NYY is New York Yankees, TB is Tampa Bay Rays, HOU is Houston Astros, CWS is Chicago White Sox, LAD is Los Angeles Dodgers, SF is San Francisco Giant, MIL is Milwaukee Brewers, ATL is Atlanta Braves.

3.2.3 Predicting Accuracy of 2022 MLB postseason

According to the data presented in Table 6, the Borda count proposed in this study correctly predicted the winning team in two out of four matchups during the 2022 MLB Wild Card stage. Moving on to the Divisional stage, it also accurately predicted the winning team in two out of four matchups. Subsequently, in the League Championship and the World Series matchups, the Borda count correctly predicted the winning team in all three matchups. The team with the highest Borda count in 2022, the Houston Astros, emerged as the World Series champions. Overall, the Borda count model proposed in this research achieved a correct prediction rate of 63.6% across all eleven matchups analyzed in the 2022 MLB postseason.

Table 6.
2022 MLB Postseason Predictions

Series		Wild Card			Disision			League			World		
	BC	P	A	J	P	A	J	P	A	J	P	A	J
TOR	217												
SEA	175	TOR	SEA	0									
HOU	243				HOU	HOU	1						
CLE	178							HOU	HOU	1			
TB	172	CLE	CLE	1									
NYY	238				NYY	NYY	1						
NYM	229										HOU	HOU	1
SD	174	NYM	SD	0									
LAD	279				LAD	SD	0						
STL	175												
PHI	193	PHI	PHI	1				PHI	PHI	1			
ATL	233				ATL	PHI	0						

Correct ratio is 63.6%

Note: BC is Borda count, P is predict, A is actual, J is Judge, 1 is correct, 0 is incorrect, TOR is Toronto Blue Jays, SEA is Seattle Mariners, HOU is Houston Astros, CLE is Cleveland Guardians, TB is Tampa Bay Rays, NYY is New York Yankees, NYM is New York Mets, SD is San Diego Padres, LAD is Los Angeles Dodgers, STL is St. Louis Cardinals, PHI is Philadelphia Phillies, ATL is Atlanta Braves.

4 Conclusion

The accuracy of predictions has always been a topic of interest for many researchers and bettors in the field of sports management. The model proposed in this study solely relies on the regular-season performance of MLB teams that advanced to the postseason. By employing a straightforward criteria normalization technique in combination with Borda counts, the model calculates the winning probabilities for each team, thus enabling the prediction of the outcomes of various team matchups.

According to the results obtained in this research, the prediction accuracy for the postseason from 2020 to 2022 fell within the range of 55.6% to 66.7%. These results closely resemble the accuracy achieved by Huang and Li (2021) and Jia et al. (2013) in their utilization of extensive data and algorithms to predict game outcomes. For an extended period, predicting the results of baseball games has been a topic of great interest for many scholars and fans. They rely on various team performance indicators to forecast match outcomes, with many methods yielding prediction accuracies of approximately 50%. As a result, the method proposed in this study, which relies on limited data to predict baseball game results, demonstrates commendable accuracy and can be considered an effective approach.

Predicting game outcomes has always been a highly regarded and popular topic, but most studies tend to focus on longer events with numerous games, generating large amounts of data. However, short-term events also hold their appeal, such as professional league playoffs, FIFA World Cup, or the Baseball World Cup, which are significant international tournaments. Therefore, this research proposes a method that allows individuals to access data and employ computational techniques to predict the results of such events.

The results of this study demonstrate that the application of Borda count yields a good predictive capability for determining the winning ability of each participating team, thus offering valuable insights for predicting team advancements in various sporting events.

References:

[1] Bailey, S. R., Loeppky, J., & Swartz, T. B. (2020). The prediction of batting aver-ages in major league baseball. *Stats, 3*(2), 84-93. https://doi.org/10.3390/stats3020008

[2] Chen, C. C., Lee, Y. T. & Tsai, C. M. (2014). Professional Baseball Team Starting Pitcher Selection Using AHP and TOPSIS Methods. *International Journal of Performance Analysis in Sport, 14*(2), 545-563. https://doi.org/10.1080/24748668.2014.11868742

[3] Chen, C. C., Yang, T. C., Chiu, W. S., Wang, J. C., Chow, T. H., & Chen, M. Y. (2022). The Research of Key Indicators of Performance to Predict to Advance the Knock Out Stage in the International Football Tournament. *Mathematical Problems in Engineering, 2022, Article ID3396238*. https://doi.org/10.1155/2022/3396238

[4] Dalton-Barron, N., Palczewska, A., Weaving, D., Rennie, G., Beggs, C., Roe, G., & Jones, B. (2022). Clustering of match running and performance indicators to assess between-and within-playing position similarity in professional rugby league. *Journal of Sports Sciences, 40*(15), 1712-1721. https://doi.org/10.1080/02640414.2022.2100781

[5] Huang, M. L., & Li, Y. Z. (2021). Use of Machine Learning and Deep Learning to Predict the Outcomes of Major League Baseball Matches. *Applied Sciences, 11*(10), 4499. https://doi.org/10.3390/app11104499

[6] Jia, R., Wong, C., & Zeng, D. (2013). *Predicting the Major League Baseball Season*. CS229 Machine Learning Final Project, 1-5.

[7] Kaiser, B. (2019). Strategy and paradoxes of Borda count in Formula 1 racing. *Decyzje, 31*, 115-132.

[8] Nutting, A. W. (2016). Discrimination and information: Geographic bias in college basketball polls. *Eastern Economic Journal, 42*, 80-103. https://doi.org/10.1057/eej.2014.41

[9] Pradhan, S. (2018). Ranking regular seasons in the NBA's Modern Era using grey relational analysis. *Journal of Sports Analytics, 4*(1), 31-63. https://doi.org/10.3233/JSA-160165

[10] Romero, F. P., Lozano-Murcia, C., Lopez-Gomez, J. A., Angulo Sanchez-Herrera, E., & Sanchez-Lopez, E. (2021). A data-driven approach to predicting the most valuable player in a game. *Computational and Mathematical*

Methods, e1155. https://doi.org/10.1002/cmm4.1155

[11] Saqlain, M., Jafar, N., Hamid, R., & Shahzad, A. (2019). Prediction of Cricket World Cup 2019 by TOPSIS Technique of MCDM-A Mathematical Analysis. *International Journal of Scientific & Engineering Research, 10*(2), 789-792.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.