

# Wisdom of crowds in sports forecasting.

How social media crowds' sports forecasts hold against traditional methods?

Author: Javier Ocampo 470673

Supervisor: dr. Tong Wang

Second Assessor: Benjamin Tereick

Date Final Version: the 1st of August, 2019

**ERASMUS UNIVERSITY ROTTERDAM**  
Erasmus School of Economics  
MSc Economics and Business  
Master thesis Behavioural Economics

The views stated in this thesis are those of the author and not necessarily those of Erasmus School of Economics or Erasmus University Rotterdam.

## **Abstract**

Today, people are spending hour upon hour on various social media platforms. As a result of this, for the vast majority of people, social media is the primary source from where they get their information about current events, sports, celebrity gossip, etc. Given the significant role that social media plays in our everyday lives and the fact that our dependence on social media seems to be growing at a seemingly never-ending rate, it is logically then of great importance to study whether tools such as Instagram polls contain valuable information or it is all just noise.

The wisdom of crowds is based on the idea that good group judgement can be generated from the aggregation of individual judgments and this study aims to find out whether Instagram polls can be used as a reliable source for the wisdom of the crowd. Furthermore, it is hoped that the findings of this thesis will promote the use of these polls to forecast the outcome of sports events and further investigate the application of Instagram polls to forecast the success of new products, movies box office, tourism preferences, etc.

A database of Instagram polls of a sports event was generated to study its effectiveness in forecasting sports events. Tests were conducted to determine if the Instagram polls performed better forecasts than a guess chance. The preliminary results of the tests were found to be favourable for Instagram polls. In addition, tests were conducted determine if the surveys could provide better forecasts than TV specialists and other benchmarks, in this case, the hypothesis that the surveys were different from the experts' forecasts or other benchmarks could not be rejected. Although it could not be shown that the crowd through the Instagram surveys was better at forecasting than the benchmarks, it can not be said that the benchmarks are better than the crowd, which validates that the Instagram crowd can be as good source of wisdom as the benchmarks.

## Table of content

<b>1. Introduction</b> .....	4
<b>2. Literature Review</b> .....	5
2.1. Origins of WOC.....	5
2.2. Crowdsourcing. ....	5
2.3. Conditions required for WOC.....	7
2.4. The wisdom of crowds in sports. ....	8
<b>3. Main Question.</b> .....	9
<b>4. Methodology.</b> .....	10
4.1. Hypothesis.....	10
4.2. Instagram Polls. ....	11
4.3. Benchmarks.....	12
4.3.1. Experts predictions.....	12
4.3.2. FIFA/Coca-Cola World Ranking. ....	13
4.3.3. Instagram Media Followers.....	13
4.3.4. Betting Odds.....	14
4.3.5. SPI model. ....	14
4.3.6. CARMELO model. ....	15
<b>5. Results.</b> .....	15
5.1 Descriptive statistics.....	15
5.2. Main Results.....	17
5.3. Expertise Accuracy ROC - AUC.....	20
<b>6. Discussion.</b> .....	26
6.1 Results discussion.....	26
6.2. Limitations and Recommendations.....	29
<b>7. Conclusions.</b> .....	32
<b>8. Bibliography.</b> .....	33
<b>9. Appendix.</b> .....	36

## 1. Introduction

What if you want to make a bet on a football match, then you see a poll on Instagram and the team you want to bet on has very little support on the poll, is this information relevant? Should you change your mind? This is what we are going to find out.

The wisdom of crowds is based on the idea that good group judgement can be generated from the aggregation of individual judgments. Also, crowd judgment is more accurate than most individual judgments (Lee & Lee, 2017). This premise gives us to believe that if we should consider the result of the survey in our judgment.

But according to literature, there are four conditions to assure that the crowd is reliable. First, the crowd must be knowledgeable. Second, the subjects must have an incentive to be correct in their predictions. Third, the observations must be independent. Fourth, the crowd must be diverse (Simmons, Nelson, Galak, & Frederick, 2010).

The first condition is only partially achieved. If the poll is done via an Instagram account related to the event, then by in large the subjects are to some degree interested in the event. The second condition is not achieved because there is no incentive to promote the need to be right. The third condition is reached as subjects will individually complete the poll from their cell phone without knowing for which option the other participants voted for. Finally, the fourth condition is also reached as subjects will be from all around the globe.

The purpose of this study is to find out whether an Instagram poll is a good source for the wisdom of the crowd. Is this information better than a guess or chance? Is the judgment of the crowd better than say the judgment of an expert? How does this judgment compare to the other benchmarks?

The outline of the thesis is as follows. First, the thesis starts with a review of the existing research and literature on the topic. Then, chapter 3 presents the research question, chapter 4 presents the research hypothesis and the used mythology is described. In chapter 5, focus on the descriptive statistics and test results. Chapter 6 discusses the limitations, results and recommendations for future research. Finally, the conclusions are presented in chapter 7.

## **2. Literature Review.**

### 2.1. Origins of WOC.

In *Vox Populi*, Francis Galton for the first time tested the theory of "The Wisdom of Crowds," in which one would be able to obtain important and precise information from a big crowd. The Wisdom of Crowds is the idea that through the aggregation of estimations from a large crowd, would result in a more precise estimation when compared to the estimation of one subject. In this paper, the author studied a weight-judging competition in which 800 people guessed the weight of oxen, at the end the crowd was correct within 1 per cent of the real value, while one an individual estimate would fluctuate from -3.7 per cent and +2.4 per cent from the real value (Galton, 1949). The results obtained confirmed that a crowd can be a good source of knowledge and a precise forecaster.

### 2.2. Crowdsourcing.

Crowdsourcing is defined as being the practice of obtaining information or input into a task or project by enlisting the services of a large number of people, either paid or unpaid, typically via the internet (Oxford dictionaries, 2019).

Crowdsourcing is not a concept that is exclusively applicable to forecasting, as many companies are using it as a means of reducing their marketing costs. For example, instead of hiring a photographer for a hospital campaign and paying hundreds of dollars, a company would instead opt to get their pictures from a website called iStockphoto, for about \$1 each. Another way of how companies are using crowdsourcing is through the reduction in costs that R&D entails. This is perhaps best demonstrated by InnoCentive, which is a website where companies can place their unsolved problems. Either subjects or a crowd, can they try and solve these kinds of problems and get paid for their service. Companies pay anywhere from \$10.000 to \$100.000 for this service, which is far more cost-effective than solving this problem cost in-house or outsourcing (Howe, 2006).

#### 2.2.1. Prediction Markets.

Prediction markets are exchange markets, in which a group of people speculate on a vast array of events, ranging from stock prices to election results. Market prices indicate how likely a crowd think an event may occur.

Some example of prediction markets:

- Iowa Electronic Market: This prediction market was the original electronic market. IEM was developed by the University of Iowa in 1988 and its first experiment was the 1988 presidential election. At that time IEM allowed trading a contract that would pay \$0.025 per for each percentage point of the popular vote of the election (Wolfers & Zitzewitz, 2004). Currently, IEM allows traders to invest real money (\$5-\$500) and operates twenty-four hours a day. Traders exchange different contracts in the areas of politics, economics, movies, etc. The payoffs of the contracts traded depend on real-life outcomes, like election results, an economic indicator, box office, etc(Iowa Electronic Markets, 2018). From the 1988 presidential election to the 2000 presidential election, IEM predicted the outcome of the elections with an average absolute error of around 1.5 percentage points. For example, polls made by Gallup had a 2.1 percentage point of error (Wolfers & Zitzewitz, 2004).
- Hollywood Stock Exchange: In this market, subjects trade on different areas of the movie industry like total box office, opening week box office, who will win Oscar's, tv ratings, etc. To perform these trades, subjects use virtual currency. In the year 2015, HSX predicted almost all the nominees of the main categories of The Oscar's (Hickey, 2015).
- Transfermarkt: Is a German-based website in which users estimate the transfer value of players. Players values have been used to predict football outcomes with success. (Peeters, 2018).

Prediction markets are designed to continuously give the latest forecast prior to the results of the events (Atanasov et al., 2015). In a binary option, traders bet on the probability of an event. For example: In a tennis match a contract that pays \$1 if player A wins. The price in which that contract is trading is derived from probability off the likelihood of the event occurring. In this case, if the contract is trading in \$0.20, the traders are assigning a probability of 20% for player A to win the match. In theory, for a market to work two features must be present. First, the market must have enough traders, so the aggregation is relevant to predict the outcome of the event. Second, the market should facilitate the aggregation of disparate information (Berg, Forsythe, Nelson, & Rietz, 2008).

### 2.2.2. Prediction Polls.

The difference between a prediction poll and a survey is that in the prediction poll the subject is going to be asked to predict an event, in a sentiment survey subjects are asked to say which event they prefer to happen. As an illustration, in prediction poll, one would be asked “Who do you think is going to win the next presidential election?” whereas in a sentiment survey you would be asked, “For which candidate are you voting in the presidential election?”(Atanasov et al., 2015).

Examples of prediction polls:

- Economic indicators expectation run by Central Banks. For example, in Costa Rica, the Central Bank runs a survey every month, in which the Central Bank asks economists for their expectations of different economic indicators, usually for the next twelve months. Over one hundred economists answer the survey, the different averages of these predictions are officially published as expectation indicators of the Central Bank.
- Sports prediction competitions. ESPN publish the predictions of their expert analyst for the NFL matches and keep score of these predictions. ESPN includes “Pigskin predictions” as one of their analysts, “Pigskin” is the name of the prediction game in which regular people place their predictions. During the 2018 NFL season after 17 weeks, the crowd (Pigskin) had better forecast results than six experts and worse than four expert analysts (ESPN, 2019).
- Election prediction polls. This is one of the most common uses of prediction polls. It is important to differentiate between prediction pools, where subjects are asked “Who do you think is going to win” and sentimental surveys, where subjects are asked, “Which candidate are you voting for in the next election?” (Atanasov et al., 2015).

### 2.3. Conditions required for WOC.

According to Simmons et al. four conditions are required to assure that the crowd is reliable. First, the crowd must be knowledgeable. Second, the subjects must have an incentive to be correct in their predictions. Third, the observations must be independent. Fourth, the crowd must be diverse. (Simmons et al., 2010).

#### 2.4. The wisdom of crowds in sports.

The wisdom of the crowd in sports has been part of many investigations and their conclusions are important in many day-to-day situations. One of these findings is the favourite bias. This tendency is shown in betting with the spread. The spread is when a certain amount of points are subtracted from the favourite team. If you bet for the favourite team, to win the bet, the favourite team must win with a bigger point differential than the spread. If the favourite win with fewer points than the spread or lose, then the underdog bettors win the bet. However, if the score difference is exactly the same as the spread, then the bettors get their money back. An example of this would be, team A is the favourite and the spread is +4. The final score is 105 points for Team A and 100 points for Team B. Those who had bet on Team A, would have won. However, had the final score being 103 points for Team A and 100 points for Team B, those who had bet on Team B would have won this time. Lastly, everyone would have got their money back had the score been 104 points for Team A and points for Team B (Levit, 2004).

Furthermore, it used to be a widely held popular belief that bookmakers moved the line according to betting placement, keeping an equal amount of money going to either option, due to the precision of betting lines it was thought that betting lines were one of the best examples of the wisdom of crowds. Levitt showed that this was a wrongly held belief by showing how bets are placed and the shape of the distribution is not a normal distribution, the distribution was one-sided (Levit, 2004).

In a different investigation, the authors showed that the favourite bias is present in gambling. When a line is well placed, a bettor can be indifferent to either option, but they have to pick one, so there is a tendency that can be attributed to favourite bias. To test this favourite bias, the authors moved the line to make betting for the underdog more attractive. Even after this change, bettors kept betting on the favourite, even if this bet was clearly not a good bet (Simmons et al., 2010).

Another common bias in gambling is sentimental bias. Sentimental bias is when emotions make you take a decision that is not the best for you. In finance, sentiment bias happens when the return on stocks is influenced by how well known the company is. Subjects have a preference to invest in companies that they know over companies that they don't.

In sports, subjects will not bet against the team they support. This creates an incentive for the bookmakers to move the line against the team with a bigger fanbase to increase the revenue of the book (Forrest & Simmons, 2008).

Contrary to previous studies, Herzog and Hertwig used the sentimental bias in their favour. The authors use collective recognition heuristic to forecast the winners of football and tennis matches. To create this recognition variable, the authors interviewed subjects and asked them to identify the maximum number of players that they could. Teams with more recognizable players for football were the favourites in the forecasts and recognition was a variable included in the forecast tennis model. In this study, authors showed that using collective recognition heuristic as a forecaster is better than a guess chance or than FIFA Rankings, but is not a better forecaster than betting odds (Herzog & Hertwig, 2011).

Another example of the wisdom of crowds in sports is using individual valuation of players on an aggregate level to determine the strength of a team. In the paper *Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results*, the author studied the aggregation of individual estimations of the value of players made by users of Transfermarkt to forecast the outcome of international football matches. The authors based their approach in the values that users assigned in Transfermarkt website. Transfermarkt is a German-based website that provides sports information like fixtures, results and the most relevant to the wisdom of crowds players value estimations (Peeters, 2018).

### **3. Main Question.**

Is social media a good source of wisdom of crowds?

Today, people are spending hours on many social media platforms. As a result of this, for most people, social media is the primary source from where they get their information about current events, sports, celebrity gossip, etc. Given the significant role that social media plays in our everyday lives and the fact that our dependence on social media seems to be growing at a seemingly never-ending rate. The wisdom of crowds is based on the idea that good group judgement can be generated from the aggregation of individual judgments and this study aims

to find out whether Instagram polls can be used as a reliable source for the wisdom of the crowd.

A database of Instagram polls of a sports event was generated to study its effectiveness in forecasting sports events. The Instagram surveys were tested to identify if are a better choice than guessing. In addition, tests were run to further determine if the polls could provide better forecasts than experts' forecasts or other benchmarks.

#### **4. Methodology.**

##### 4.1. Hypothesis

Based on previous literature regarding the wisdom of crowds I wanted to test the following hypothesis.

H<sub>0</sub>: The proportion of correct predictions forecasted correctly by Instagram polls  $\leq 0.5$

In binary choice, the expected value of guessing correctly is 0.5. Instagram Polls have to achieve better results than a simple chance guess to be considered an advantageous source of wisdom of crowds.

H<sub>1</sub>: The mean of the matches forecasted correctly by Instagram polls = the mean of the matches forecasted correctly by experts.

The wisdom of crowds is based on the idea that good group judgement can be generated from individual judgments. Comparing the results of the crowd versus the results of experts can confirm if the Instagram crowds can be as good or even better than the experts.

H<sub>2</sub>: The mean of the matches forecasted correctly by social media = The mean of the matches is forecasted correctly by a benchmark.

More traditional methods have proven to be effective in predicting, but with the advantageous reach of social media, it is important to determine if the Instagram surveys can be an adequate alternative for these more traditional methods.

Finally, social media variables accuracy versus other benchmarks. These benchmarks are some of the most precise and most referenced sports forecasters.

For Instagram polls to work as a significant platform to generate prediction polls, according to Simmons, four conditions must be met for the wisdom of crowds: 1) knowledgeable, 2) motivated to be accurate, 3) independent and 4) diverse. In this thesis it will be examined if these Instagram surveys are as precise as several benchmarks.

#### 4.2. Instagram Polls.

To test the hypothesis, it was necessary to generate a database of Instagram polls on football and basketball. For the football database, from June 14<sup>th</sup>, 2018 up to June 28<sup>th</sup>, 2018 a database was made by collecting data from polls run by the @fifaworldcup Instagram account. In this account, polls were made for every match. The FIFA page posted the polls during the morning hours and the cut to collect the information was before the start of each match. The games in the first two rounds were played on three different hours, with the first match starting at 3 pm, the next at 5 pm and the last at 8 pm. For the third round, two games were played at 4 pm and two other games at 8 pm.

From this sample, the level of accuracy is going to be tested in three different ways. First, a binomial test is going to be carried out to determine if Instagram polls are better than flipping a coin to forecast the winner of the match. Second, a McNemar test is going to be carried to determine if Instagram polls forecast is equally precise as a specialist forecast or not. Finally, a McNemar test is going to be performed to determine if Instagram polls forecasts are equally precise as the forecast from other benchmarks (FIFA World Ranking, Betting Odds, SPI model and Instagram Followers media).

For the basketball database data was collected from the polls made on the @nbaontnt Instagram account every time the television network is going to broadcast a game during the period from April 6<sup>th</sup>, 2018 up to Feb 14<sup>th</sup>, 2019. NBA games broadcast on national television (USA) are normally played at 2 am European time, if a second game is broadcasted, it usually starts two hours later. The surveys began after noon and the cut time at which the information was collected was at the time of the start of the first match of the day, usually about 12 hours after the polls were posted. The question asked on the polls was which team is going to be the winner of the game.

The level of forecast accuracy of these polls will be tested at different levels. First, a binomial test will be carried out to determine if it is better or not to flip a coin or the Instagram polls to forecast the result of a match. Second, the percentage of correct predictions will be compared against the level of correct predictions of other benchmarks (CARMELO model and odds), and determine if there is a difference between polls, a mathematical model or betting odds. Third, the Receiver Operating Characteristic (ROC) curve will measure how good the forecaster is identifying favourites as the winners. Taking the voting level as the probability of winning of the team. As the voting percentage by the fans increases, the probability of predicting the result of the game correctly also should increase the same principle applies to the winning probability of the CARMELO model.

In both cases, the polls were posted @fifaworldcup and @nbaontnt accounts the day of the matches, before the start of each game the results were recorded. The participation on the polls is private, and only the account manager can see the number of participants, but as the @nbaontnt account has about 3 million followers and @fifaworldcup account has about 11 million followers, it is assumed that there was enough participation on these polls.

#### 4.3. Benchmarks.

In the database every time one of the benchmarks makes a correct forecast, a score of 1 is awarded and each time the forecast is incorrect, a score of zero is awarded. The benchmarks can be a binary prediction or probabilistic prediction (from the probability, the binary prediction is inferred).

##### 4.3.1. Experts predictions.

Every day during the FIFA World Cup, experts from all around the world gave their predictions of the outcome of the matches. For this research, four different sources are going to be considered. One prediction from one of the panellists of ESPN FC (United States football TV show), one from ESPN Deportes (United States football TV network in Spanish), another from ESPN Mexico and finally predictions from Mark Lawrenson from BBC Sports in the United Kingdom.

Data base made with data from: ESPN (ESPN, 2018j, 2018o, 2018m, 2018i, 2018b, 2018n, 2018e, 2018h, 2018a, 2018g, 2018k, 2018l, 2018f, 2018d, 2018c) and BBC (Bevan, 2018c,

2018b, 2018a). Only matches in which the expert picked a team to win and one of the teams won were included in the sample excluding all ties, making the sample a binary prediction.

#### 4.3.2. FIFA/Coca-Cola World Ranking.

The FIFA/Coca-Cola World Ranking is a ranking made by FIFA every month. The ranking is based on the past performance of each team. For the FIFA World Cup, rankings determine the tiers for the group draw. The seven best-ranked teams plus the host country are in the first tier of the draw, the next eight are in the second tier and so on. Thus, this ranking is relevant as it is used to separate the best teams into different groups after the draw.

In this thesis, the team with a higher ranking is going to be the favourite to win a match in a binary prediction. The last ranking before the FIFA World Cup was published on July 7th, 2018 (FIFA, 2018).

#### 4.3.3. Instagram Media Followers

This benchmark is important since in a certain way it allows us to compare the accuracy of the surveys against the predictions we would have if a binary prediction were determined only by which team has the players with a higher median number of followers on Instagram.

Data was collected to perform two different tests. For the first one, before the FIFA World Cup, gathering all the Instagram accounts available of the players that were about to play in the FIFA's World Cup 2018 to generate a database. Then, before every round, the database was updated with the most recent amount of Instagram followers of every account available to ensure that the database had the latest information.

The first round started on June 14th, 2018, the second round on June 19th, 2018 and the third on June 25th, 2018. All data from Instagram was download through RStudio and exported to Microsoft Excel. Then a ranking was made, in which the team whose players had more followers was regarded as the favourite to win the match. In case the match ended in a draw, it was taken out of the sample.

In *The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition*. The authors interviewed subjects at the University of Basel to come up with their collective recognition heuristic variable. To give more representation to all the other countries of the

world in the collective recognition variable and avoid giving an advantage to European countries, the use of social media seems appropriate.

#### 4.3.4. Betting Odds.

Odds are the probability that a particular event will or not happen. In gambling, odds are a probability expressed in numbers. From the betting odds a probabilistic prediction will be obtained, as in the FIFA/Coca-Cola World Ranking, every time that a team was the favourite in the betting odds, won the match value of 1 was assigned, 0 if the favourite lost and if the match ended in a draw, it was taken out of the sample.

For the analysis of the World Cup, before every round, the betting odds were updated using the information provided by [www.oddshark.com](http://www.oddshark.com). This website provides an average of odds from a significant source of bookmakers (oddshark, 2018). And for basketball, the betting odds were taken from [www.oddsportal.com](http://www.oddsportal.com). Oddsportal.com is a website which compares odds from over eighty different bookmakers. (Oddsportal.com, 2019).

Historically bookmakers have been very accurate in determining favourites and betting lines. Therefore, in investigations about forecasting accuracy in sports, betting odds are the gold standard and are taken as a reference point.

#### 4.3.5. SPI model.

FiveThirtyEight (a company that is part of ESPN) makes the soccer power index or SPI ratings. This index is often referenced in football broadcasts and media. It is an estimate of the strength of the teams. The index is composed of an offensive rating with the goals that a team would be likely to score and a defensive rating with the expected amount of goals that a team would concede. Then, the index of one team is compared against the index of the other team to estimate the probability of a victory for team A, the probability of victory of team B and the probability of a draw. From the probabilistic prediction, the team with higher probabilities to win the match will be designated as the favourite and hence a binary prediction between favourite vs. non-favourite is inferred. Before every round, the data from SPI was taken from [www.fivethirtyeight.com](http://www.fivethirtyeight.com) to have the latest forecast (Boice, Silver, et al., 2018).

#### 4.3.6. CARMELO model.

Also known as the career-arc regression model estimator with local optimization. CARMELO is a model made by FiveThirtyEight. This model uses player ratings (a combination of real plus-minus and box plus/minus) to track how much talent there is on each team, it also considers injuries, trades and team chemistry (Boice et al., 2019; Boice, Koeze, Silver, & Paine, 2018). As in the SPI model, the index of each team is compared against the other team index to estimate the probability of a victory for team A and the probability of victory of team B. From the probabilistic prediction, the team with higher probabilities to win the match will be designated as the favourite and hence a binary prediction between favourite vs. non-favourite is inferred.

### **5. Results.**

#### 5.1 Descriptive statistics.

##### *FIFA WORLD CUP Instagram polls.*

In total 48 matches were played in the group stage of the FIFA World Cup 2018. Of the 48 matches, 39 matches ended up having a winner and 9 ended up in a tie. All games that were a tie were taken out of the sample. Of these 39 games, Instagram polls forecasted correctly the winner 27 times (69.23%). While the benchmarks, SPI model from FiveThirtyEight and the betting odds from bookmakers forecasted the winner correctly 29 times (74.36%) giving them a better performance than Instagram polls. On the other hand, FIFA Ranking forecasted the correct winner 26 times (66.67%) and Instagram Media Followers 24 times (61.54%) underperforming the Instagram polls (Figure 1).

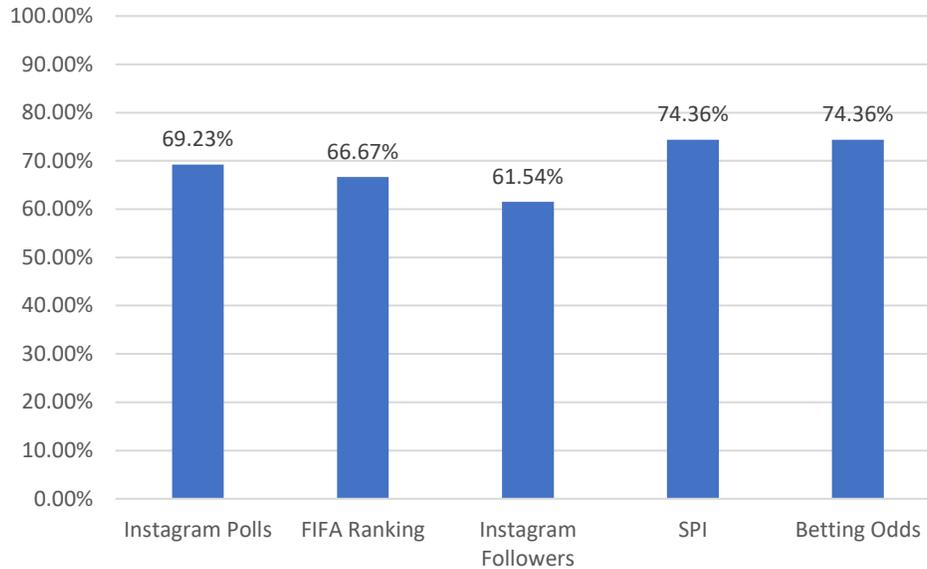


Figure 1: Forecast FIFA World Cup Accuracy benchmarks.

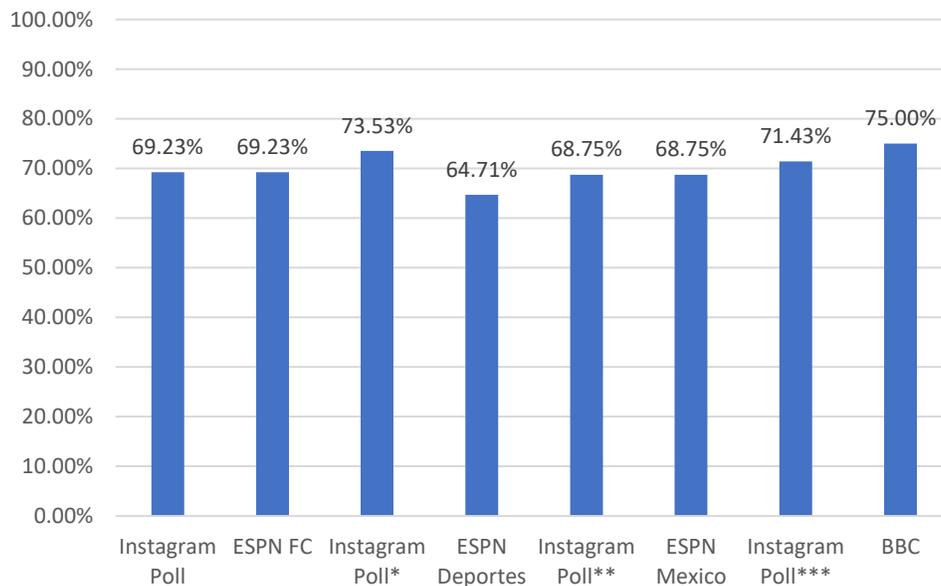


Figure 2: Forecast correct-incorrect matrix specialists.

On a matched sample, where only the games in which both the poll and the specialist predicted one of the teams to win and the game ended with one of the teams winning the match (no draws are in the sample). ESPN experts and Instagram Polls predicted the correct winner with a rate of 69.23% on a sample of 39 games, ESPN Deportes 64.71% vs Instagram Polls 73.53% on 34 games, ESPN Mexico and Instagram Polls both 68.75% on 32 games versus and BBC 75% vs Instagram 71.43% on 28 games (Figure 2).

*NBA data.*

The sample consisted of 62 matches, in which the Instagram polls correctly predicted 39 matches (62.90%), the Betting Odds correctly predicted 40 out of 62 (64.52%), the CARMELO model from FiveThirtyEight predicted correctly the winner 38 matches (61.29%). In the subsample of games played late in the season, when fans pay more attention to games, Instagram polls correctly predicted 23 out of 27 matches (79.31%), Betting Odds 22 out of 27 (81.48%) and the CARMELO model predicted 20 out of 27 (68.97%) (Figure 3).

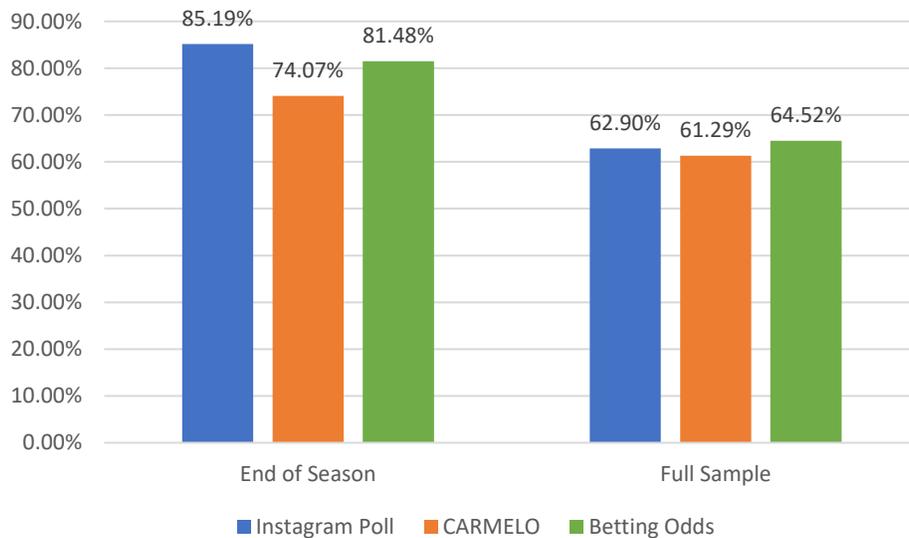


Figure 3: Forecast correct-incorrect matrix NBA playoffs.

## 5.2. Main Results.

*Instagram polls against guess chance.*

Binomial probability test is used to test how likely the outcome happens, given the probabilities on the null hypothesis ( $H_0 \leq 0.5$ ). Binomial Probability test is used when the population exists only in two classes (correct or incorrect or favourite and non-favourite, etc) (StataCorp LLC, 2019).

From the Binomial test,  $H_0$  is rejected for FIFA WORLD CUP polls as  $0.6923\% > 0.5$  at confidence levels of 0.10, 0.05 and 0.01. Also,  $H_0$  is rejected for NBA polls as  $0.6290\% >$

0.5 at confidence levels of 0.10, 0.05 and 0.01. For the sub-sample of NBA Playoffs,  $H_0$  is rejected as  $0.8519\% > 0.5$  at confidence levels of 0.10, 0.05 and 0.01. In contrast,  $H_0$  cannot be rejected for NBA prepositions as  $0.4592 < 0.5$  at any confidence level. Also,  $H_0$  cannot be rejected for NBA prepositions playoffs as  $0.4667 < 0.5$  at any confidence level (Appendix 1).

*Instagram polls against human experts.*

Table 1: McNemar test matrix, Instagram polls vs specialist.

Variable	Instagram Poll			p-value	
	correct	incorrect	total		
<b>ESPN FC</b>	correct	21	6	27	1.0000
	incorrect	6	6	12	
	Total	27	12	39	
<b>ESPN Deportes</b>	correct	21	1	22	0.3750
	incorrect	4	8	12	
	Total	25	9	34	
<b>ESPN Mexico</b>	correct	18	4	22	1.0000
	incorrect	4	6	10	
	Total	22	10	32	
<b>BBC</b>	correct	19	2	21	1.0000
	incorrect	1	6	7	
	Total	20	8	28	

The McNemar test is used to compare two paired samples when the data is nominal and dichotomous (success or failure) (Mccrum-Gardner, 2008). Is the test used for  $H_1$  and  $H_2$ . From the McNemar test, when comparing Instagram polls against the specialist forecasts  $H_1$  cannot be rejected for any of the variables at confidence levels of 0.10, 0.05 and 0.01 as the p-value for Instagram polls vs ESPN FC as it is shown in Table 1 is 1.0000, Instagram polls vs ESPN Deportes p-value is 0.3750, Instagram vs ESPN Mexico p-value is 1.0000 and Instagram polls vs BBC p-value is 1.0000. So, it is concluded that the proportion of matches correctly forecasted by Instagram polls is statistically no different from that of the predictions made by the experts during the FIFA World Cup 2018.

*Instagram polls against benchmarks.*

Table 2: McNemar test matrix, Instagram polls vs benchmarks.

Variable	Instagram Poll			p-value	
		correct	incorrect		total
<b>Betting Odds</b>	correct	26	3	29	0.6250
	incorrect	1	9	10	
	total	27	12	39	
<b>Ranking</b>	correct	24	2	26	1.0000
	incorrect	3	10	13	
	total	27	12	39	
<b>SPI</b>	correct	24	5	29	0.7266
	incorrect	3	7	10	
	total	27	12	39	
<b>Instagram Media</b>	correct	22	2	24	0.4531
	incorrect	5	10	15	
	total	27	12	39	

The results of the McNemar test showed that  $H_2$  cannot be rejected at a confidence level of 0.01, 0.05 or 0.10 as it is shown in Table 2, the p-value of Instagram polls versus Betting Odds is 0.6250, versus Ranking, is 1.0000, SPI is 0.7266 and when Instagram polls are compared against Instagram Media Followers the p-value is 0.4531. Therefore, it can be concluded that during the World Cup Instagram polls provide just as accurate forecasts as those provided by the benchmarks.

Table 3: McNemar test NBA end of the season sample matrix, Instagram polls vs benchmarks.

Variable	Instagram Poll			p-value	
		correct	incorrect		total
<b>CARMELO</b>	correct	18	2	20	0.4531
	incorrect	5	2	7	
	total	23	4	27	
<b>Betting Odds</b>	correct	21	1	22	1.0000
	incorrect	2	3	5	
	total	23	4	27	

Also seen from on Table 3 the results of the McNemar test when Instagram polls are compared to the benchmarks on the NBA playoffs sub-sample,  $H_2$  cannot be rejected at a confidence level of 0.01, 0.05 or 0.10 as the p-value for Instagram polls versus CARMELO is 0.4531 and for Instagram polls versus Betting Odds is 1.0000. Therefore, it is concluded that Instagram polls provided just as accurate forecasts during the NBA season finale as those provided by benchmarks. Also, on the full season sample,  $H_2$  cannot be rejected at a confidence level of 0.01, 0.05 or 0.10 as Instagram polls versus CARMELO forecast has a p-value of 1.0000 and Instagram polls versus Betting Odds 1.000 is it is shown on Table 4. Therefore, it is also concluded that Instagram polls provided just as accurate forecasts during the NBA season as those provided by benchmarks.

Table 4: McNemar test NBA full sample matrix, Instagram polls vs benchmarks.

Variable	Instagram Poll			p-value	
	correct	incorrect	total		
<b>CARMELO</b>	correct	29	9	38	1.0000
	incorrect	10	14	24	
	total	39	23	62	
<b>Betting Odds</b>	correct	32	8	40	1.0000
	incorrect	7	15	22	
	total	39	23	62	

### 5.3. Expertise Accuracy ROC - AUC.

The statistics presented in this section are important as they are related to the ROE curve presented in this section. Shown here is the percentage of voting that the favourite teams had and what percentage of successful forecasts were obtained according to the different ranges of favouritism.

Table 5: Favouritism distribution.

	Instagram Poll		CARMELO	
	End of Season	Full Sample	End of Season	Full Sample
Wrong	56.5%	66.13%	68.0%	67.96%
Right	67.0%	68.15%	68.0%	68.92%

In the full sample when the crowd picked the right winner, on average the correct option got 68.15% of the votes. And when the crowd was wrong, the incorrect option got 66.13% of the votes. At the end of the season, when the crowd picked the winner correctly, on average 67% voted for it and when the crowd picked the wrong option, that option was on average 56.5% of the votes (Table 5).

In the CARMELO model from FiveThirtyEight, when the model predicted right or wrong the winner of the match at the end of the season, in both cases during the regular season the model gave a chance of success of 68% to the favourite. In the full sample of the Carmelo model, when the model forecasted right the favourite team to win the game, gave to it a 68.92% probability to win, and when the CARMELO model predicted picked a team to win and the result was a lost for that team, the probability assigned by the model was 67.96%.

Shown in Figure 4, is that the accuracy of the forecasts did not increase with an increase in the voting percentage from the crowd and the favouritism percentage on the CARMELO model from FiveThirtyEight.

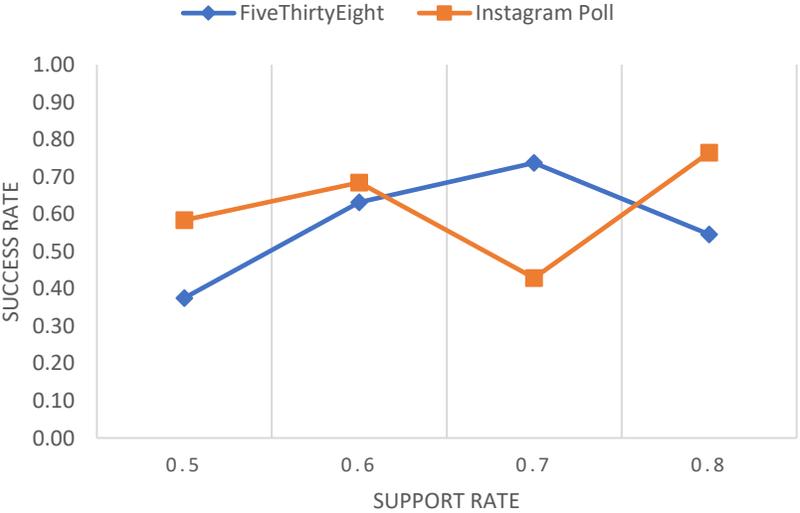


Figure 4: Full sample NBA matches, correct prediction rate per crowd percentage support.

Also, shown in Figure 5 is that the accuracy of the Instagram polls on the subsample of the playoffs is very high, as the accuracy increased as the voting difference between favourite and non-favourite also increased. For the CARMELO model, the accuracy is not increasing when the favouritism percentage increases.

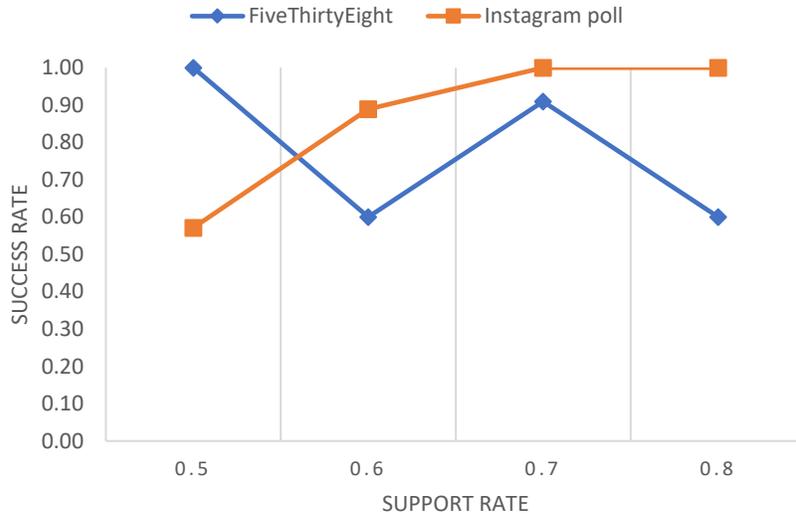


Figure 5: NBA playoffs matches sub-sample, correct prediction ratio per crowd support level playoffs.

*NBA Propositions.*

Table 6: Favouritism distribution propositions.

	Season	Playoffs
Wrong	64.92%	65.29%
Right	67.51%	65.90%

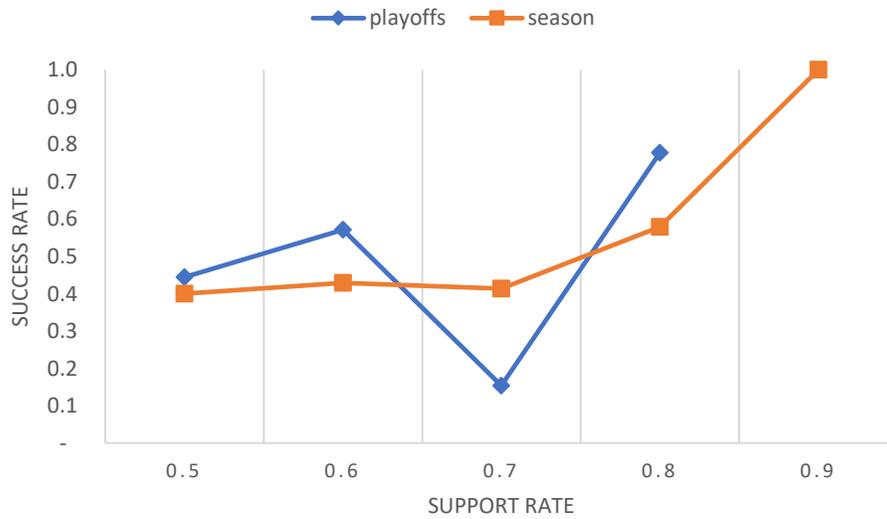


Figure 6: NBA prop Instagram polls, success rate per support rate level.

During the season, on average when the crowd was correct on average the favourite got 2.59 p.p. more votes than when the crowd was wrong. In the playoffs sub-sample, this difference is even smaller, just 0.61 p.p. (Table 6). Regarding to the propositions, the accuracy of the crowd increases slightly with the increase in the voting percentage received by the favourite. On the other hand, during the playoff show, in which there is great precision on the part of the crowd to select the winner of the matches, for the propositions no precision is shown in Figure 6.

*Area Under Curve (AUC) – Receiver Operating Characteristics (ROC) curve results.*

ROC is a probability curve and AUC is the area under the ROC. In a good model, the result of the AUC is close to 1, while in a model that generates a random performance the AUC is close to 0.5. The model classifies the outcomes in four different classes: A positive result, that was classified as positive is a true positive, a positive classified as negative is a false negative, a negative result classified as negative is a true negative and a negative classified as positive is counted as a false positive. The closer the AUC to 1 is the better is the model to distinguish between classes (Fawcett, 2006).

The ROC graph can be used to visualize, organise and selecting classifiers based on their performance. In a ROC graph, a diagonal line is drawn on the graph which indicates a random guess. If our model line is above that line it means that it has a better performance than a random guess, the opposite happens when our model line is below the diagonal line (Fawcett, 2006).

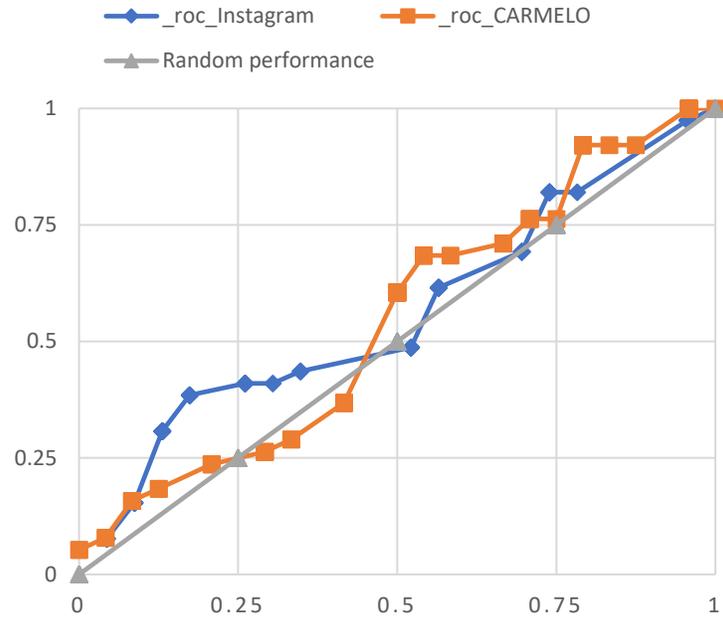


Figure 7: All NBA matches ROC curves for Instagram polls and CARMELLO model.

On the ROC for the full sample of the NBA matches. The area under the curve of the Instagram crowd model has an observed coefficient of 0.5206243 a bias of -0.000275 and a standard error of 0.0742411. The AUC of CARMELLO has an observed coefficient of 0.5120614, a bias of -0.00116 and a standard error of 0.0783813 (Appendix 9).

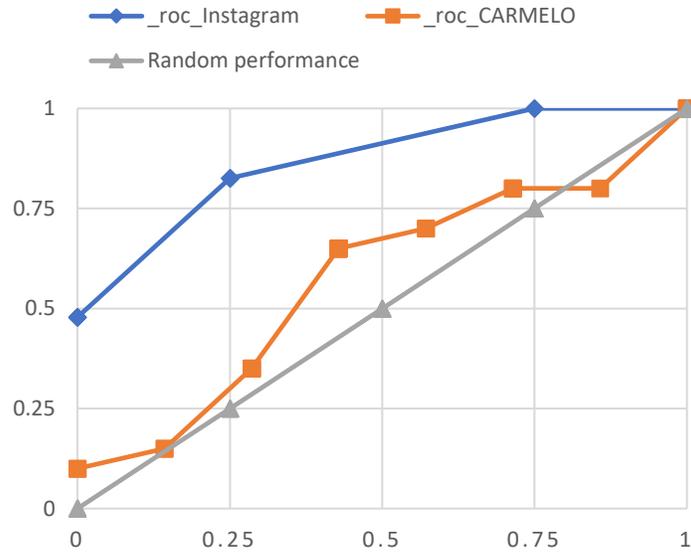


Figure 8: NBA playoffs matches sub-sample ROC curves for Instagram polls and CARMELO model.

On the subsample of NBA playoffs matches, the area under the curve of the Instagram polls from the playoffs model has an observed coefficient of 0.7826087, a bias of -0.002595 and a standard error of 0.1303103. The AUC of CARMELO has an observed coefficient of 0.5071429, a bias of 0.0019694 and a standard error of 0.1318163 (Appendix 10).

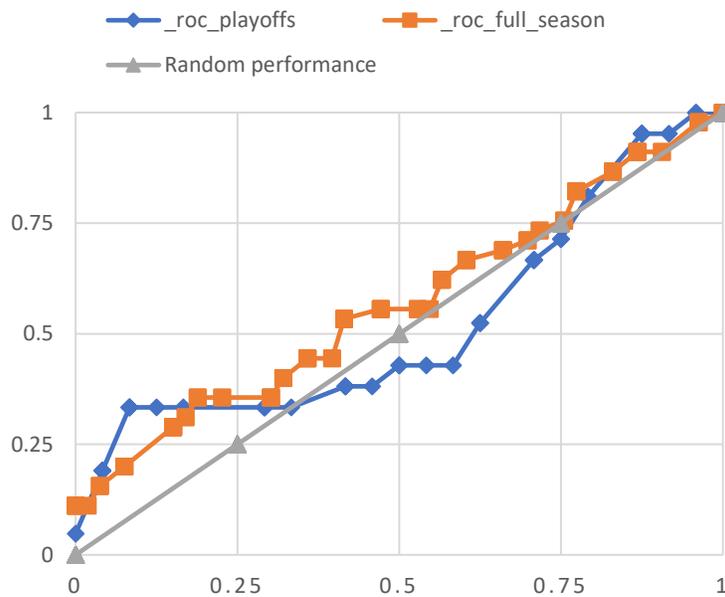


Figure 9: NBA propositions ROC curves for Playoffs sub-sample and All Season sample.

NBA propositions AUC regression. The area under the curve of the playoffs sample has an observed coefficient of 0.4960317, a bias of 0.0007585 and a standard error of 0.090173. The AUC of Full Sample has an observed coefficient of 0.5509434, a bias of 0.0010587 and a standard error of 0.0594852 (Appendix 11).

## **6. Discussion.**

### 6.1 Results discussion

The wisdom of crowds is based on the idea that good group judgement can be generated from individual judgments. But how accurate and reliable is the wisdom of crowds in predicting sports outcomes? Can we trust better the judgments of the crowds over prediction models or betting odds? In which situations is it reliable and in which situations is it unreliable? How much does the size of the majority increase the precision of the forecast? Those were some of the questions I attempted to answer.

The first step focused on seeing whether the crowd could be more accurate than a random coin toss when it comes to predicting the winner of a match. To solve this question, binomial tests were performed on football and basketball samples. The football events were exclusive to the winner's prediction, while the basketball events were about predicting the winner as well as other events that occur during a basketball game.

According to the binomial tests,  $H_0$  the hypothesis that the forecasts of the crowd were equal to tossing a coin or a 50% probability for the prognosis of football and basketball matches was rejected. In the case of propositions in basketball, the hypothesis cannot be rejected. This gives an initial green light to the crowd that can be a good source of information to know which team is the favourite to win a football match (69.23% success rate) and a basketball match (62.90% success rate on the full sample and 85.19% on the playoffs sample). But it is a red light in the case of forecasting basketball propositions as the forecast success rate was under 50% (45.92% full sample and 46.67% playoff sub-sample).

After demonstrating that the crowd outperforms the coin toss when it comes to predicting the winner of a match, the next step for the football sample was to be compared against the individual choices of the experts. The ESPN FC specialist was right on 69.23% of the matches versus 69.23% rate from the Instagram crowd, the ESPN Deportes specialist was right on 64.71% versus 73.53% rate from the Instagram crowd, ESPN Mexico 68.75% versus 68.75% rate from the Instagram crowd and BBC predicted the matches with a 75% rate versus a 71.43% rate from the Instagram crowd.

On the first instance, the crowd had better performance than one of the experts, the same performance as two of the experts and one worse performance than an expert. After carrying out the respective McNemar tests,  $H_1$  the hypothesis that the forecasts of the crowd have a different performance to the forecasts of the experts was not rejected.

After showing that following the advice of the crowd manages to keep up with the advice of experts. It was time to compare the accuracy of the forecasts of the crowd against the accuracy of the forecasts of other benchmarks. This time the Instagram polls had better performance in their forecasts (69.23%) than the FIFA Ranking (66.67%), and the media number of followers on Instagram (61.54%), but the forecasts were not better than those of the betting odds (74.36%) or the SPI model (74.36%). However, once again after carrying out the McNemar tests, we cannot reject  $H_2$  the hypothesis that the Instagram crowd has a different success predicting the winner of the matches to none of the benchmarks.

At the same time, the Instagram surveys of the NBA were compared to their benchmarks. On this occasion the surveys were studied in two ways: 1) The total of the observations. 2) A sub-sample with the observations made at the end of the season (playoffs) since at this moment the attention by the public is much higher than at any other moment during the season and the crowd could be more aware of the decisions it makes.

This difference between the decisions of the crowd during the full sample and the subsample of the playoffs is evident, with the forecasts of the sub-sample of the playoffs being more accurate by 22.28 p.p. In the sample, the Instagram polls guessed the winner of 62.90% matches correctly, the odds guessed right on 64.52% and the CARMELO model predicted the winner correctly on 61.29% of the matches. In the playoff subsample, the polls guessed

correctly 85.19% matches, the odds 81.48% and the CARMELO model 74.07% In this subsample, there is an advantage on the part of the crowd.

As with football polls, McNemar tests are conducted to rule out the hypothesis  $H_2$  that Instagram polls are not equally effective in predicting the winner of a basketball when match with their benchmarks. And just like football polls, the hypothesis could not be ruled out when compared with Odds or the CARMELO model, in the full sample or in the playoff subsample.

Finally, a ROC regression was performed to study the influence of the voting percentage on the effectiveness of the forecasts. Testing if greater voting percentage means greater precision in the forecast. The same was done for the level of favouritism indicated by the CARMELO model. In both cases, the coefficients in the regressions of the crowd were higher than the coefficients of the CARMELO model. In the complete sample (Figure 6), the difference between the AUC for Instagram Polls support rate and CARMELO winning probability rate is not much, in both cases are close to random performance value of 0.5.

In the playoffs sub-sample (Figure 7), the AUC for the variable of Instagram Polls support rate is around to 0.8, On this case, the model is accurate in identifying a favourite that won their games. On the other hand, the regression for CARMELO winning probability rate was again close to random performance value of 0.5.

The ROC curve regression was also performed for the NBA prepositions surveys. Here, the AUC for both variables (Playoffs subsample and Full sample) had a coefficient close to random performance value of 0.5. This added to that in the binomial test showed that these surveys are as effective in their forecast as flipping a coin. Therefore, we can conclude that the crowd can be trusted in situations where there is popular knowledge on the topic, for example, the winner of a match, the crowd easily without much investigation: know which team is better placed in the standings, if they have won or not in his last games. But we cannot trust that the crowd will be a good guide to the prepositions forecast. These situations require a deeper follow-up of what happens during basketball games day by day. For example, how many triples a player scores per game, how many rebounds, etc.

It should be noted that although none of the tests could show that when predicting the winners of the matches, the forecasts obtained by the Instagram polls were better or more accurate than those of the analyzed benchmarks. Nor can it be said that the forecasts were worse than those of the benchmarks. So, the next step is to distinguish which method is more appropriate, which requires fewer resources, which is easier to perform, among other variables. Similarly, we must think that as subjects increasingly use social media and ever-decreasing attention to traditional media such as television or even countries where betting sites are restricted, for example, the Netherlands. This results in the information of social media being one of the easiest source of information to access.

## 6.2. Limitations and Recommendations.

### *Sample size and events restrictions.*

The size of the sample is limited for several reasons. Firstly, because of the coverage of the events. The events must be large enough to have Instagram surveys with enough participation from the public as well as enough benchmarks to be able to make a comparison between the methods. For example, during the Fifa World Cup, there is great availability of forecasts by experts, since everyday experts give forecasts of the results of the matches, different mathematical models, betting odds, pools, among others.

Secondly, although the World Cup is the ideal event in Coverture, it is an event that lasts a month and takes place every four years. So, the number of observations is limited. In addition, events are not exclusively binary, there is the possibility of a winner, there may also be ties. These ties were eliminated from the sample.

On the other hand, the sample of surveys of the NBA has the limitation in that not all the matches of each day are covered, neither do the experts. Surveys are only made during matches that are aired on national television in the United States. On average, two to four matches are broadcast per week. Therefore, a large amount of time is required to have a large sample.

### *Conditions required for WOC.*

Among the conditions for a crowd to be reliable is that they must have knowledge. This is difficult to control in the Instagram poll since there is no way to disengage between a subject that is an avid follower or one with a limited passing interest in the events.

Another condition that is compromised is the incentive to be correct. The subjects carry out the surveys and there is nothing that keeps the record of their predictions if they were correct or not.

The other two conditions of which the observations have to be independent and diverse are greatly satisfied. This must be taken into account when making conclusions based on the obtained results.

### *The participation of enough subjects on the polls is assumed.*

The collection of information from the surveys was done by recording the voting percentages at the beginning of the events. But we do not have exact data on the number of votes for each of the options and the total number of participants. The community managers of the accounts were asked for the survey data but unfortunately, no response was obtained.

Given that surveys were carried out by accounts with millions of followers, a participation of a high number of subjects is assumed. So, it is not questioned that there is enough participation of subjects to be a crowd.

### *Binomial polls.*

Currently, Instagram surveys are limited to binomial polls. So, there is some limitation for the forecasts in events in which the result can have more than binomial outcomes. For example, if in the future the tie option were included, it would not be necessary to eliminate from the sample the matches that have ended in a draw or in which experts predicted a draw and could not be comparable. When surveys can be conducted with multiple options, they would be a more versatile option and a better tool to extract wisdom from the crowd.

### *Implications.*

Conducting these Instagram polls can be positive for traditional media. TV Networks can conduct polls on matches that they will broadcast keeping track of the results of these audience poll, which can result in greater engagement with their audience. Creating content that can be seen not only on television but via videos on Instagram, Facebook or YouTube. An example of this would be ESPN Nation, which was based on polls that ESPN made on their website. This program aired from 2012 to 2018. Doing the surveys on Instagram instead of a website may be more convenient if there is more internet traffic on Instagram for a TV Network.

In the same way, there are many internet sites specialized in generating sports forecast models. These forecasts could add the opinion of the crowd within the model or provide information on which team is favourite for the crowd on their website.

For the public, it is not just a way to feel more involved during party transmissions. It is well known that betting is a market that moves 100 billion dollars annually. For the public that participates in betting, the crowd can be a source of important information so that they can make more informed and make smarter bets.

These following facts illustrate the sheer size and scale of the Gambling Industry in just the United Kingdom:

- 18% of the bettors follow an account of betting companies on some form of social media. (Mar 2018) (Gambling Commission, 2018a).
- 45.8% of people participated in at least one form of gambling in the last 4 weeks. (Sep 2017)(GamblingCommission, 2018c).
- 18.4% participated in at least one form of online gambling in the past 4 weeks. (Sep 2017) (Gambling Commission, 2018c).
- 53% of online gamblers have placed a bet from their cell phone or tablet (Jun 18) (Gambling Commission, 2018a).
- £ 14.4 bn Total gross gambling yield. (Apr 2017 – Mar 2018) (Gambling Commission, 2018b).

- £5.4 bn Total GGY for the remote sector (Apr 2017 – Mar 2018) (Gambling Commission, 2018b).

Finally, it would be interesting to expand the study by measuring the accuracy of Instagram polls on forecasting the success of a new product, the box office revenue of a movie or the award nominations of a television show, among other predictions that are currently carried out in electronic markets.

## **7. Conclusions.**

In this research, it was shown that Instagram polls are an advantageous alternative in collecting individual opinions and getting a good group judgment. The validity of these crowd judgments was tested in the context of forecasting sporting events. The crowd judgments were about predicting the winner of the matches. On both football and basketball polls showed that the forecasts were better than guessing.

Although, statistically it was not possible to verify that the forecasts of the crowd obtained through the surveys of Instagram were better than that of the forecasts of experts or benchmarks. Likewise, it is not entirely negative that the results are not as good as the other ways of obtaining forecasts if the surveys are a simpler way of obtaining forecasts than benchmarks.

Similarly, in the subsample of NBA playoff polls even if it is not statistically significant that there is a better performance from the crowd than in the full NBA sample. This can be favoured by more attention to the events that are happening. Similar results are observed in the ROC regression since it was the one that gave the greatest area under the curve.

Likewise, the forecasts in which the crowd had worse performance were the propositions in basketball games. Which can also be explained by the amount of information that the crowd knows. In this case, the crowd had little information, so the forecasts were not precise.

Further research could go more in deep about measuring the accuracy of Instagram polls to forecast the success of new products, the box office of a movie or the success of a television

show, the promotion of tourist destinations, among other predictions that are currently carried out in electronic markets.

## 8. Bibliography.

- Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P. E., ... Mellers, B. (2015). Distilling the Wisdom of Crowds: Prediction Markets versus Prediction Polls. *Academy of Management Proceedings*, 2015(1), 15192–15192. <https://doi.org/10.5465/AMBPP.2015.15192abstract>
- Berg, J., Forsythe, R., Nelson, F., & Rietz, T. (2008). Chapter 80 Results from a Dozen Years of Election Futures Markets Research. *Handbook of Experimental Economics Results*, 1(C), 742–751. [https://doi.org/10.1016/S1574-0722\(07\)00080-7](https://doi.org/10.1016/S1574-0722(07)00080-7)
- Bevan, C. (2018a). World Cup 2018: Lawro's predictions - final group games. Retrieved from <https://www.bbc.com/sport/football/44594386>
- Bevan, C. (2018b). World Cup 2018: Lawro's predictions - second group games. Retrieved January 13, 2019, from <https://www.bbc.com/sport/football/44458950>
- Bevan, C. (2018c). World Cup 2018: Lawro predicts the opening round of group games. Retrieved January 13, 2019, from <https://www.bbc.com/sport/football/44397861>
- Boice, J., Dottle, R., Koeze, E., Wezerek, G., Silver, N., & Paine, N. (2019). 2018-19 NBA Predictions. Retrieved February 20, 2018, from [https://projects.fivethirtyeight.com/2019-nba-predictions/games/?ex\\_cid=rrpromo](https://projects.fivethirtyeight.com/2019-nba-predictions/games/?ex_cid=rrpromo)
- Boice, J., Koeze, E., Silver, N., & Paine, N. (2018). 2017-18 NBA Predictions. Retrieved August 1, 2018, from <https://projects.fivethirtyeight.com/2018-nba-predictions/games/>
- Boice, J., Silver, N., Dottle, R., Wezerek, G., Scheinkman, A., & Wolfe, J. (2018). 2018 World Cup Predictions. Retrieved June 25, 2018, from <https://projects.fivethirtyeight.com/2018-world-cup-predictions/matches/>
- ESPN. (2018a). How did our experts fare picking Costa Rica vs. Serbia, Germany vs. Mexico, Brazil vs. Switzerland. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759552/world-cup-predictions-costa-rica-vs-serbia-germany-vs-mexico-brazil-vs-switzerland](http://www.espn.com/sports/soccer/story/_/id/23759552/world-cup-predictions-costa-rica-vs-serbia-germany-vs-mexico-brazil-vs-switzerland)
- ESPN. (2018b). World Cup predictions: Alex Pareja and Steph Brantz share the honours. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759834/world-cup-predictions-australia-vs-peru-denmark-vs-france-iceland-vs-croatia-nigeria-vs-argentina](http://www.espn.com/sports/soccer/story/_/id/23759834/world-cup-predictions-australia-vs-peru-denmark-vs-france-iceland-vs-croatia-nigeria-vs-argentina)
- ESPN. (2018c). World Cup predictions: Belgium vs. Tunisia, South Korea vs. Mexico, Germany vs. Sweden. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759800/world-cup-predictions-belgium-vs-tunisia-south-korea-vs-mexico-germany-vs-sweden](http://www.espn.com/sports/soccer/story/_/id/23759800/world-cup-predictions-belgium-vs-tunisia-south-korea-vs-mexico-germany-vs-sweden)
- ESPN. (2018d). World Cup predictions: Brazil vs. Costa Rica, Nigeria vs. Iceland, Serbia

- vs. Switzerland. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759776/world-cup-predictions-brazil-vs-costa-rica-nigeria-vs-iceland-serbia-vs-switzerland](http://www.espn.com/sports/soccer/story/_/id/23759776/world-cup-predictions-brazil-vs-costa-rica-nigeria-vs-iceland-serbia-vs-switzerland)
- ESPN. (2018e). World Cup predictions: Damien Didonato edges Stewart Robson on final day of group play. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759930/world-cup-predictions-damien-didonato-edges-stewart-robson-final-day-group-play](http://www.espn.com/sports/soccer/story/_/id/23759930/world-cup-predictions-damien-didonato-edges-stewart-robson-final-day-group-play)
- ESPN. (2018f). World Cup predictions: Egypt vs. Uruguay, Morocco vs. Iran, Portugal vs. Spain. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759178/world-cup-predictions-egypt-vs-uruguay-morocco-vs-iran-portugal-vs-spain](http://www.espn.com/sports/soccer/story/_/id/23759178/world-cup-predictions-egypt-vs-uruguay-morocco-vs-iran-portugal-vs-spain)
- ESPN. (2018g). World Cup predictions: Experts all nail England, Belgium to win on high-scoring Day 5. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759650/world-cup-predictions-sweden-vs-south-korea-belgium-vs-panama-tunisia-vs-england](http://www.espn.com/sports/soccer/story/_/id/23759650/world-cup-predictions-sweden-vs-south-korea-belgium-vs-panama-tunisia-vs-england)
- ESPN. (2018h). World Cup predictions: France vs. Australia, Argentina vs. Iceland, Peru vs. Denmark, Croatia vs. Nigeria. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23701494/world-cup-predictions-france-vs-australia-argentina-vs-iceland-peru-vs-denmark-croatia-vs-nigeria](http://www.espn.com/sports/soccer/story/_/id/23701494/world-cup-predictions-france-vs-australia-argentina-vs-iceland-peru-vs-denmark-croatia-vs-nigeria)
- ESPN. (2018i). World Cup predictions: Gab Marcotti and Stevie Nicol share top honours on Day 12. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759818/world-cup-predictions-saudi-arabia-vs-egypt-uruguay-vs-russia-iran-vs-portugal-spain-vs-morocco](http://www.espn.com/sports/soccer/story/_/id/23759818/world-cup-predictions-saudi-arabia-vs-egypt-uruguay-vs-russia-iran-vs-portugal-spain-vs-morocco)
- ESPN. (2018j). World Cup predictions: Julien Laurens spot-on with scoreline projections for Australia-Denmark and Peru-France. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759746/world-cup-predictions-denmark-vs-australia-france-vs-peru-argentina-vs-croatia](http://www.espn.com/sports/soccer/story/_/id/23759746/world-cup-predictions-denmark-vs-australia-france-vs-peru-argentina-vs-croatia)
- ESPN. (2018k). World Cup predictions: Julien Laurens wins low-scoring affair, while Stewart Robson only expert to pick Russia. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759693/world-cup-predictions-colombia-vs-japan-poland-vs-senegal-russia-vs-egypt](http://www.espn.com/sports/soccer/story/_/id/23759693/world-cup-predictions-colombia-vs-japan-poland-vs-senegal-russia-vs-egypt)
- ESPN. (2018l). World Cup predictions: Leonard Bertozzi, Mauricio Pedroza share the spoils on Day 7. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759739/world-cup-predictions-portugal-vs-morocco-uruguay-vs-saudi-arabia-iran-vs-spain](http://www.espn.com/sports/soccer/story/_/id/23759739/world-cup-predictions-portugal-vs-morocco-uruguay-vs-saudi-arabia-iran-vs-spain)
- ESPN. (2018m). World Cup predictions: Two pundits get all three results correct. Retrieved January 10, 2019, from [http://www.espn.com/sports/soccer/story/\\_/id/23759810/world-cup-predictions-england-vs-panama-japan-vs-senegal-poland-vs-colombia](http://www.espn.com/sports/soccer/story/_/id/23759810/world-cup-predictions-england-vs-panama-japan-vs-senegal-poland-vs-colombia)
- ESPN. (2018n). World Cup predictions Costa Rica's late equaliser hands Javier Trejo Garay improbable comeback victory. Retrieved January 10, 2019, from

- [http://www.espn.com/sports/soccer/story/\\_/id/23759918/costa-rica-late-equaliser-hands-javier-trejo-garay-improbable-comeback-victory](http://www.espn.com/sports/soccer/story/_/id/23759918/costa-rica-late-equaliser-hands-javier-trejo-garay-improbable-comeback-victory)
- ESPN. (2018o). World Cup predictions Day 1: Russia vs. Saudi Arabia. Retrieved from [http://www.espn.com/sports/soccer/story/\\_/id/23759154/world-cup-predictions-day-1-russia-vs-saudi-arabia](http://www.espn.com/sports/soccer/story/_/id/23759154/world-cup-predictions-day-1-russia-vs-saudi-arabia)
- ESPN. (2019). NFL Expert Picks - Week 17. Retrieved February 28, 2019, from <https://www.google.com/search?q=nfl+week+17&oq=nfl+week+17&aqs=chrome..69i57j69i60j0l4.4629j1j7&sourceid=chrome&ie=UTF-8>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- FIFA. (2018). MEN'S RANKING - JUNE 2018. Retrieved June 15, 2018, from <https://www.fifa.com/fifa-world-ranking/ranking-table/men/rank/id11839/>
- Forrest, D., & Simmons, R. (2008). Sentiment in the betting market on Spanish football. *Applied Economics*, 40(1), 119–126. <https://doi.org/10.1080/00036840701522895>
- Galton, F. (1949). Presents Vox Populi ( The Wisdom of Crowds ). *Nature*, 75, 450–451. <https://doi.org/https://dx.doi.org/10.1002/clc.22367>
- Gambling Commission. (2018). Behaviour, awareness and attitudes. Retrieved February 26, 2019, from <https://www.gamblingcommission.gov.uk/news-action-and-statistics/Statistics-and-research/Statistics/Behaviour-awareness-and-attitudes.aspx>
- Gambling Commission. (2018a). Gambling key facts. Retrieved February 26, 2019, from <https://www.gamblingcommission.gov.uk/news-action-and-statistics/Statistics-and-research/Statistics/Gambling-key-facts.aspx>
- Gambling Commission. (2018b). Who we are and what we do. Retrieved February 26, 2019, from <https://www.gamblingcommission.gov.uk/PDF/survey-data/Who-we-are-and-what-we-do.pdf>
- Herzog, S. M., & Hertwig, R. (2011). The wisdom of ignorant crowds : Predicting sport outcomes by mere recognition. *Judgment and Decision Making*, 6(1), 58–72.
- Hickey, W. (2015). Betting Markets ( Mostly ) Nailed The Oscar Nominations. Retrieved October 30, 2018, from <https://fivethirtyeight.com/features/betting-markets-mostly-nailed-the-oscar-nominations/>
- Howe, J. (2006). The rise of crowdsourcing. Retrieved July 1, 2019, from <https://www.wired.com/2006/06/crowds/>
- Iowa Electronic Markets. (2018). What Is the IEM? - Iowa Electronic Markets. Retrieved October 30, 2018, from <https://iemweb.biz.uiowa.edu/media/summary.html>
- Lee, M. D., & Lee, M. N. (2017). The relationship between crowd majority and accuracy for binary decisions. *Judgment and Decision Making*, 12(4), 328–343.
- Levit, S. (2004). Why Are Gambling Markets Organised so Differently from Financial Markets? *The Economic Journal*, 114(495), 223–246.

- Mccrum-Gardner, E. (2008). Which is the correct statistical test to use ? *British Journal of Oral and Maxillofacial Surgery*, 46, 38–41.  
<https://doi.org/10.1016/j.bjoms.2007.09.002>
- Oddsportal.com. (2019). NBA Results & Historical Odds.
- oddsshark. (2018). world cup odds. Retrieved July 20, 2018, from  
<https://www.oddsshark.com/soccer/world-cup-odds>
- Oxford dictionaries. (2019). crowdsourcing definition. Retrieved July 1, 2019, from  
<https://en.oxforddictionaries.com/definition/crowdsourcing>
- Peeters, T. (2018). Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*, 34(1), 17–29.  
<https://doi.org/10.1016/j.ijforecast.2017.08.002>
- Simmons, J. P., Nelson, L. D., Galak, J., & Frederick, S. (2010). Intuitive Biases in Choice vs. Estimation: Implications for the Wisdom of Crowds. *Ssrn*, 38(June).  
<https://doi.org/10.2139/ssrn.1553935>
- StataCorp LLC. (2019). bitest — Binomial probability test. Retrieved from  
<https://www.stata.com/manuals13/rbitest.pdf>
- Wolfers, J., & Zitzewitz, J. (2004). PREDICTION MARKETS. *Biophysical Journal*, 87(5), 3088–3099.

## 9. Appendix.

### Appendix 1: Binomial test Football.

	N	Observed K	Expected K	Assumed p	Observed p
Instagram Poll	39	27	19.5	0.5	0.69231
Pr ( $k \geq 27$ )	0.011851	(one-sided test)			
Pr ( $k \leq 27$ )	0.995262	(one-sided test)			
Pr ( $k \leq 12$ or $k \geq 27$ )	0.023703	(two-sided test)			

### Appendix 2: Binomial test NBA full sample.

	N	Observed K	Expected K	Assumed p	Observed p
<b>Season</b>	62	39	31	0.5	0.62903
Pr ( $k \geq 39$ )	0.027949	(one-sided test)			
Pr ( $k \leq 39$ )	0.984992	(one-sided test)			
Pr ( $k \leq 23$ or $k \geq 39$ )	0.055897	(two-sided test)			

Appendix 3: Binomial test NBA playoffs.

	<b>N</b>	<b>Observed K</b>	<b>Expected K</b>	<b>Assumed p</b>	<b>Observed p</b>
<b>Playoffs</b>	27	23	13.5	0.5	0.85185
Pr (k>=23)	0.000155	(one-sided test)			
Pr (k<=23)	0.999975	(one-sided test)			
Pr (k<=4 or k>=23)	0.000311	(two-sided test)			

Appendix 4: Binomial test NBA full sample props.

<b>Season</b>	<b>N</b>	<b>Observed K</b>	<b>Expected K</b>	<b>Assumed p</b>	<b>Observed p</b>
	98	45	49	0.5	0.45918
Pr (k>=45)	0.818323	(one-sided test)			
Pr (k<=45)	0.239846	(one-sided test)			
Pr (k<=45 or k>=53)	0.4796692	(two-sided test)			

Appendix 5: Binomial test NBA playoffs props.

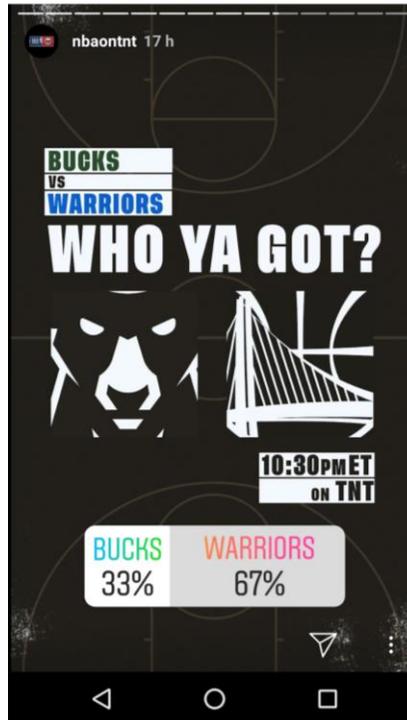
<b>Playoffs</b>	<b>N</b>	<b>Observed K</b>	<b>Expected K</b>	<b>Assumed p</b>	<b>Observed p</b>
	45	21	22.5	0.5	0.46667
Pr (k>=21)	0.724258	(one-sided test)			
Pr (k<=21)	0.382996	(one-sided test)			
Pr (k<=21 or k>=24)	0.765992	(two-sided test)			

Appendix 6: Example of a Instagram poll from @fifaworldcup



The image corresponds to the women's World Cup match between Germany and Spain played on June 9, 2019. The image is merely illustrative to show the surveys conducted by @fifaworldcup. It is important to mention that during the men's World Cup the polls had a neutral background and not a player from one of the teams, which can generate distortion in the results of the polls.

Appendix 7: Example of a poll from @nbaontnt



The image corresponds to the match played on November 8, 2018. The final score was Bucks 134 - Warriors 111. Resulting in an incorrect prediction by the crowd since the Warriors were the favourites in the poll. This poll is part of the sample.

Appendix 8: Example of a prep poll from @nbaontnt

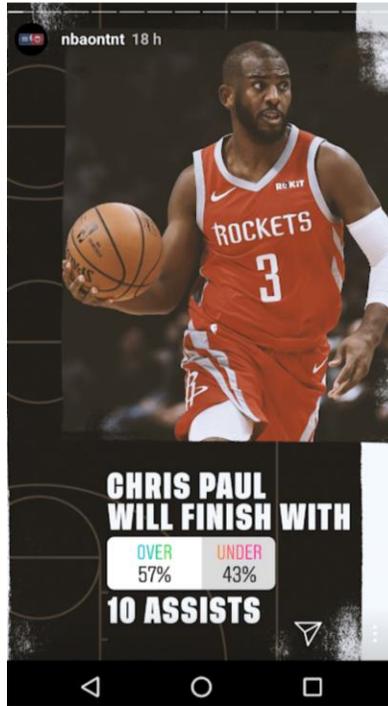


Image from November 15, 2018. Chris Paul finished the game with 7 assists, which were not enough for the crowd to be correct in his prediction that he was going to make over 10 assists. This poll is part of the sample.

Appendix 9: AUC NBA full sample regressions table for Instagram Polls and CARMELO model.

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]		
Instagram	0.5206243	-0.000275	0.0752411	0.3731544	0.6680942	(N)
				0.3658991	0.6674874	(P)
				0.3699495	0.6747968	(NC)
CARMELO	0.5120614	-0.00116	0.0783813	0.3584369	0.6656859	(N)
				0.3511706	0.6596543	(P)
				0.3597884	0.6681081	(NC)

Status:

- Voting percentage.
- Favouritism percentage.

Classifier:

- Successful predictions crowd.
- Successful predictions CARMELO model.

Appendix 10: AUC regression models NBA playoffs table for Instagram Polls and CARMELO model.

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]		
Instagram	0.7826087	-0.002595	0.1303103	0.5272052	1.038012	(N)
				0.4722222	0.9861111	(P)
				0.4615385	0.9722222	(NC)
CARMELO	0.5071429	0.0019694	0.1318163	0.2487876	0.7654981	(N)
				0.2571429	0.7727273	(P)
				0.2619048	0.7763158	(NC)

Status:

- Voting percentage.
- Favouritism percentage.

Classifier:

- Successful predictions crowd.
- Successful predictions CARMELO model.

Appendix 11: ROC regression models NBA propositions for NBA playoffs and NBA full sample.

AUC	Observed Coef.	Bias	Bootstrap Std. Err.	[95% Conf. Interval]		
Playoffs	0.4960317	0.0007585	0.090173	0.3192958	0.6727677	(N)
				0.3186291	0.6713095	(P)
				0.3100961	0.6659483	(NC)
Full sample	0.5509434	0.0010587	0.0594852	0.4343546	0.6675322	(N)
				0.4428217	0.669061	(P)
				0.4433034	0.669234	(NC)

NBA propositions regressions:

Status:

- Voting percentage playoffs.
- Voting percentage full sample.

Classifier:

- Successful predictions crowd playoffs sample.
- Successful predictions crowd full sample.