MATHEMATICAL AND STATISTICAL ANALYSIS OF EXCITEMENT SCORE IN PENALTY SHOOTOUTS

by
EZGİ GENÇ

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfilment of
the requirements for the degree of Master of Science

Sabancı University
December 2021

MATHEMATICAL AND STATISTICAL ANALYSIS OF EXCITEMENT SCORE IN PENALTY SHOOTOUTS

Approved by:

Date of Approval: December 17, 2021

ABSTRACT

MATHEMATICAL AND STATISTICAL ANALYSIS OF EXCITEMENT SCORE IN PENALTY SHOOTOUTS

EZGİ GENÇ

INDUSTRIAL ENGINEERING M.Sc. THESIS, DECEMBER 2021

Thesis Supervisor: Asst. Prof. Sinan Yıldırım

Keywords: excitement, penalty shootouts, machine learning, winning expectancy

Quantitative measurement of excitement during sports games is a field of study that has not been explored much. In this thesis, our aim was to find the expected excitement score of penalty shootouts mathematically. The excitement score formulations were generated based on the variation in the winning probabilities of teams after each penalty. Probability of success for each penalty originates from a Beta distribution in cases where players' scoring probabilities are unknown. Expectation-Maximization (EM) algorithm was utilized to find the parameter estimations of Beta distribution.

A survey was conducted to understand what makes shootouts exciting for the viewers and participants were asked to choose between two penalty shootout scenarios by examining scenario features at each question. Subsequently, the Bradley-Terry model was used to rank the scenario preferences of the viewers. This ranking was then used to make comparisons with the ranking of the scenarios obtained by using excitement scores. Predictive models were built using machine learning algorithms including Logistic Regression, Random Forest, AdaBoost Classifier and XGBoost to find feature importances. An alternative excitement score calculation was formed by taking the incremental excitement into consideration. Lastly, the excitement score was calculated for cases where scoring probabilities of teams vary at each round for a realistic approach. Discrete-time Markov chain process was used to find winning probabilities of each team. The results demonstrated that our excitement score calculations were successful in determining the least exciting shootouts. Moreover, features deemed as important by the viewers were also crucial mathematically.

ÖZET

PENALTI ATIŞLARINDA HEYECAN SKORUNUN MATEMATİKSEL VE İSTATİKSEL ANALİZİ

EZGİ GENÇ

ENDÜSTRİ MÜHENDİSLİĞİ YÜKSEK LİSANS TEZİ, ARALIK 2021

Tez Danışmanı: Dr. Sinan Yıldırım

Anahtar Kelimeler: heyecan, penaltı atışları, makine öğrenmesi, kazanma beklentisi

Spor müsabakaları sırasında duyulabilecek heyecanın nicel ölçümü çok fazla incelenmemiş bir araştırma alanıdır. Bu tezin amacı penaltı atışlarının beklenen heyecan skorunu matematiksel olarak bulmaktır. Heyecan skoru formülleri, her penaltı atışından sonra takımların kazanma olasılıklarındaki değişime göre oluşturulmuştur. Oyuncuların penaltı atma olasılığının bilinmediği durumlarda, her penaltının başarı olasılığı bir beta dağılımından gelmektedir. Beklenti-Maksimizasyonu algoritması bu beta dağılımının parametrelerini tahmin etme amaçlı kullanılmıştır.

Penaltı atışlarını izleyiciler için nelerin heyecanlı kıldığını anlayabilmek için bir anket tasarlanmış ve her soruda katılımcılardan iki penaltı atışı senaryosu arasından izlemeyi tercih edecekleri senaryoyu, senaryo özelliklerini de dikkate alarak seçmeleri istenmiştir. İzleyicilerin senaryo tercihlerini sıralamak için Bradley-Terry modeli kullanılmıştır. Bu sıralama, heyecan skorları kullanılarak elde edilen sıralama ile karşılaştırma yapmak için kullanılmıştır. Tahmine dayalı modeller, Lojistik Regresyon, Rassal Orman, AdaBoost sınıflandırıcı ve XGBoost gibi makine öğrenimi algoritmaları kullanılarak oluşturulmuştur. Artımlı heyecan dikkate alınarak alternatif heyecan skoru hesaplaması yapılmıştır. Son olarak, gerçekçi bir yaklaşım için takımların penaltı atma olasılıklarının her turda farklılık gösterdiği baz alınarak heyecan skoru hesaplanmıştır. Takımların kazanma olasılıklarını bulmak için ayrık zamanlı Markov zinciri kullanılmıştır. Sonuçlar, heyecan skoru hesaplamalarımızın en az heyecan verici atışları belirlemede başarılı olduğunu gösterdi. Ayrıca izleyiciler tarafından önemli görülen özellikler matematiksel olarak da önemli bulundu.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor Dr. Sinan Yıldırım. He continuously supported and guided me in every step of the way. I am very grateful for his patience, his immense knowledge and sincerity. I could not have asked for a better mentor.

I would also like to thank my jury members, Dr. Ezgi Karabulut Türkseven and Dr. Okan Örsan Özener, for their very valuable comments and for their time.

I am also very grateful to have such a supportive group of friends. Many thanks to my Sabancı grad pals, Irmak, Semih and Eren, who faced the same struggles and showed nothing but understanding throughout the journey. I would also like to express my gratitude and love for Emre, Yağmur, Arda, Ayça, Sezen, Hüs, Bora and Engin for their continuous support and companionship through the years. May we have many more occasions to celebrate in the future.

Finally, I am very thankful to my parents for their never-ending support and love. Without them, I would not be able to accomplish the things that I have always dreamed of. Thank you so much for standing by me and helping me, I am forever grateful.

To our beloved Deniz

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Sports games have a significant place in most people's daily lives. They provide a sense of community, help people to socialise by bringing them together and allow them to escape from their daily struggles. One of the most popular sports games in the world is football. It can bring excitement, sadness or joy into our lives. That is why, it may be useful to look for what excites us during a football match.

In this study, we focus on a special segment of football matches, the penalty shootouts. They can be the determinant part of a match; hence finding out which aspects of a shootout make it more exciting to the viewers can be used as a game refinement method in the future. The scope of this study can be explained in two parts. First is to find the excitement score of penalty shootouts mathematically and second is to compare mathematical results with the real viewer opinions by using survey data and statistical tools.

The remainder of this thesis is organized as follows. In Chapter 2, we provide a literature review for measuring emotions in various sports and analyse their formulations. In Chapter 3, we introduce the mathematical formulations of winning probability and expected excitement to calculate the excitement score for constant, varying and unknown scoring probabilities. We also provide a maximum likelihood estimation for Beta distribution with real-life data for cases where scoring probabilities of the players are unknown. In Chapter 4, we provide the results of a survey that we conducted to compare mathematical results with the real-life. We build a Bradley-Terry preference model by using survey data to make comparisons with the mathematical findings. In addition, we use machine learning applications and SHAP values to derive the feature importances and to understand which features of a penalty shootout affect the viewers' decision making process the most while selecting a penalty shootout to watch. Finally, in Chapter 5, we present the concluding remarks and possible future work for this thesis.

2. RELATED WORK

Understanding people's emotions and responses towards different circumstances has been an area of interest in the academic community. Quantitative techniques are used vigorously to evaluate human behaviour and feelings. However, a field of study that has not been examined much is the quantitative measure of viewer emotions in various sports competitions. Our study focuses on the quantitative measurement of excitement in penalty shootouts. We provide mathematical calculations derived from the variability of the winning expectancy to measure the excitement of a penalty shootout. Consequently, we can understand the motivation behind the viewer's preference to watch a game and use it to improve game rules and overall game entertainment.

In the literature, researchers have employed different strategies to improve game entertainment. One of the first studies in this field by Iida, Takeshita & Yoshimura (2003) developed the foundations of game refinement theory which is a game theory that focuses on improving the attractiveness of games. Following Iida, Takeshita and Yoshimura's research in 2003, there have been a number of studies that focused on improving the game entertainment by using game refinement theory. The first formal definition of game refinement theory was made by Iida, Takahara, Nagashima, Kajihara & Hashimoto (2004). In their paper, Iida, Takeshita and Yoshimura explained game refinement theory as a way to measure game uncertainty to understand the value of a game. Higher uncertainty would mean that accrued information on the game result over time by the viewer would be lower, thus game would be more exciting and fascinating. They believed that the gaming rules needed to be refined and optimized over time in order to keep the viewers' or players' attention. This process of evolution can be presented as the game refinement theory. Similarly, a study by Sutiono, Purwarianti & Iida (2014) explains the game refinement theory based on the concept of information of game outcome uncertainty. Another study by Nossal & Iida (2014) applies game refinement theory to the current scoring system in badminton and compares results with the old scoring system. In the old scoring system, also known as side-out, only the server side could score the point and if they lose the rally, no point was awarded to both sides, thus, causing an increase in the match times. In the new scoring system, serving side is not taken into consideration and the side which wins the rally scores the point regardless. Nossal and Iida's findings suggest that the new scoring system makes the game approximately 29% more interesting. A different study by Chetprayoon, Iida & Takahashi (2017) compared best-of-three and best-of-five set competitions in tennis by using game refinement theory and concluded that best-of-three set competitions were approximately 21% more entertaining for each tournament that they have explored. Junki, Rido & Hiroyuki (2014) used game refinement theory to compare three variants of volleyball. In time, volleyball's scoring system has changed to a 25-point rally point system for a better game understandability. Previously, it had different scoring systems such as 30-point rally system and 15-point side out scoring system. Comparison of three variants resulted with 25-point rally point system being the most exciting scoring system since it has the highest game-refinement value. In contrast to game refinement theory, there have been several other studies that focused on finding emotional scores of sports games rather than improving their game rules. While some of these studies focus on measuring the emotion factor over analysing the game videos by considering different aspects to understand highlights of a game, others propose quantitative models to calculate the excitement.

Hanjalic (2005) designed a video abstraction algorithm to keep the most interesting parts of a sports event. He generated an excitement time curve and described it as a function of the excitement level variety over the video frames caused by the stimulus that is represented by a feature. These features were later explained as the excitement components that mimicked the changes in a user's excitement. In order to achieve a more explanatory curve, Hanjalic only included the video segments that contained an event of a particular strength in the curve. This strength was determined by the total excitement level of an event. Another study in the same field was performed by Lee, Kim & Kim (2009). They combined video analysis with statistical modeling. Semantic video analysis was utilized to extract score information of basketball games. Following that, excitement of the shots was estimated from a statistical model created by the authors. As a result, excitement information was used to determine the highlights of the game.

As stated previously, some of the other studies developed quantitative models for excitement calculation. A study from Pollard (2016) concentrated on finding excitement of a point in a tennis game. He determined that excitement of a point in each score is associated with a pre-defined importance of the point at that score. Formulation of the excitement score, therefore, can be constructed as $p \mid qI \mid +$ q$\mid -pI \mid$ where p is player A's probability of winning a point with importance I and q is equal to $1-p$. In the framework of a football game, Vecer, Ichiba & Laudanovic (2007) examines the probabilistic excitement of a football game as follows

Excitement = TV(Probability of Team 1 Wins) + TV (Probability of Team 2 Wins)

where, for a probability function of time, $f(\cdot)$, $\mathrm{TV}(f)$ corresponds to total variation formulated as

$$\mathrm{TV}(f) = \lim_{\max_i |t_{i+1}-t_i| \to 0} \sum_i | f(t_{i+1}) - f(t_i) |$$

Vecer et al. (2007) stated that, as the variability of winning expectancy increases, excitement of that game also increases. In addition, according to their findings, people tend to watch games in which opposing teams have the same strength. They concluded their work with a comparison of theoretical results and real-life examples by using data from FIFA World Cup Soccer 2006. While it demonstrates the quantitative measure of excitement in a football game, Vecer et al. (2007) paper does not elaborate on the excitement of penalty shootouts. In our study, we will apply Vecer et al. (2007)'s principles regarding the relationship between variability of winning expectancy and excitement of a game to penalty shootouts. Furthermore, we will compare mathematical findings of our research with the statistical analysis of a survey that we have conducted.

3. EXCITEMENT SCORE CALCULATION

In this chapter, we introduce the mathematical formulations related to the expected excitement score of a penalty shootout sequence between two teams, team A and team B. The act of scoring a penalty shootout is discrete and teams act in turn. At each turn, a player from the team will take a single penalty with a certain scoring probability. This chapter explores three different cases for scoring probabilities: constant (i.e. equal), varying and unknown. A round is formed by the succession of player A's and player B's trials. Finally, we introduce the Graphical User Interface that we created to calculate the excitement scores of the penalty shootout scenarios automatically.

3.1 Winning Probability Formulations

Probability of winning is a term used for explaining the winning chance of a sports team at any point during the game. It can be based on the previous performances or the current success of the team. In this section, we will focus on calculating the winning probabilities of opposing teams during their penalty shootouts at certain times.

A penalty shootout can be considered as a sequence of penalty kicks. Before 1970, laws of association football did not have a certain method used for a drawn match. The first association football tournament, known as FA Cup, used extra time, coin tosses and replays to decide the winner. Penalty shootouts were introduced officially to football after they were proposed by Joseph (also known as Yosef) Dagan to Fédération Internationale de Football Association (FIFA) in 1970. After their implementation to the game, the winner of a single-elimination tournament in which the opposing teams were drawing after the regular 90 minutes playing time and extra 30 minutes would be determined by penalty shootouts.

Some of the most important rules for the penalty shootouts that were determined by Laws of the Game are as follows:

1) Coin toss will be used to determine which team will use the penalty kick first.

2) Teams will select five different players to shoot penalties from the players that were on the pitch at the end of the match. Each penalty shootout must be performed by a different player.

3) Teams will take turns until each has taken five kicks. If one of the teams has scored more successful kicks than the other could possibly reach with all of its remaining kicks, the penalty shootouts end despite the number of kicks remaining.

4) After these five rounds of kicks the teams have scored, if they have an equal number of goals, they will perform sequential kicks until one of the teams has one more goal than the other with the same number of penalty kick attempts. This process is also known as sudden death.

After the revision of penalty shootout rules, the winning probabilities of opposing teams can be calculated, starting with the explanation of some of the notations that will be used throughout this chapter.

Table 3.1 General Notation of Mathematical Formulations

T	Total number of regular penalty rounds
t	Total number of penalties taken by teams up to current round
t_1	Penalties taken by Team A up to round t
t_2	Penalties taken by Team B up to round t
x	Current score of team A
y	Current score of team B
S_{t_1,t_2}	Score when team A used t_1 penalties and team B used t_2 penalties
$\mathbf{a}$	Scoring probabilities of team A
$\mathbf{b}$	Scoring probabilities of team B

Let S_{t_1,t_2} be the score after teams A and B have just taken $t_1 \geq 1$ and $t_2 \geq 1$ penalties, respectively. Given scoring probabilities $\mathbf{a} = (a_1, a_2, \ldots)$ and $\mathbf{b} = (b_1, b_2, \ldots)$ for teams A and B, respectively, we define

$$\begin{aligned} A^{\mathbf{a},\mathbf{b}}_{t_1,t_2}(x,y) &= P(\text{Team A wins}|S_{t_1,t_2} = (x,y)) \\ B^{\mathbf{a},\mathbf{b}}_{t_1,t_2}(x,y) &= P(\text{Team B wins} \mid S_{t_1,t_2} = (x,y)), \end{aligned} \qquad 1 \leq t_1, t_2 \leq T. \tag{3.1}$$

where the superscript denotes the scoring probabilities of team A and team B. Since summation of teams winning probabilities will be equal to 1, it can be observed that

$$B^{\mathbf{a},\mathbf{b}}_{t_1,t_2}(x,y) = 1 - A^{\mathbf{a},\mathbf{b}}_{t_1,t_2}(x,y) \tag{3.2}$$

Consequently, the winning probabilities of team A and team B at the beginning of the game are given by $A^{\mathbf{a},\mathbf{b}}_{0,0}(0,0)$ and $B^{\mathbf{a},\mathbf{b}}_{0,0}(0,0)$.

Remark 3.1. *It is useful at this point to introduce the notation for winning probabilities from now on.*

Unless necessary, we will drop the superscript a,b from the notation if it is clear from the context that the scoring probabilities are the default probabilities a and b. The notation in (3.1) *will be useful when we want to calculate the winning probabilities if a pair of scoring probabilities are swapped in a or b. To be concrete, for a sequence $\mathbf{v} = (v_1, v_2, \ldots)$, we let $\sigma_t(\mathbf{v}) = (v_t, v_2, \ldots, v_{t-1}, v_1, v_{t+1}, \ldots)$, the sequence obtained by swapping v_1 and v_t in $\mathbf{v}$. It will be revealed soon that for expected excitement calculations, we will need winning probabilities such as*

$$A^{\sigma_t(\mathbf{a}),\mathbf{b}}_{1,0}(1,0), \quad A^{\sigma_t(\mathbf{a}),\mathbf{b}}_{1,0}(0,0), \quad B^{\mathbf{a},\sigma_t(\mathbf{b})}_{0,1}(0,1), \quad B^{\mathbf{a},\sigma_t(\mathbf{b})}_{0,1}(0,0)$$

When all a_t's and b_t's are the same, we will again drop the superscript from the notation.

The following intermediate results, presented as Lemmas 3.1, 3.2, 3.3 and will be helpful for the subsequent derivations for the expected excitement.

Lemma 3.1. *By using the notation in this part, we have the following trivial inequalities of winning probabilities*

1. $A_{t,t-1}(x,y) \leq A_{t-1,t-1}(x,y)$,
2. $A_{t,t-1}(x+1,y) \geq A_{t-1,t-1}(x,y)$,
3. $A_{t,t}(x,y) \geq A_{t,t-1}(x,y)$,
4. $A_{t,t-1}(x,y) \leq A_{t,t}(x,y+1)$.

First inequality demonstrates that the winning probability of a team after an unsuccessful penalty shootout cannot be greater than its previous winning probability. In contrast, second inequality demonstrates that the winning probability of a team after a successful penalty shootout must be greater than or equal to its previous winning probability.

Third and fourth inequalities explain the winning probability of team A after opposing team's penalty shootout. Third inequality applies for the case when opposing team misses its penalty shootout while fourth inequality is for the case when the opposing team scores its penalty shootout.

In order to define upcoming Lemmas and the incremental expected excitement in this section properly, we have to show scoring probabilities at a certain time. Let us define

$$p_{t_1,t_2}(x,y) = P(S_{t_1,t_2} = (x,y))$$

as the probability of having the score (x,y) after using t_1 and t_2 penalty kicks, respectively.

Lemma 3.2. *For any $1 \leq t_1, t_2 \leq T$, we have*

$$A_{0,0}(0,0) = \sum_{x,y} p_{t_1,t_2}(x,y) A_{t_1,t_2}(x,y), \quad B_{0,0}(0,0) = \sum_{x,y} p_{t_1,t_2}(x,y) B_{t_1,t_2}(x,y).$$

Lemma 3.3. *For any $t \geq 1$, have*

$$\sum_{x,y} p_{t-1,t-1}(x,y) A_{t,t-1}(x+1,y) = A_{1,0}^{\sigma_t(\mathbf{a}),\mathbf{b}}(1,0)$$
$$\sum_{x,y} p_{t-1,t-1}(x,y) A_{t,t-1}(x,y) = A_{1,0}^{\sigma_t(\mathbf{a}),\mathbf{b}}(0,0)$$
$$\sum_{x,y} p_{t,t-1}(x,y) B_{t,t}(x,y+1) = B_{0,1}^{\mathbf{a},\sigma_t(\mathbf{b})}(0,1)$$
$$\sum_{x,y} p_{t,t-1}(x,y) B_{t,t}(x,y) = B_{0,1}^{\mathbf{a},\sigma_t(\mathbf{b})}(0,0).$$

3.2 Expected Excitement Formulations

In this section, we define a mathematical formula for the expected excitement of a penalty shootout, adopting the approach of Vecer et al. (2007). Let $S = (S_{1,0}, S_{1,1}, S_{2,1}, S_{2,2}, \ldots)$ denote a random game with where the elements of the sequence are the scores. According to Vecer et al. (2007), the total excitement given a random game is as follows

$$\mathcal{E}(S) = \sum_{t=1}^{\infty} \{|A_{t,t-1}(S_{t,t-1}) - A_{t-1,t-1}(S_{t-1,t-1})| + |A_{t,t}(S_{t,t}) - A_{t,t-1}(S_{t,t-1})|\} \quad (3.3)$$

whose realization given $S = s$ can be written as

$$\mathcal{E}(s) = \sum_{t=1}^{\infty} \{|A_{t,t-1}(s_{t,t-1}) - A_{t-1,t-1}(s_{t-1,t-1})| + |A_{t,t}(s_{t,t}) - A_{t,t-1}(s_{t,t-1})|\} \quad (3.4)$$

Table 3.2 Notations for the Mathematical Formulations of the Expected Excitement

$S = (S_{1,0}, S_{1,1}, S_{2,1}, S_{2,2}, \ldots)$	Random game with random sequence of scores
$s = (s_{1,0}, s_{1,1}, s_{2,1}, s_{2,2}, \ldots)$	Realization of S

Equations in (3.3) and (3.4) depict the summation of the variation in winning probability of team A after their t^{th} penalty shootout and the variation in winning probability of team A after team B's t^{th} penalty shootout. They represent the change in team A's winning probability per round. These equations could have been written for team B as well, however, penalty shootouts will result with a winner in each case. Therefore, multiplying these equations by 2 would be sufficient.

The expected excitement will be expectation of $\mathcal{E}(S)$, which can be written as

$$\begin{aligned}\mathcal{E} &= \mathbb{E}[\mathcal{E}(S)] \\ &= \sum_{t=1}^{\infty} \{\mathbb{E}[|A_{t,t-1}(S_{t,t-1}) - A_{t-1,t-1}(S_{t-1,t-1})|] + \mathbb{E}[|A_{t,t}(S_{t,t}) - A_{t,t-1}(S_{t,t-1})|]\} \\ &= \sum_{t=1}^{\infty} E_{t,1} + E_{t,2}. \qquad (3.5)\end{aligned}$$

where $E_{t,1}$ and $E_{t,2}$ in (3.5) are defined as the first and second expectation terms in the sum, called *incremental expected excitements* at round t for Team A's and Team B's penalty takes, respectively.

Let $f_{A,t}(x',y'|x,y)$ (resp. $f_{B,t}(x',y'|x,y)$) be the transition probability for the score from (x,y) to (x',y') when team A (resp. B) takes a penalty at round t. Thus, incremental expected excitements at time t can be written as

$$\begin{aligned}E_{t,1} &= \sum_{x,y,x',y'} p_{t-1,t-1}(x,y) f_{A,t}(x',y'|x,y) |A_{t-1,t-1}(x,y) - A_{t,t-1}(x',y')| \\ E_{t,2} &= \sum_{x,y,x',y'} p_{t,t-1}(x,y) f_{B,t}(x',y'|x,y) |B_{t,t-1}(x,y) - B_{t,t}(x',y')|\end{aligned}$$

which shows the expected absolute difference in winning probabilities. (x,y) represents the score before a certain penalty shootout and (x',y') represents the score after a certain penalty shootout. Thus, the total expected excitement can be written as

$$\sum_{t=1}^{\infty} (E_{t,1} + E_{t,2}). \qquad (3.6)$$

We find the total excitement in two steps. First, for regular rounds we derive the incremental expected excitement and summing those we will get $\sum_{t=1}^{T}(E_{t,1}+E_{t,2})$. Secondly, we derive the probability of a draw after T rounds and the expected excitement for the tie-break period, which, when multiplied together, yield the rest of the sum, $\sum_{t=T+1}^{\infty}(E_{t,1}+E_{t,2})$.

3.2.1 Incremental Expected Excitement for Regular Rounds

For the regular rounds, we have the result, whose proof is based on Lemmas 3.1, 3.2, and 3.3.

Proposition 3.1. *For the regular rounds, i.e., for $1 \le t \le T$, the incremental excitements for round t can be written as*

$$\begin{aligned} E_{t,1} &= a_t A_{1,0}^{\sigma_t(a),b}(1,0) - (1-a_t)A_{1,0}^{\sigma_t(a),b}(0,0) + (1-2a_t)A_{0,0}(0,0) \\ E_{t,2} &= b_t B_{0,1}^{a,\sigma_t(b)}(0,1) - (1-b_t)B_{0,1}^{a,\sigma_t(b)}(0,0) + (1-2b_t)B_{0,0}(0,0) \end{aligned} \tag{3.7}$$

Proof. For the regular rounds, i.e., for $1 \le t \le T$, we have

$$\begin{aligned} E_{t,1} &= \sum_{x,y} p_{t-1,t-1}(x,y)\,[a_t \mid A_{t-1,t-1}(x,y) - A_{t,t-1}(x+1,y) \mid \\ &\quad + (1-a_t) \mid A_{t-1,t-1}(x,y) - A_{t,t-1}(x,y) \mid] \\ &= \sum_{x,y} p_{t-1,t-1}(x,y)\,[a_t\{A_{t,t-1}(x+1,y) - A_{t-1,t-1}(x,y)\} \\ &\quad + (1-a_t)\{A_{t-1,t-1}(x,y) - A_{t,t-1}(x,y)\}]\,. \end{aligned}$$

The absolute values can be resolved by using the trivial inequalities in Lemma 3.1 and Lemma 3.2. In addition, we use first two equations in Lemma 3.3 to rewrite $E_{t,1}$ as follows

$$\begin{aligned} E_{t,1} &= a_t \sum_{x,y} p_{t-1,t-1}(x,y)A_{t,t-1}(x+1,y) - a_t \sum_{x,y} p_{t-1,t-1}(x,y)A_{t-1,t-1}(x,y) \\ &\quad + (1-a_t)\sum_{x,y} p_{t-1,t-1}(x,y)A_{t-1,t-1}(x,y) - (1-a_t)\sum_{x,y} p_{t-1,t-1}(x,y)A_{t,t-1}(x,y) \\ &= a_t A_{1,0}^{\sigma_t(a),b}(1,0) - (1-a_t)A_{1,0}^{\sigma_t(a),b}(0,0) + (1-2a_t)A_{0,0}(0,0) \end{aligned}$$

Similarly, for the second part of a round in which B uses their penalty shootout,

$E_{t,2}$ can be written as

$$\begin{aligned}
E_{t,2} &= \sum_{x,y} p_{t,t-1}(x,y)\left[b_t|A_{t,t}(x,y+1)-A_{t,t-1}(x,y)|+(1-b_t)|A_{t,t}(x,y)-A_{t,t-1}(x,y)|\right] \\
&= \sum_{x,y} p_{t,t-1}(x,y)\left[b_t\{A_{t,t-1}(x,y)-A_{t,t}(x,y+1)\}+(1-b_t)\{A_{t,t}(x,y)-A_{t,t-1}(x,y)\}\right] \\
&= b_t\sum_{x,y} p_{t,t-1}(x,y)A_{t,t-1}(x,y)-b_t\sum_{x,y} p_{t,t-1}(x,y)A_{t,t}(x,y+1) \\
&+(1-b_t)\sum_{x,y} p_{t,t-1}(x,y)A_{t,t}(x,y)-(1-b_t)\sum_{x,y} p_{t,t-1}(x,y)A_{t,t-1}(x,y) \\
&= (2b_t-1)A_{0,0}(0,0)+(1-2b_t)+b_tB_{0,1}^{a,\sigma_t(b)}(0,1)-(1-b_t)B_{0,1}^{a,\sigma_t(b)}(0,0) \\
&= b_tB_{0,1}^{a,\sigma_t(b)}(0,1)-(1-b_t)B_{0,1}^{a,\sigma_t(b)}(0,0)+(1-2b_t)B_{0,0}(0,0).
\end{aligned}$$

Hence we conclude. □

3.2.2 Expected Excitement of the Tie-break

The tie-break period of a penalty shootout is played if teams have a draw at the end of the T rounds. In the tie-break, the game continues until either team is ahead at the end of a round. Let $\mathcal{E}_0$ be the excitement of a tie-break game, and

$$D = P(\{S_{T,T} = (x,x) : x \in \{0,\dots,T\})$$

be the probability of a draw after. Then, the total expected excitement of a penalty shootout can be written as

$$\mathcal{E} = \sum_{t=1}^{T}(E_{t,1}+E_{t,2})+D\mathcal{E}_0 \tag{3.8}$$

In the most general case, calculating $\mathcal{E}_0$ requires an infinite sequence of scoring probabilities, since the tie-break game continues indefinitely. To simplify, let us assume that the scoring probabilities of team A and B in each round of the tie-break are constants and given as a and b. Under this assumption, we can derive a closed form expression for $\mathcal{E}_0$.

Proposition 3.2. *The total excitement of the tie-break with scoring probabilities a and b is given by*

$$\mathcal{E}_0 = \frac{4ab(1-a)(1-b)}{(a+b-2ab)}. \tag{3.9}$$

Proof. The expected excitation of a tie-break game can be written as

$$\mathcal{E}_0 = \sum_{i=0}^{\infty} p_d^i E_r \tag{3.10}$$

where p_d is the probability of a draw in a round, given by

$$p_d = ab + (1-a)(1-b), \tag{3.11}$$

which corresponds to either both teams' successfully scoring the penalty shootout or missing the penalty shootout at the same round d. The excitement per round, E_r is given by

$$\begin{aligned}
E_r =& [a|(1-b) + bp_A - p_A| + (1-a)|(1-b)p_A - p_A| \\
& + ab|p_A - ((1-b) + bp_A)| + a(1-b)|1 - ((1-b) + bp_A)| \\
& + (1-a)b|0 - (1-b)p_A| + (1-a)(1-b)|p_A - (1-b)p_A|] \\
=& [a(1-b)(1-p_A) + (1-a)bp_A + 2ab(1-p_A)(1-b) \\
& + 2(1-a)(1-b)bp_A]
\end{aligned}$$

where p_A corresponds to the probability of team A winning the penalty shootouts eventually, given by

$$p_A = \sum_{i=0}^{\infty} p_d^i P(\text{A wins in a round}) = \frac{a(1-b)}{1-p_d} = \frac{(1-b)a}{a+b-2ab} \tag{3.12}$$

Combining the above, we have the desired result. (See the section Calculations for (3.9) under Appendix A for a detailed derivation.) □

3.3 Formulas for Equal Scoring Probabilities

For this section, we will consider the scoring probabilities at each round as constants and calculate the conditional expected excitement of a penalty shootout sequence.

It is important to note that $\mathcal{E}$ is invariant under permutations of $a_{1:T}$ and $b_{1:T}$. Additionally, if $a_t = a$ and $b_t = b$ for all $t \geq 1$, we have

$$\mathcal{E} = T(E_{1,1} + E_{1,2}) + DE_0 \tag{3.13}$$

where D can be simply found as $\sum_{i=0}^{T} \text{Binom}(i;T,a)\text{Binom}(i;T,b)$.

3.3.1 Conditional Excitement Formulations

In this section, we focus on calculating the remaining expected excitement of a game given its score at a given time. We call this quantity the *conditional expected excitement*. Calculating the conditional expected excitement for a game given its score at a time can provide useful insights in determining the most exciting aspects of the shootouts. In fact, we use the derived calculations in this section in our experimental work in Chapter 4.

For this section, we assume that the scoring probabilities of teams are constant at each round, denoted by a and b. The more general case can also be handled, although with increased complexity of calculations.

Suppose that in a game with T regular rounds, the current time is (t_1, t_2) and the score is (x, y). The conditional expected excitement given $\mathcal{S} = \{S_{t_1,t_2} = (x,y)\}$ can be shown to be

$$\mathcal{E}(\mathcal{S}) = \max\{0, T - t_1\} F(\mathcal{S}) + \max\{0, T - t_2\} G(\mathcal{S}) + D(\mathcal{S})\mathcal{E}_0 \tag{3.14}$$

where

$$\begin{aligned} F(\mathcal{S}) &= aA_{t_1+1,t_2}(x+1,y) - (1-a)A_{t_1+1,t_2}(x,y) + (1-2a)A_{t_1,t_2}(x,y), \\ G(\mathcal{S}) &= bB_{t_1,t_2+1}(x,y+1) - (1-b)B_{t_1,t_2+1}(x,y) + (1-2b)B_{t_1,t_2}(x,y). \end{aligned} \tag{3.15}$$

and $D(\mathcal{S})$ is the draw probability after $\max\{T, \max\{t_1, t_2\}\}$ rounds.

Although for $t_1, t_t \leq T$, the interpretation of the above formula is clear; the cases that correspond to a time in the tie-break needs elaboration.

- If $t_1 = t_2 \geq T$, that means that the game is in a tie-break stage and teams have taken equal number of penalty shootouts, the remaining part of the shootouts is as exciting as any tie-break game, in which excitement is denoted by $\mathcal{E}_0$.
- If $t_1 = t_2 + 1 \geq T$, that means that the first team have taken one more penalty than the second team during the overtime, and score is (x, y) with $x - y = s \in \{0, 1\}$, it can be considered as a game where $T = 1$ and the score is $(s, 0)$ at time $t_1 = 1$ and $t_2 = 0$.

3.4 Calculating the Winning Probabilities

It is clear from the above analysis that, in order to calculate expected excitements, we need to be able to calculate winning probabilities of several types. A general recipe for that is to model the game as a discrete time Markov chain with a discrete state-space and utilize this Markov chain for the calculation of the score probabilities at the rounds (which are the time steps of the chain), and deduce the winning probabilities from the score probabilities.

In this section, we introduce discrete time discrete Markov chains in general and show their application for our work.

3.4.1 Discrete-time Markov Chains

Discrete-time Markov chains (DTMCs) are stochastic processes that consist of a sequence of random variables. A DTMC is characterized by a discrete set of states visited by the process at a discrete set of times. The distinguishing property of a Markov chain is the conditional independence of the next state on the entire past given the current state, which is also known as the one-step memory property.

Assume that $\{X_n : n = 0, 1, \dots\}$ is a discrete-time stochastic process with a discrete state space $S = \{1, \dots, s\}$. The joint probability model of stochastic process will be represented as

$$P(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)$$

for $n = 0, 1, \dots$ and $x_0, x_1, \dots, x_n \in S$. This process is a homogeneous Markov chain if for all times $n \geq 0$ and all states $x_0, \dots i, j \in S$,

$$\begin{aligned} P(X_{n+1} = j \mid X_0 = x_0, X_1 = x_1, \dots X_n = i) &= P(X_{n+1} = j \mid X_n = i) \\ &= p_{ij} \end{aligned}$$

where p_{ij} is the one-step transition probability and denotes the probability of moving from state i to state j in one-step.

We could represent this chain by a $s \times s$ square matrix M with the elements of p_{ij} where $i, j \in S$. This square matrix is called the transition matrix and each row will sum up to one due to the fact that leaving state i, the chain must move to one of the states in $j \in S$. So, for all $i \in S$

$$\sum_{j \in S} p_{ij} = 1.$$

To summarize, the future sequence of the chain i.e. $\{X_{n+1}, X_{n+2}, \ldots\}$ only depends on X_n, that is, it is independent of any past states, namely $X_0, \ldots, X_{n-1}$.

For our study, DTMC could be used for the score, which is a bivariate random quantity. However, to calculate the winning probabilities, it suffices to track the probabilities of the score difference, which is univariate quantity and easier to use. Therefore, we consider a DTMC for the goal difference over the rounds. DTMC is suitable for this process since each player's success in scoring can be considered independent of the other players.

For example, suppose we have a penalty shootout with $T = 5$ rounds. Then the state space is $\{-5, -4, -3, -2, -1, 0, +1, +2, +3, +4, +5\}$. The state "$-5$" stands for the case where the score difference between team A and team B is five goals and team B is ahead of team A. The state "+3" state shows that the score difference between team A and team B is three goals and team A is ahead of team B. Actually, according to the rules of penalty shootouts, maximum score difference allowed between two teams is $\lceil \frac{T+1}{2} \rceil$ goals. Even though a difference of 4 or 5 goals is not possible, since there are a couple of rare cases that we need to be aware of to represent all possibilities clearly, we will not make any arrangements on the state space. If we were to rearrange the state space with the maximum difference, we would miss the winning probability of team B in a case where $T = 5$, current score is 3-0 and teams A and B have used 3 and 2 penalties, respectively. We display the exemplary state diagrams for a penalty shootout sequence with $T = 3$ rounds in Figure 3.1. For this case, we have 7 different states. State diagrams were drawn under the assumption of constant scoring probabilities, a and b, respectively.

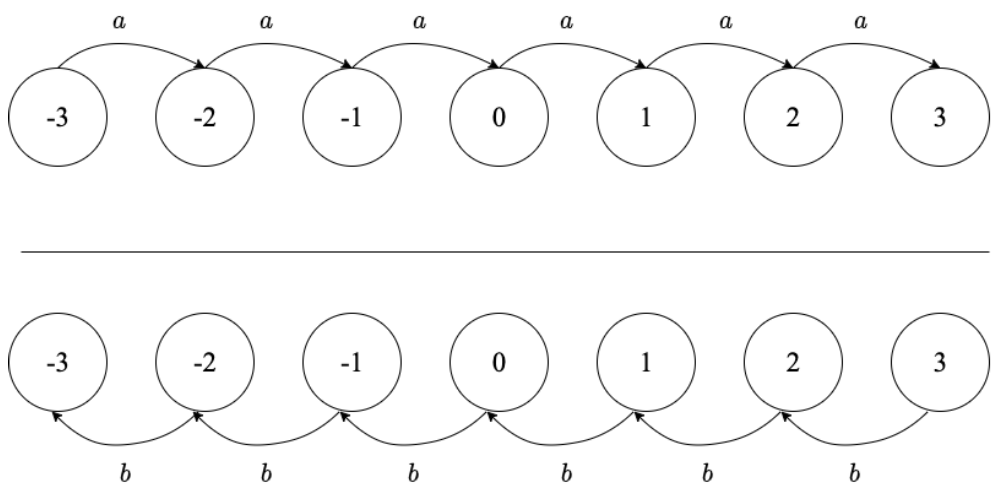

Figure 3.1 State diagram for a single round. Top: Team A's turn, Bottom: Team B's turn

The restriction of equal scoring probabilities can be relaxed, leading to a non-homogeneous Markov chain where the transition probabilities depend on the time step n, i.e.,

$$p_{ij}^{(n)} = P\left(X_n = j \mid X_{n-1} = i\right),$$

that is, $p_{ij}^{(n)}$ is the probability of moving from state i to state j at time step n.

A non-homogeneous DTMC is relevant to this work. Since the player to score the penalty will be different for each round, the probability of scoring the penalty may change. This suggests that the probability of moving from a score state i to score state j will be different at each step.

In the following, we will show the necessary derivations to calculate the expected excitement of a penalty shootout when the scoring probabilities vary across rounds. The derivations will reveal that the calculations depend on being able to calculate winning probabilities at any point of the penalty shootouts. To calculate those winning probabilities, the DTMC for the score difference will be utilized.

In Algorithm 1, we provide the general algorithm to calculate the eventual winning probabilities for Team A and Team B, as well as the drawing probability after T regular rounds, given a score (x, y) at time (t_1, t_2), meaning that $S_{t_1,t_2} = (x, y)$.

Algorithm 1 Conditional winning probability after T rounds given $S_{t_1,t_2} = (x,y)$

Input: Scoring probabilities $a_1,\dots,a_T$; $b_1,\dots,b_T$; tie-break scoring probabilities a_0, b_0; regular rounds T; current score $\mathcal{S} = \{S_{t_1,t_2} = (x,y)\}$.

Output: Eventual winning probabilities $A_{t_1,t_2}(x,y)$, $B_{t_1,t_2}(x,y)$; draw probability after T rounds $D(\mathcal{S})$.

Set $s = 2T+1$.

Set current difference $d = x-y$, and

$$\pi(k) = \begin{cases} 1 & k = d+T+1 \\ 0 & \text{else} \end{cases}, \quad k = 1,\dots,s.$$

for $t = 1,2,\dots,T$ **do**

if $t > t_1$ **then**

Set the $s \times s$ matrix M_A such that $M_A(i,j) = \begin{cases} a_t & j = i+1 \\ 1-a_t & 1 < j = i < s \\ 1 & j = i = 1 \text{ or } j = i = s \\ 0 & \text{otherwise} \end{cases}$

Update $\pi \leftarrow \pi M_A$

end if

if $t > t_2$ **then**

Set the $s \times s$ matrix M_B such that $M_B(i,j) = \begin{cases} b_t & i = j+1 \\ 1-b_t & 1 < j = i < s \\ 1 & j = i = 1 \text{ or } j = i = s \\ 0 & \text{otherwise} \end{cases}$

Update $\pi \leftarrow \pi M_B$

end if

end for

$A_{t_1,t_2}(x,y) = \{\pi(T+2)+\ldots+\pi(s)\} + \pi(T+1)\frac{a_0(1-b_0)}{a_0+b_0-2a_0b_0}$.

$B_{t_1,t_2}(x,y) = \{\pi(1)+\ldots+\pi(T)\} + \pi(T+1)\frac{b_0(1-a_0)}{a_0+b_0-2a_0b_0}$.

$D(\mathcal{S}) = \pi(T+1)$.

return $A_{t_1,t_2}(x,y)$, $B_{t_1,t_2}(x,y)$, and $D(\mathcal{S})$

Example 3.1. *In this example, we show the steps of Algorithm 1 for a simple case. Consider the game with $T = 5$ regular rounds whose scoring probabilities are given in Table 3.3. Assume that a_0 and b_0 are the averages of $a_1,\dots,a_5$ and $b_1,\dots,b_5$, respectively. That is, i.e. $a_0 = \frac{1}{T}\sum_{t=1}^{T} a_t$ and $b_0 = \frac{1}{T}\sum_{t=1}^{T} b_t$. For our example, these*

Table 3.3 DTMC Example with $T = 5$ Rounds

Symbol	Explanation	Value
T	Total number of rounds	5
$a_1,\dots,a_5$	scoring probabilities of team A players	[0.60 0.70 0.55 0.80 0.45]
$b_1,\dots,b_5$	scoring probabilities of team B players	[0.90 0.65 0.70 0.50 0.60]

are

$$a_0 = \frac{0.6+0.7+0.55+0.8+0.45}{5} = 0.62, \quad b_0 \ = \frac{0.9+0.65+0.7+0.5+0.6}{5} = 0.67.$$

Our goal is to calculate $A_{0,0}(0,0)$, which is the winning probability of team A at the beginning of a penalty shootout.

Total number of states for T rounds is $s = 2T+1$, which is $s = 11$ for $T = 5$. The initial probability distribution for the score difference is $\pi_{0,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, since $S_{0,0} = (0,0)$.

Since the scoring probability of Team A at the first round is 0.60, transition probability matrix for the score difference at Team A's turn will be

$$M_{A,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

. Multiplying $\pi_{0,0}$ by M_A, we have the probability distribution of the score difference after team A's penalty kick at the first round as

$$\pi_{1,0} = \pi_{0,0} M_{A,1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Next, we can construct the transition probability matrix of for the score difference *at Team B's turn in the first round as*

$$M_{B,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.9 & 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

. The probability distribution of the score difference after team B's penalty kick at the first round can be found in a similar manner, as

$$\pi_{1,1} = \pi_{1,0} M_{B,1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.36 & 0.58 & 0.06 & 0 & 0 & 0 & 0 \end{bmatrix}$$

By following the same steps for the next four rounds, the winning probability after $T = 5$ *regular rounds is calculated as* $\pi_{5,5}(7) + \pi_{5,5}(8) + \pi_{5,5}(9) + \pi_{5,5}(10) + \pi_{5,5}(11) = 0.3028$*. However, this is not the overall winning probability* $A_{0,0}(0,0)$*, since the game can continue to tie-break rounds. The winning probability for Team A in tie-break rounds,* p_A *is given in* (3.12)*. Applying it to our example, we have*

$$p_A = \frac{0.33 \times 0.62}{0.62 + 0.67 - 2 \times 0.62 \times 0.67} = 0.4455.$$

Therefore, we have

$$A_{0,0}(0,0) = (\pi_{5,5}(7) + \pi_{5,5}(8) + \pi_{5,5}(9) + \pi_{5,5}(10) + \pi_{5,5}(11)) + \pi_{5,5}(6) p_A = 0.4205.$$

Similar steps can be followed to calculate $A_{1,0}^{\sigma_t(\mathbf{a}),\mathbf{b}}(1,0)$ *and* $A_{1,0}^{\sigma_t(\mathbf{a}),\mathbf{b}}(0,0)$*, the probabilities needed to calculate the incremental excitement* $E_{t,1}$*, for any* $t = 1, \ldots, T$*. For each of those, a different initial distribution should be set given the score and the transition matrices should be applied as many as the remaining trials for each team. For example,* $A_{1,0}^{\sigma_t(\mathbf{a}),\mathbf{b}}(1,0)$ *is the winning probability of Team A given that the probability sequence for Team A is* $\sigma_t(\mathbf{a})$ *and Team A missed their first penalty. The initial distribution for this calculation is* $\pi_{1,0} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$*. This initial distribution should be multiplied by* $M_{B,1} M_{A,2} M_{B,2} \ldots M_{A,5} M_{B,5}$ *where the transition matrices* $M_{A,2}, \ldots, M_{A,5}$ *are constructed from the probabilities in*

$\mathbf{a}^{-t} = (a_2,\dots,a_{t-1},a_1,a_{t+1},\dots,a_T)$ *and* $M_{B,1},\dots,M_{B,5}$ *are constructed from the probabilities in* $\mathbf{b}$*. By following those steps for* $t = 1$*, we have*

$$A^{\mathbf{a},\mathbf{b}}_{1,0}(1,0) = 0.3973 + 0.4455 \times 0.2831 = 0.5234,$$
$$A^{\mathbf{a},\mathbf{b}}_{1,0}(0,0) = 0.1609 + 0.4455 \times 0.2360 = 0.2660,$$

Therefore, the incremental excitement is calculated as

$$E_{1,1} = 0.6 \times 0.5234 - 0.4 \times 0.2660 - 0.2 \times 0.4205 = 0.1235.$$

Just as the winning probabilities for Team A, we can also calculate $B_{0,0}(0,0)$*,* $B^{\mathbf{a},\sigma_t(\mathbf{b})}_{0,1}(0,0)$ *and* $B^{\mathbf{a},\sigma_t(\mathbf{b})}_{0,1}(0,1)$*, the probabilities needed to calculate the incremental excitements* $E_{t,2}$*, for* $t = 1,\dots,T$*, by following similar steps.*

For the values in our example, $B^{\mathbf{a},\sigma_t(\mathbf{b})}_{0,0}(0,0)$*,* $B^{\mathbf{a},\sigma_t(\mathbf{b})}_{0,1}(0,1)$ *and* $B^{\mathbf{a},\sigma_t(\mathbf{b})}_{0,1}(0,0)$ *can be calculated as* 0.5794*,* 0.6053 *and* 0.3460*, respectively. Thus,*

$$E_{1,2} = 0.9 \times 0.6053 + 0.1 \times 0.3460 - 0.8 \times 0.5794 = 0.0466.$$

So, for the first round combined, the excitement will be equal to $E_{1,1} + E_{1,2}$ *= 0.1701.*

Calculations for the remaining steps can be found in Table A.4 under Appendix A.

3.4.2 Numerical Exploration of Expected Excitement

In this section, we illustrate our mathematical findings with numerical examples. Example 3.2 shows the expected excitement score for 9 cases with different scoring probability pairs. For this example, the scoring probabilities of each team remains the same at each round. Example 3.3 elaborates on calculating the expected excitement for varying scoring probabilities at each round for various T values. Some cases also provides the excitement during various stages of a penalty shootout.

The code for this section was written in Python programming language using Jupyter Notebook environment. All computational experiments in this section were carried out on a 64-bit machine with Intel Core i5-4260U processor at 1.60 GHz and 4GB RAM.

Example 3.2. *In this example, we illustrated our mathematical findings for constant scoring probabilities with computational experiments. Scoring probabilities of both teams were entered by the user. The upper limit for the number of rounds* T *was determined as 200. However, some of the cases did not converge even after* $T =$

200 *rounds due to scoring probabilities of the opposing teams. Expected excitement for various cases with their properties can be found in Table 3.4. Their graphical illustrations can be found in Figure 3.2.*

Table 3.4 Expected Excitement of Scenarios with Different Scoring Probability Pairs

Graph	Scoring Probabilities		Highest Excitement Score	Round of Highest Excitement Score
	a	b		
(1)	0.1	0.6	0.42	1
(2)	0.3	0.4	2.23	43
(3)	0.5	0.5	7.99	> 200
(4)	0.7	0.5	1.18	10
(5)	0.7	0.7	7.33	> 200
(6)	0.4	0.8	0.51	1
(7)	0.9	0.3	0.32	1
(8)	0.8	0.9	1.13	22
(9)	0.9	0.9	4.82	> 200

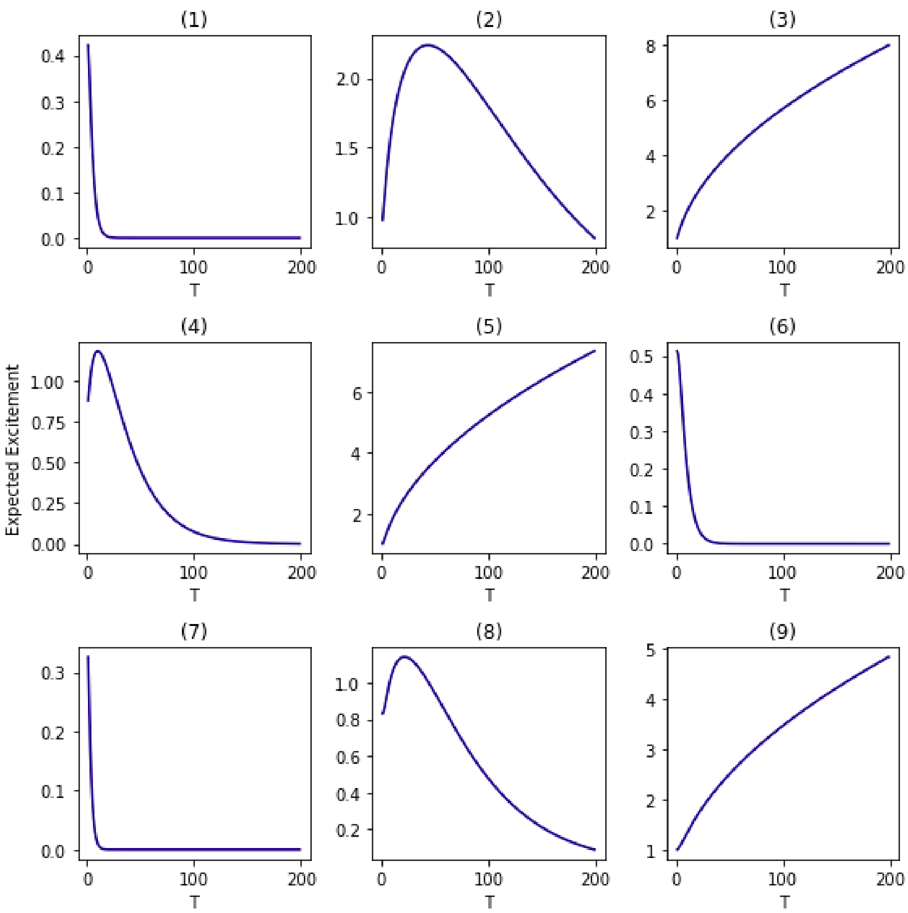

Figure 3.2 Graphical illustrations of scenarios with different scoring probability pairs

Graphs (3), (5) and (9) represent scenarios in which the opposing teams have equal *scoring probabilities. Results suggest that if opponents have the same scoring probabilities (i.e. same strength), the excitement score would constantly increase at each round. The highest excitement score was obtained for the case where both teams had the scoring probability of* 0.50*, which corresponds to Graph (3). That scenario can be considered as the ultimate penalty shootout scenario with the highest excitement score. Graphs (1), (6) and (7) show cases in which difference between scoring probabilities of the opponents are very high. In these cases, the excitement score would peak at the beginning of the shootouts and constantly decrease afterwards. These results illustrate that it would not be so exciting to watch a penalty shootout between two teams with a huge difference in scoring ability. Graph (4) shows a similar result, however, since the gap between the scoring probabilities of opposing teams is smaller, the highest excitement score for this scenario is greater than (1), (6) and (7)'s highest excitement scores. Finally, graphs (2) and (8) show teams with similar strengths. For graph (2), both teams are weak and have low scoring probabilities and for graph (8), both teams are strong and have high scoring probabilities. Comparison of these graphs shows that, mathematically, when opposing teams have similar strengths, penalty shootout scenario with the weaker teams would be more exciting. This may be related to the fact that weak teams would miss more shootouts compared to strong teams and it could cause their shootouts to last longer.*

Example 3.3. *In order to understand the excitement score calculation with varying scoring probabilities at each round, alternative penalty shootout sequences were created. These sequences, also known as scenarios, can be found in Table 3.5. Explanations for the excitement difference between these scenarios are also presented.*

Interpretations of the scenarios in Table 3.5 are as follows.

- *Scenario (1) represents a penalty shootout sequence between two weak teams that will last for 3 rounds. The current score is 0-0 and neither of the teams have shot a penalty yet. The excitement score for this scenario is 1.150. In contrast, (2) represents a penalty shootout sequence between two very strong teams that will last for 3 rounds. The excitement score for this scenario is 1.124. By comparing first two scenarios, it can be observed that excitement score for both ends of the spectrum will be very similar to each other. In (3), another scenario with the same setup is represented. However, this time both teams have an average scoring probability. Thus, excitement score slightly increases to 1.278.*
- *Scenarios (4) and (5) have the same scoring probability setups. The first scenario has the current score of 0-0 and neither of the teams have shot a penalty*

Table 3.5 Expected Excitement of Scenarios with Scoring Probability Vector Pairs

Scenario	T	Scoring Probabilities							x	y	t_A	t_B	Excitement
1	3	0.30	0.25	0.40					0	0	0	0	1.150
		0.20	0.45	0.15									
2	3	0.80	0.95	0.75					0	0	0	0	1.124
		0.85	0.90	0.80									
3	3	0.50	0.60	0.55					0	0	0	0	1.278
		0.50	0.55	0.60									
4	5	0.35	0.45	0.25	0.35	0.4			0	0	0	0	0.258
		0.85	0.90	0.70	0.80	0.80							
5	5	0.35	0.45	0.25	0.35	0.4			3	3	4	4	1.455
		0.85	0.90	0.70	0.80	0.80							
6	5	0.85	0.90	0.70	0.80	0.80			3	3	4	4	2.301
		0.90	0.85	0.80	0.90	0.70							
7	5	0.5	0.7	0.6	0.45	0.65			3	3	4	4	2.962
		0.46	0.60	0.65	0.70	0.50							
8	5	0.85	0.90	0.70	0.80	0.80	0.70		6	5	6	5	2.297
		0.90	0.85	0.80	0.90	0.70	0.80						
9	5	0.50	0.70	0.60	0.45	0.65			0	0	0	0	1.452
		0.50	0.60	0.65	0.55	0.50							
10	5	0.70	0.50	0.45	0.65	0.60			0	0	0	0	1.451
		0.65	0.50	0.5	0.55	0.60							
11	5	0.50	0.70	0.60	0.45	0.65			3	3	3	1	0.317
		0.50	0.60	0.65	0.55	0.50							
12	7	0.50	0.80	0.10	0.40	0.60	0.40	0.75	0	0	0	0	1.378
		0.45	0.30	0.80	0.90	0.85	0.35	0.55					
13	7	0.50	0.80	0.10	0.40	0.60	0.40	0.75	5	5	5	5	3.201
		0.45	0.30	0.80	0.90	0.85	0.35	0.55					

yet. On the other hand, second scenario depicts a sequence in which both teams had 3 successful shootouts. (4) has an excitement score of 0.258 whereas (5) has an excitement score of 1.455. This difference stems from the fact that team A is much weaker than team B and the outcome of the shootouts is very predictable. Thus, at the beginning of the shootout sequences the excitement is very low. However, in (5), team A has the same score as team B against all odds which makes the shootouts more exciting by increasing the variability in winning probabilities.

- *Scenario (6) represents shootouts between two strong teams and the current score is 3-3. The excitement for this scenario is 2.301. Since both teams are very strong, the outcome is unpredictable and the excitement is at its maximum. Under the same circumstances, scenario (5) has an excitement score of 1.455 because one of the teams is much weaker than the other. Another example with the same setup is scenario (7). This time, both teams have an average scoring probability. Similar to the results displayed in Table 3.4, excitement score formulation with varying scoring probabilities also considers scoring probabilities that are closer to* 50% *as the most exciting ones. That is* *why, scenario (7) has an excitement score of 2.962 and it is the most exciting shootout sequence among the scenarios with the same setup.*

- *Scenario (8) shows a penalty shootout sequence that went into overtime and it has an excitement score of 2.297. It has the same scoring probability setup as (6) for both teams which shows the fact that after a certain point, the excitement score between two highly competitive teams remains constantly high.*

- *Scenarios (9) and (10) were built to show that the excitement score is invariant under permutations of teams' scoring probabilities if the penalty shootouts have not started yet. These scenarios can be considered as two different penalty shootout sequences among the same opposing teams. However, the order of the players is shuffled between scenarios for both teams. Even though the order is shuffled, the excitement score remains the same for both scenarios.*

- *Scenarios (9) and (11) have the same scoring probabilities, but their setup is different. Scenario (9) has 0-0 as its current score whereas scenario (11)'s current score is 3-1. On the other hand, (9) has an excitement score of 1.452 and (11) has an excitement score of 0.317. This shows the importance of goal difference between two teams and aligns with the feature importance findings that will be explored in Chapter 4.*

- *Finally, two penalty shootout sequences with 7 rounds were considered. Both scenario (12) and scenario (13) have the same random scoring probabilities. (12) has 0-0 as its current score and neither of the teams have shot a penalty yet whereas (13) has 5-5 as its current score and each team have shot 5 penalties. Comparison of these scenarios provides insights for understanding the excitement in sports. If two teams with similar strengths have successfully scored each of their penalty shootouts up until the very last moment, the excitement of final penalties will be very high. Thus, the excitement of (13) will be 3.201 and it is higher than the excitement score of (12) which represents the beginning of the penalty shootouts.*

3.5 Maximum Likelihood Estimation for Scoring Probability Distribution

We know how to calculate the expected excitement of a penalty shootout given the scoring probabilities. Therefore, if we were given a probability distribution for the scoring probability of a team, we could identify the best T, which yields the maximum expected excitement $\mathcal{E}^T$ in expectation with respect to the scoring probability distribution.

However, the probability distribution of the scoring probability of a team is unknown in general. This motivates for the task of estimating the probability distribution of the scoring probability of a team in a random game. In this section, we investigate maximum likelihood estimation for this purpose and apply it to real data.

The Beta-Binomial distribution can be described as a form of Binomial distribution in which probability of success at each trial is randomly drawn from a Beta distribution. During penalty shootouts, probability of success is not known, thus, considering the randomness is very important. That is why, we need to estimate the parameters of the Beta distribution for cases where scoring probabilities of the opposing teams are unknown. The Beta distribution has two parameters, namely α and β. In order to obtain a maximum likelihood estimation for these parameters, iterative methods such as Expectation-Maximization algorithm combined with Newton-Raphson method algorithm was utilized. Notations for this part can be found in Table 3.6.

Table 3.6 Notations for Binomial Distribution

N	Total number of penalty trials
K	Number of successful penalties

3.5.1 Real Data

The data X, including 314 penalty shootouts, were generated using the results of the penalty shootouts from 1982-2018 FIFA World Cup, 1976-2020 UEFA European Cup, 2016-2020 Turkish Cup, 1993-2019 Copa América and 1991-2019 CONCACAF Cup. Data consist of the total number of penalty shootouts taken by each team and the number of successful penalties for each team. A representative sample from the data can be found in Figure 3.3.

	PenaltiesTaken(K)	SuccessfulPenalties(N)
0	5	4
1	8	7
2	9	8
3	5	5
4	6	5
5	11	9

Figure 3.3 A sample from the penalty shootouts data

3.5.2 Calculations

Suppose each data point i from the set X follows $x_i \sim Binom(n_i, p_i)$. The probability of success for each penalty comes from $p_i \sim Beta(\alpha, \beta)$, given that $\{p_i\}_{i=1}^{P}$ and $p_i \in (0,1)$. In addition, for each data point i, total number of penalties are represented as n_i and number of successful penalties are represented as k_i. The probability mass function of Binomial distribution and beta density are as follows

$$f(k_i \mid p_i) = \binom{n_i}{k_i} \cdot p_i^{k_i} \cdot (1-p_i)^{n_i-k_i} \tag{3.16}$$

$$f(p_i) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{(\alpha-1)}(1-p_i)^{(\beta-1)} = \frac{p_i^{\alpha-1}(1-p_i)^{\beta-1}}{B(\alpha,\beta)}. \tag{3.17}$$

The joint distribution can be formulated as $f(k_i, p_i) = f(p_i)f(k_i \mid p_i)$ which is equal to

$$\begin{aligned} &\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{(\alpha-1)}(1-p_i)^{(\beta-1)} \binom{n_i}{k_i} \cdot p_i^{k_i} \cdot (1-p_i)^{n_i-k_i} \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(n_i+1)\Gamma(k_i+1)}{\Gamma(n_i-k_i+1)} p_i^{(\alpha+k_i-1)}(1-p_i)^{(n_i-k_i+\beta-1)} \end{aligned} \tag{3.18}$$

Beta distribution is the conjugate prior for the Binomial distribution in Bayesian inference which provides an easier numerical computation. That is why, it was selected as the probability distribution for our study.

$$\begin{aligned} f(k_i) &= \int_0^1 f(k_i, p_i) dp_i. \\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \frac{\Gamma(n_i+1)\Gamma(k_i+1)}{\Gamma(n_i-k_i+1)} \frac{\Gamma(\alpha+k_i)\Gamma(n_i+\beta-k_i)}{\Gamma(n_i+\alpha+\beta)} \end{aligned} \tag{3.19}$$

$$
\begin{aligned}
f(p_i \mid k_i) &= \frac{f(k_i, p_i)}{f(k_i)} \\
&= \frac{\Gamma(\alpha+\beta+n_i)\Gamma(\alpha+k_i)}{\Gamma(n_i+\beta-k_i)} p_i^{(\alpha+k_i-1)}(1-p_i)^{(n_i-k_i+\beta-1)}
\end{aligned}
\tag{3.20}
$$

posterior distribution is a Beta distribution with $\alpha_{post,i} = \alpha + k_i$ and $\beta_{post,i} = \beta + n_i - k_i$.

3.5.3 Maximum Likelihood Estimation

General formulations of the maximum likelihood estimation are as follows.

Let $X_1, \ldots, X_n$ be iid random variables sampled from a distribution with $f(x \mid \theta)$ density. Suppose parameters of this distribution are represented by θ in a parameter space. Thus, the likelihood function for $X_{1:n} = x_{1:n}$ will be

$$L_n(\theta; x_{1:n}) = \prod_{i=1}^{n} f(x_i \mid \theta). \tag{3.21}$$

The log-likelihood function can be written as

$$\ell_n(\theta; x_{1:n}) = \log L_n(\theta \mid x_{1:n}) = \sum_{i=1}^{n} \log f(x_i \mid \theta). \tag{3.22}$$

The maximum likelihood estimate of θ is defined as

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, \ell_n(\theta \mid x_{1:n}). \tag{3.23}$$

Given that $\ell_n(\theta \mid x_{1:n})$ is differentiable in θ, take the logarithmic derivative with respect to θ and set it to 0.

$$\frac{\partial \ell_n(\theta \mid x_{1:n})}{\partial \theta} = 0. \tag{3.24}$$

For our study, maximum likelihood estimation of the parameters of the Beta distribution given a set of probabilities will be relevant. Maximum likelihood estimation for Beta distribution given the a sequence of variables $p_{1:n}$ with each $p_i \in (0,1)$ consists of the following steps.

1. Let $\theta = (\alpha, \beta)$. Derive the log-likelihood function.

$$\begin{aligned}\ell(\alpha,\beta;p_{1:n}) &= \sum_{i=1}^{n} \ln f(p_i \mid \theta) \\ &= \sum_{i=1}^{n} \ln\left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{(\alpha-1)}(1-p_i)^{(\beta-1)}\right) \\ &= (\alpha-1)\sum_{i=1}^{n}\ln p_i + (\beta-1)\sum_{i=1}^{n}\ln(1-p_i) + \\ &\quad n\left(\ln\Gamma(\alpha+\beta) - \ln\Gamma(\alpha) - \ln\Gamma(\beta)\right)\end{aligned} \tag{3.25}$$

2. Take the logarithmic derivatives w.r.to α and β. It should be noted that

$$\psi(x) = \frac{\partial \ln\Gamma(x)}{\partial x} = \frac{\Gamma'(x)}{\Gamma(x)}$$

For α:

$$\begin{aligned}&\frac{\partial}{\partial\alpha}\left((\alpha-1)\sum_{i=1}^{n}\ln p_i + (\beta-1)\sum_{i=1}^{n}\ln(1-p_i) + n\left(\ln\Gamma(\alpha+\beta) - \ln\Gamma(\alpha) - \ln\Gamma(\beta)\right)\right) \\ &= n\left(\psi(\alpha+\beta) - \psi(\alpha)\right) + \sum_{i=1}^{n}\ln p_i\end{aligned} \tag{3.26}$$

For β:

$$\begin{aligned}&\frac{\partial}{\partial\beta}\left((\alpha-1)\sum_{i=1}^{n}\ln p_i + (\beta-1)\sum_{i=1}^{n}\ln(1-p_i) + n\left(\ln\Gamma(\alpha+\beta) - \ln\Gamma(\alpha) - \ln\Gamma(\beta)\right)\right) \\ &= n\left(\psi(\alpha+\beta) - \psi(\beta)\right) + \sum_{i=1}^{n}\ln(1-p_i)\end{aligned} \tag{3.27}$$

3. Set the derivatives to 0 and solve for α and β.

3.5.4 Expectation Maximization Algorithm & Newton-Raphson Method

The expectation-maximization (EM) algorithm can be considered as an iterative method to obtain local maximum likelihood or maximum a posteriori (MAP) parameter estimations. It was first introduced in 1977 and described as a solver for models including latent variables, known data points and unknown parameters (Dempster, Laird & Rubin, 1977). It consists of two steps: E-step and M-step.

E-step: EM algorithm will start with initialization of parameters. Suppose you have a set X of observed data points, set of unobserved data points Δ and a vector of unknown parameters θ. The likelihood function given the complete data will be $L(\theta;\Delta,X) = p(\Delta,X \mid \theta)$. For the expectation step, a function Q will be defined as the expected value of the log-likelihood function of θ at each step n.

$$Q\left(\theta \mid \theta^{(n)}\right) = \mathbb{E}[\log L(\theta;\Delta,X)] \ = \mathbb{E}[\ \ell(\theta;\Delta,X)] \tag{3.28}$$

M-step: This step will consist of maximizing the function Q with respect to θ at each iteration until it converges.

$$\theta^{n+1} = \underset{\theta}{\operatorname{argmax}}\, Q\left(\theta \mid \theta^{(n)}\right) \tag{3.29}$$

For our study, we used the EM algorithm to estimate the parameters of a Beta-Binomial distribution that is used to model the number of successful penalties. This model corresponds to assigning a Beta distribution for the scoring probability of each team in each penalty shootout, and assuming that given those probabilities the number of successful penalties has a Binomial distribution. The E-step of the EM algorithm for this model requires taking the posterior expectations of each $\log p_i$ and $\log(1-p_i)$ given the data point n_i, k_i (trials, successes) where $p_i \sim Beta(\alpha,\beta)$ for the set of n observed data points in Section 3.5.2.

In Section 3.5.2, it was determined that if $f(p_i)$ is a beta density with parameters α and β, $f(p_i \mid K_i = k_i)$ is a beta density with parameters $\alpha + k_i$ and $\beta + n_i - k_i$.

Initial parameters are defined as $\theta_0 = (\alpha_0, \beta_0)$. E-step at the j'th iteration of EM is calculated as follows

$$S_i = \mathbb{E}_{\theta_{j-1}}[\ \log p_i \mid K_i = k_i] = \psi(\alpha_{j-1} + k_i) - \psi(\alpha_{j-1} + \beta_{i-1} + n_i)$$
$$Y_i = \mathbb{E}_{\theta_{j-1}}[\ \log(1-p_i) \mid K_i = k_i] \ = \psi(\beta_{j-1} + n_i - k_i) - \psi(\alpha_{j-1} + \beta_{j-1} + n_i)$$

For the EM algorithm for the Beta-Binomial distribution, the M-step reduces to maximizing a function over α and β of the form

$$\bar{S}(\alpha-1) + \bar{Y}(\beta-1) - \ln\Gamma(\alpha+\beta) + \ln\Gamma(\alpha) + \ln\Gamma(\beta)$$

where

$$\overline{S} = \frac{\sum_{i=1}^{n} S_i}{n}, \quad \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}. \tag{3.30}$$

are derived in the E-step preceding the M-step using the parameters of the previous iteration. Since the maximization is not tractable, the Newton-Raphson algorithm can be used.

Newton-Raphson method approximates the roots of a function $f(x)$ by calculating

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{3.31}$$

in which x_n and x_{n+1} represents n^{th} and $(n+1)^{th}$ iterations.

Using this method in the M-step of the EM algorithm, we have an inner loop for the maximization process for θ where the iterates of this loop are calculated as

$$\theta^{(r+1)} = \theta^{(r)} - G^{-1}(\theta^{(r)})g(\theta^{(r)}) \tag{3.32}$$

starting from $\theta^{(0)} = \theta_{j-1}$. When the inner loop is run for a certain number of iterations, say $m > 0$, the last iterate $\theta^{(m)}$ is taken as $\theta_j = (\alpha_j, \beta_j)$, the estimate of the j'th iteration of the EM algorithm.

In (3.32), the gradient vector and the Hessian matrix are defined as

$$g(\theta) = \begin{pmatrix} \psi(\alpha+\beta) - \psi(\alpha) + \overline{S} \\ \psi(\alpha+\beta) - \psi(\beta) + \overline{Y} \end{pmatrix} \tag{3.33}$$

and

$$G(\theta) = \begin{pmatrix} \psi_1(\alpha+\beta) - \psi_1(\alpha) & \psi_1(\alpha+\beta) \\ \psi_1(\alpha+\beta) & \psi_1(\alpha+\beta) - \psi_1(\beta) \end{pmatrix} \tag{3.34}$$

where ψ_1 represents trigamma function. The derivations behind these equations are shown in (3.26) and (3.27).

For our model, we set the initial parameters as $\alpha_0 = 0$ and $\beta_0 = 0$. After 10000 iterations, parameter estimations for α and β converges to 106.08 and 33.58, respectively. The graph representing their convergence can be seen in Figure 3.4.

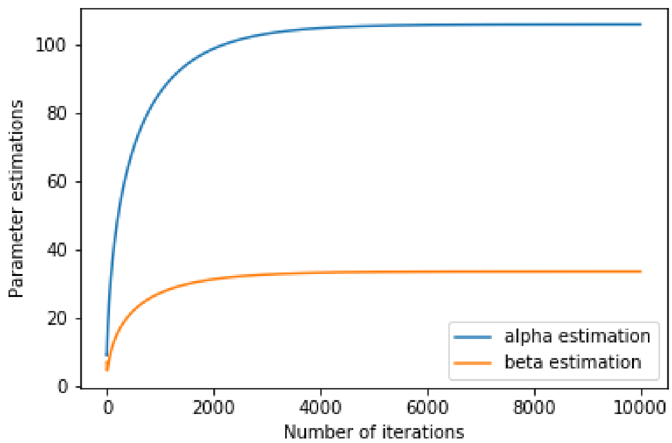

Figure 3.4 α and β parameter estimations

By obtaining these estimated values for the Beta distribution parameters, we could approximate the expectation of the expected excitement with respect to the distribution of the scoring probabilities, which is given by

$$\int \mathbb{E}[\mathcal{E}_T(S)|a,b] Beta(a;\alpha,\beta) Beta(b;\alpha,\beta) da db$$

Since we estimated α and β using MLE, we have the distributions of the scoring probabilities a and b. Hence, we can calculate an importance sampling estimate of the expectation above by sampling a and b from their Beta distributions, respectively. Algorithm 2 displays the process of calculating the expectation of the expected excitement with probability distributions.

Algorithm 2 Excitement Score Calculation with Estimated Beta Parameters

Input: $\alpha = 106.08$, $\beta = 33.58$, $T = 200, M = 5$

```
Start with E_max = 0
for T = 1,2,...,T_max do
    for m = 1,2,...,M do
        Draw a ~ Beta(106.08,33.58) and b ~ Beta(106.08,33.58)
        Calculate E^(m) = E[ℰ^T(S) | a,b]
    end for
    E(T) = (Σ_{m=1}^{M} E^(m)) / M
    if E_max < E(T) then
        E_max ← E(T)
        T_best ← T
    end if
end for
max_excitement ← E_max
best_penalty_round ← T_best
return max_excitement, best_penalty_round
```

Where: $E_{\max} = 0$; $E^{(m)} = \mathbb{E}[\mathcal{E}^T(S) \mid a,b]$; $E(T) = \frac{\sum_{m=1}^{M} E^{(m)}}{M}$; $T_{best} \leftarrow T$.

3.6 Excitement Score Calculator

Graphical User Interface (GUI) is an interface which allows users to interact with electronic devices by using graphical icons. It provides a comfortable experience for the user, since navigating through your desktop or your applications would be easier with icons representing the function of a program. First prototype of GUI was developed by Xerox Paro Alto Research Center in 1979.

In this study, GUI was created to provide a user-friendly experience in calculating the excitement scores of different penalty shootout scenarios. In Figure 3.5, we provide an example case where current score is 6-6, probability of scoring a penalty shootout for teams are 70% and 60% and number of completed penalty shootouts for teams are 7 and 7. For the excitement score calculation, formulations in Chapter 3 were used. In return, GUI returns the excitement score for this scenario as 0.9527.

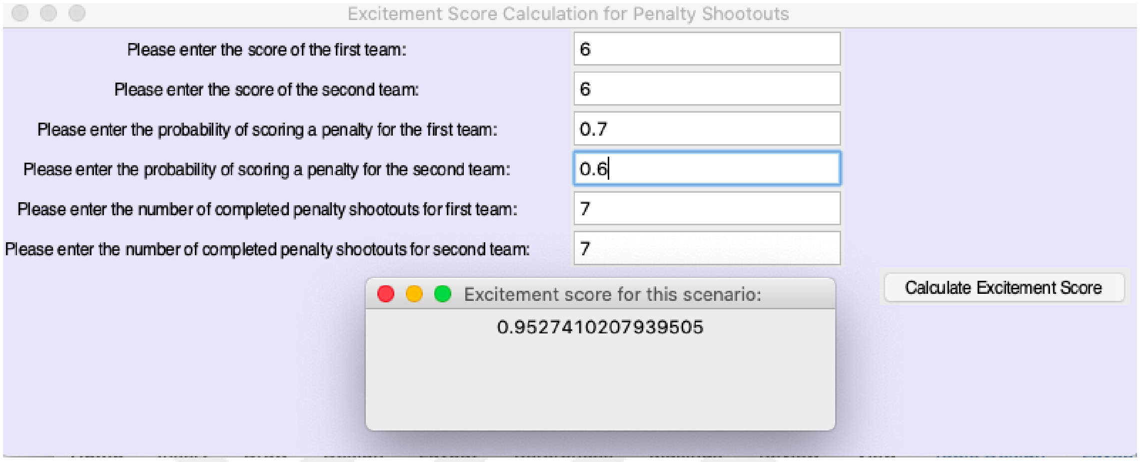

Figure 3.5 Graphical user interface for excitement score calculation

4. MODELING

In this chapter, we compare the findings of mathematical excitement calculations with the viewers' perception in real-life by examining the data of a survey that we have conducted. This survey consists of the paired comparisons of 20 penalty shootout scenarios. The excitement score of each penalty shootout scenario in this survey is calculated and ranked by using the formulations in Chapter 3. For the real-life part, preferences of the participants regarding which penalty shootout scenario they found more exciting are evaluated using the Bradley-Terry model. Furthermore, using machine learning applications, we explore which features of a penalty shootout influence people's decision making process.

4.1 Excitement Survey

We have conducted a survey with 310 participants. This survey consisted of questions regarding pairwise comparison of 20 different penalty shootout scenarios. All of these scenarios have been compared against each other, which corresponds to a total of 190 paired comparison, thus, our survey consisted of 190 questions. Among these 190 possible questions, each participant answered 15 questions that were randomly assigned to them. Example survey questions can be found in Appendix A shown as Table A.1, Table A.2, Table A.3.

As the outcome of this survey, we had the preferences of the participants for different paired scenario comparisons. Therefore, we were able to rank these scenarios based on the number of times they were preferred. As a more principled approach, we also utilised Bradley-Terry model to rank preferences of the participants. Attributes of each penalty shootout scenario are explained in Table 4.1. The penalty shootout scenarios created for this study are represented in Table 4.2.

4.1.1 Survey Participants

Out of 310 respondents, 117 of them were average football viewers and 193 of them can be considered as football fans. Since we would like to observe different participant groups, preference models were also built separately for these participant classes in Section 4.3.

Table 4.1 Attributes and Definitions

Round	Current time in each penalty shootout. Round ends when both teams have taken the penalty shootout.
Score	Current score of the penalty shootouts
Scoring Probability	Probability of scoring the penalty shootout for each team in each round
Penalty Kick In	Indicates which team will take the penalty kick next

Table 4.2 Survey Scenarios

Scenario	**Round**	**Score**	**Scoring probability (Team A)**	**Scoring probability (Team B)**	**Penalty kick in**
Scenario 1	4	3-1	70%	50%	Team A
Scenario 2	1	1-0	60%	50%	Team B
Scenario 3	5	4-4	70%	60%	Team A
Scenario 4	3	1-1	55%	55%	Team A
Scenario 5	7	4-4	90%	30%	Team A
Scenario 6	4	2-2	45%	55%	Team A
Scenario 7	3	0-1	80%	80%	Team A
Scenario 8	1	0-0	85%	75%	Team A
Scenario 9	6	6-5	80%	60%	Team B
Scenario 10	2	1-1	90%	90%	Team B
Scenario 11	8	6-6	70%	60%	Team A
Scenario 12	5	2-1	25%	80%	Team A
Scenario 13	4	4-2	75%	55%	Team B
Scenario 14	1	1-0	80%	90%	Team B
Scenario 15	10	8-8	75%	85%	Team B
Scenario 16	5	4-3	55%	45%	Team B
Scenario 17	5	3-3	85%	85%	Team A
Scenario 18	3	3-1	40%	70%	Team B
Scenario 19	6	5-5	70%	70%	Team A
Scenario 20	2	2-0	90%	30%	Team B

4.1.2 Survey Data Generation

We were able to obtain pairwise comparisons of the scenarios due to the structure of this survey. That is why, the survey data consist of 4656 entries and these entries are generated depending on how many times a scenario is selected and not selected in their respective pairwise comparison. Sample from the raw data can be found in Figure A.1 under Appendix A. It can be seen that Scenario 1 was selected 11 times and not selected 17 times during its comparison with Scenario 2. In order to make the data computationally easier to calculate, we have merged each comparison as a difference of feature values and created a processed data. A sample from the processed data can be found in A.2 under Appendix A in which feature values of Scenario 1 and Scenario 3 are merged.

In total, 8 features were considered for the machine learning algorithms to come and their detailed explanations can be found in Table 4.7. The values of these features were derived from the difference between compared scenarios, again, due to pairwise comparisons and to make calculations computationally easier.

4.2 Bradley-Terry Model

The Bradley-Terry model was introduced by Bradley & Terry (1952) and it can be described as a probability model used for pairwise evaluations. Suppose there exists a pair of individuals m and k from a population. The probability estimation for the pairwise evaluation where $m > k$ depicts that m is preferred to k which can be written as

$$\mathbb{P}(m > k) = \frac{\alpha_m}{\alpha_m + \alpha_k}. \tag{4.1}$$

α_m and α_k are designated real-valued scores assigned for m and k, respectively. These score factors can be expressed in many ways. The variation used by Bradley and Terry is the exponential score function which can be written as $\alpha_m = e^{\beta_m}$. Thus, $\mathbb{P}(m > k)$ can also be represented as

$$\mathbb{P}(m > k) = \frac{e^{\beta_m}}{e^{\beta_m} + e^{\beta_k}}. \tag{4.2}$$

Alternatively, Bradley-Terry model can be expressed in a log-linear formulation to reduce the model into a logistic regression form which is denoted in (4.3).

$$
\begin{aligned}
&logit(\mathbb{P}(m > k)) = \sigma^{-1}(\mathbb{P}(m > k)) = \log\left(\frac{\mathbb{P}(m > k)}{1 - \mathbb{P}(m > k)}\right) \\
&= \beta_m - \beta_k
\end{aligned}
\tag{4.3}
$$

where $\beta_m = \log(\alpha_m)$ is the ability parameter with logarithmic scale.

The Bradley-Terry model is a very common method for comparing the teams in a sports tournament with regard to their power. In that case, α_m can represent the strength of a team which may be estimated from the total number of games team m have won. Various studies used this model for forecasting the outcome of a sports tournament or to compare the powers of teams in a league. In some studies, Bradley-Terry model was used to predict the results of tennis matches (Ian & Morton, 2011) or to estimate abilities and rankings of basketball teams by considering their winning percentages and home-field advantages (Cattelan, Varin & Firth, 2013).

For psychometric applications, it can be used to compare preferences of subjects, including subject-specific attributes. In their research, Dittrich, Katzenbeisser & Reisinger (2000) used Bradley-Terry model and its subclass applications to rank newspaper preferences of people considering subject-specific variables such as age, income and gender. In our study, the Bradley-Terry model was used to determine the preference scores α_m of the scenarios based on the responses of the survey participants. Scenario 1 was taken as a baseline for easier interpretation and its log-ability is equal to $\beta_1 = 0$. In order to calculate the ability parameters and create the Bradley-Terry model, R package BradleyTerry2 was used. Results can be seen in Table 4.3 and in Figure 4.1 .

According to the ability estimates, the most preferred scenario is Scenario 19 and the least preferred scenario is Scenario 20. Bradley-Terry scenario rankings are presented with the excitement score rankings and voting-based ranking in Section 4.8.

Table 4.3 Bradley-Terry Model Scenario Ability Estimates

Scenario	Estimate	Standard Error	Quasi-Standard Errors
Scenario 1	0.000	0.000	0.101
Scenario 2	0.507	0.136	0.092
Scenario 3	1.106	0.139	0.093
Scenario 4	0.878	0.136	0.090
Scenario 5	0.799	0.141	0.097
Scenario 6	0.795	0.135	0.090
Scenario 7	0.766	0.136	0.091
Scenario 8	0.706	0.137	0.092
Scenario 9	0.906	0.136	0.090
Scenario 10	0.942	0.138	0.092
Scenario 11	1.413	0.141	0.096
Scenario 12	0.761	0.139	0.094
Scenario 13	0.136	0.137	0.094
Scenario 14	0.538	0.138	0.092
Scenario 15	1.389	0.142	0.098
Scenario 16	0.996	0.136	0.091
Scenario 17	1.409	0.140	0.096
Scenario 18	0.693	0.137	0.092
Scenario 19	1.456	0.143	0.010
Scenario 20	-0.044	0.139	0.009

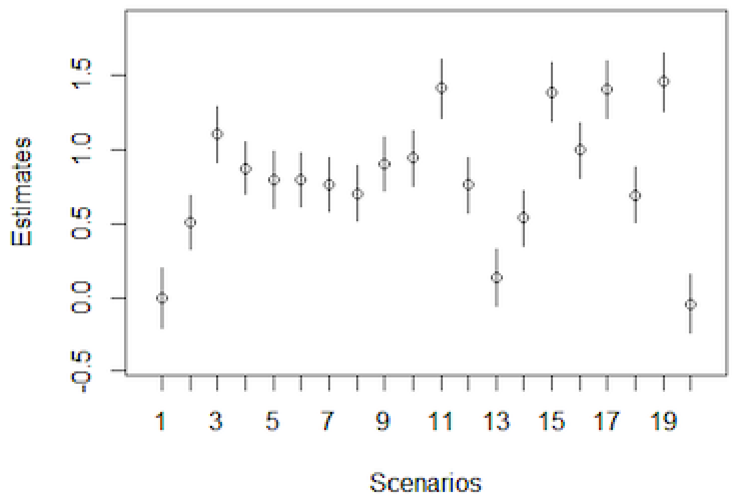

Figure 4.1 Estimation intervals based on quasi-standard errors

4.3 Bradley-Terry Preference Rankings for Survey Participant Groups

As we previously stated in Section 4.1.1, survey participants consist of two groups: average football viewers and football fans. Their preference ranking results can be seen in Table 4.4.

Table 4.4 Bradley-Terry Preference Rankings Based on Participant Groups

Rank	Scenario	Ability - Football Fans	Scenario	Ability - Average Viewer
1	Scenario 19	1.944	Scenario 10	1.126
2	Scenario 15	1.843	Scenario 17	0.972
3	Scenario 11	1.841	Scenario 11	0.888
4	Scenario 17	1.718	Scenario 19	0.809
5	Scenario 3	1.522	Scenario 15	0.766
6	Scenario 16	1.460	Scenario 8	0.693
7	Scenario 4	1.258	Scenario 14	0.629
8	Scenario 5	1.235	Scenario 9	0.571
9	Scenario 6	1.186	Scenario 3	0.557
10	Scenario 12	1.155	Scenario 18	0.498
11	Scenario 9	1.152	Scenario 2	0.461
12	Scenario 7	1.023	Scenario 7	0.375
13	Scenario 10	0.915	Scenario 4	0.311
14	Scenario 18	0.832	Scenario 6	0.261
15	Scenario 8	0.749	Scenario 16	0.213
16	Scenario 14	0.555	Scenario 5	0.205
17	Scenario 2	0.551	Scenario 12	0.181
18	Scenario 13	0.126	Scenario 13	0.139
19	Scenario 20	0.076	Scenario 1	0.000
20	Scenario 1	0.000	Scenario 20	-0.279

By examining the outcome, we can explore the similarities and differences between these groups. To begin with, ability estimates for the football fans are much higher than the estimates for the average viewers. This can be interpreted as a sign of football fans enjoying most penalty shootouts and it leads to them being indecisive while choosing one to watch. For the average football viewer, making a decision is a lot easier and excitement a penalty shootout brings is much less.

The most and the least exciting scenarios are very similar for both groups with a couple of minor differences. For average football viewers, the most exciting scenario is Scenario 10, which shows a rally between two very strong teams. It is understandable for an average viewer to prefer a shootout with the strongest teams. They may want to spend their time watching a competitive and probably popular shootout scenario since they are not a constant football fan.

4.4 Expected Excitement Scores of Survey's Penalty Shootout Scenarios

By using the mathematical formulations in Chapter 3, we found the excitement scores of the penalty shootout scenarios in Table 4.2. Python programming language was utilised for the coding of this section. Excitement scores are shown in Table 4.5.

Table 4.5 Scenario Rankings for Excitement Score

Rank	Scenario	Excitement Score
1	Scenario 4	1.246
2	Scenario 2	1.150
3	Scenario 8	1.103
4	Scenario 6	1.073
5	Scenario 14	1.009
6	Scenario 19	1.000
7	Scenario 17	1.000
8	Scenario 18	0.995
9	Scenario 11	0.952
10	Scenario 3	0.952
11	Scenario 7	0.702
12	Scenario 12	0.668
13	Scenario 16	0.630
14	Scenario 9	0.606
15	Scenario 10	0.580
16	Scenario 15	0.224
17	Scenario 5	0.173
18	Scenario 13	0.134
19	Scenario 1	0.051
20	Scenario 20	0.004

According to the mathematical formulation rankings, the most exciting penalty shootout scenario is Scenario 4. It is consistent with our experiments since scoring probabilities in this scenario are the same (%55) and they are very close to %50, which tends to be the probability that provides the most variability in a penalty shootout. Additionally, the current score of Scenario 4 is 1-1, which means variability in the winning probabilities of both teams will be high.

On the other hand, the least exciting penalty shootout scenario is Scenario 20. Again, it is consistent with the previous results since there is a huge scoring probability difference between two teams and the current score shows that the variability in the winning probabilities of both teams is not high. In this scenario, Team A is very close to winning the penalty shootouts which mathematically makes the penalty shootouts less exciting to watch.

4.4.1 Modifications in Conditional Expected Excitement Score Formulation

Conditional expected excitement formulation in (3.14) can be modified and alternative excitement scores can be obtained. In order to have an outcome that is closer to the survey results, the remaining number of penalties can be removed from the formulation. Even though general rule in football is to shoot 5 penalties each and mathematical applications should take that into consideration, in some cases penalty shootouts extend beyond that limitation. In our formulations, that extension is evaluated by calculating the tie-break excitement, however, according to the viewer, the number of penalty shootouts remaining is a mystery. That is why, an alternative to (3.14) can be proposed as follows

$$\hat{\mathcal{E}}(\mathcal{S}) = F(\mathcal{S}) + G(\mathcal{S}) + D(\mathcal{S})\mathcal{E}_0. \tag{4.4}$$

where $\mathcal{S} = \{S_{t_1,t_2} = (x,y)\}$ is the given score. The excitement scores of the penalty shootout scenarios obtained by using (4.4) are shown in Table 4.6.

Table 4.6 Scenario Rankings for the Alternative Excitement Score

Rank	**Scenario**	**Alternative Excitement Score**
1	Scenario 16	1.098
2	Scenario 19	1.001
3	Scenario 17	1.000
4	Scenario 11	0.952
5	Scenario 3	0.952
6	Scenario 9	0.905
7	Scenario 6	0.713
8	Scenario 12	0.668
9	Scenario 4	0.546
10	Scenario 18	0.538
11	Scenario 14	0.483
12	Scenario 8	0.437
13	Scenario 2	0.418
14	Scenario 7	0.390
15	Scenario 10	0.319
16	Scenario 15	0.198
17	Scenario 5	0.173
18	Scenario 13	0.114
19	Scenario 1	0.035
20	Scenario 20	0.001

4.5 Feature Importance

In this section, machine learning algorithms were utilised to examine the survey results in more detail and to identify the factors affecting the scenario preferences of the participants. This survey can be considered as a binary classification problem, in which the participants indicate the penalty shootout scenario that they prefer to watch. Thus, data utilised for this research have binary output and consist of participants' responses. In every question, specific details related to compared scenarios were provided. These details were then considered as the features of that scenario. General properties of the data were summarised in Section 4.1.2. Description of the features and the output of the data can be found in 4.7.

Table 4.7 Survey Inputs and Output

ID	Name	Type	Description
1	Total Goal	Numeric	Total number of successful penalty kicks in a scenario
2	Round	Numeric	Current penalty round of a scenario
3	Goal Difference	Numeric	Successful penalty kick difference between two teams in a scenario
4	Scoring Probability - Team A	Numeric	Penalty scoring probability of the first team in a scenario
5	Scoring Probability - Team B	Numeric	Penalty scoring probability of the second team in a scenario
6	Probability Difference	Numeric	Penalty scoring probability difference between the first team (A) and the second team in a scenario (B)
7	Penalty Turn	Binary	Displays which team has the penalty kick turn in a scenario
8	**Output:** Selected	Binary	Displays if that scenario is selected by the survey participant.

Various machine learning algorithms have been used for binary classification problems. In their research, Yamak, Saunier & Vercouter (2015) have used Support Vector Machine, Random Forest, Naive Bayesian, K-Nearest Neighbor and Adaptive Boosting algorithms for a binary classification problem in which they aim to detect a special case of social media manipulation, multiple identity accounts. Deep learning algorithms were also used in binary classification problems as more advanced methods. Abdulla & Alashoor (2020) have used Artificial Deep Neural Networks (ADNN) in a binary classification problem setting for the malicious packet detection whereas Khullar, Salgotra, Singh & Sharma (2021) have used 2-dimensional Convolutional Neural Network (CNN) algorithm to diagnose ADHD using resting state MR images.

Machine learning algorithms were also utilised as a tool to determine importance features in a model. In a study by Guo, Zhou, Zhang & Yang (2018), the main objective was to forecast the short-term electricity load with the intention of improving electricity consumption efficiency. In addition, Guo, Zhou, Zhang and Yang have used random forest and gradient boosting algorithms to find which factors were influencing the electricity consumption. Another study by Jia, Lin & Liu (2019) have analysed the attributes affecting the earthquake fatalities in China mainland and utilised deep learning methods to estimate potential fatalities by using these attributes. For their research regarding feature importance, Jia, Lin and Liu have used random forest, classification, regression tree and adaptive boosting algorithms.

In our research, we have created different binary classification models using logistic regression (LR), random forest (RF), eXtreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), K-Nearest Neighbors(k-NN) and artificial neural network (ANN) algorithms. Among these algorithms; logistic regression, random forest, XGBoost and adaptive boosting algorithms were utilised in order to determine important features. Grid Search and Random Search were used for hyperparameter tuning and 10-fold cross-validation was used for model validation. Accuracy was used as the main performance metric. However, the models also have precision and recall values as well as a confusion matrix.

4.5.1 Accuracy

In this study, if a survey participant preferred a certain penalty shootout scenario at a question it was labeled as 1. If that scenario was not preferred by the participant, it was labeled as 0. To illustrate, consider a scenario that was selected by 14 different participants and not selected by 11 different participants among those who have answered that question. So, there were 14 rows labeled as 1 and 11 rows labeled as 0. By doing so, accuracy can be considered as the percentage of correctly labeled outcomes by the model versus total number of outcomes.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}}$$

Recall computes how frequently a model accurately recognizes a positive outcome for people who have selected a certain scenario at a question.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Precision computes how frequently a model accurately recognizes a negative outcome for people who have not selected a certain scenario at a question.

$$\text{Precision} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$$

In this study, data were split by using 70/30 ratios for machine learning models. 70% of the data was used for training set and 30% of the data was used for test set. There was no need for a separate validation set since 10-fold cross validation was performed.

4.5.2 Models

Machine learning algorithms that we utilised to create the models are presented in this section.

4.5.2.1 Logistic Regression

Logistic regression is a supervised learning algorithm technique that is heavily used for classification problems. It is a predictive analysis algorithm in which observations are assigned to two or more discrete classes. Generally, it has binary dependent variables, however, it can be performed for multinomial and ordinal variable types too.

Logistic regression model starts with the set of known independent variables and the corresponding dependent variable for each observation. There is an intercept term and each independent variable has a coefficient (also known as weight) which is predicted by the model. Training part consists of model learning the intercept and coefficients. In order to obtain the best coefficients, log-likelihood function for all observations must be maximized. Mathematical formulation of these operations are as follows.

Suppose there exists a set of k predictors represented as X = $(X_1, X_2, \ldots, X_k)$. In addition, there exists a response y_i for each observation $i = 1 \ldots n$. The linear combination of predictors with θ coefficients is $\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \cdots +, \theta_k X_k$. Thus, linear function $f(x)$ will be written as

$$f(x) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \cdots +, \theta_k X_k. \tag{4.5}$$

However, this linear function is an unbounded linear equation and we need the probability of an observation being assigned to one of the classes. In addition, this probability must vary between 0 and 1. That is why, sigmoid function will be used as a logarithmic transformation to map predicted values between 0 and 1.

$$\text{Sigmoid function} = \sigma(z) = \frac{1}{1 + e^{-z}} \tag{4.6}$$

For this case, $p(x)$ is the sigmoid function of $f(x)$ and it can be interpreted as the predicted probability that the output for a given set of predictors is equal to 1. Thus, $1 - p(x)$ can be interpreted as the predicted probability that the output for a given set of predictors is equal to 0. As a side note, given that $p(x) = p$, the odds of an event can be represented as $\frac{p}{1-p}$.

$$p(x) = p(Y = 1 \mid X) \tag{4.7}$$

$$\sigma(f(x)) = p(x) = \frac{1}{1+e^{(-f(x))}} = \frac{1}{1+e^{(-(\theta_0+\theta_1 X_1+\theta_2 X_2+\cdots+,\theta_k X_k))}} \tag{4.8}$$

With a threshold value equal to 0.5, logistic regression can be used as a linear classifier. The model predicts y as 1 when $p(x) \geq 0.5$ as 0 when $p(x) < 0.5$. To obtain more accurate predictions for the coefficients at each iteration, we will consider likelihood for fitting. Likelihood can be written as

$$\Pr(Y \mid \theta; X) = L(\theta \mid y; x) = \prod_{i=1}^{n} \Pr(y_i \mid x_i, \theta) = \prod_{i=1}^{n} p_\theta(x_i)^{y_i}(1-p_\theta(x_i))^{(1-y_i)}. \tag{4.9}$$

For simplicity, the log-likelihood function will be calculated as

$$L(\theta \mid y; x) = \ell(\theta \mid y; x) = \sum_{i=1}^{n} y_i \log(p_\theta(x_i)) + (1-y_i)\log(1-p_\theta(x_i)). \tag{4.10}$$

In order to find optimal coefficients, we need to obtain θ values that maximizes log-likelihood function. To make mathematical calculations easier, an optimization algorithm such as gradient ascent or gradient descent can be used. By taking partial derivative of the log-likelihood with respect to each parameter using chain-rule, we can get closer to optimal coefficient values at each iteration.

4.5.2.2 Random Forest

Random forest algorithm is a supervised learning technique that is used for both regression and classification problems. It was introduced by Breiman (2001) as a combination of decision trees and bagging. It is considered as an ensemble learning method due to using a combination of decision trees for training. Random forest is a heavily used algorithm since it can handle missing data, reduces overfitting and increases precision. For classification problems, predicted output is determined by the majority-voting system.

4.5.2.3 Extreme Gradient Boosting

Extreme Gradient Boosting (XGBoost) is a fairly new algorithm released in 2014. It is a supervised learning technique that can be used for both regression and classification problems. Similar to random forest, it is an ensemble learning method since it uses a combination of decision trees by adding them one at a time to correct the predictions of the previous models. This method is also known as boosting. Finally, these models are fit by gradient descent optimization algorithm. As oppose to gradient boosting algorithm, XGBoost uses similarity score and gain for building trees. Due to its ability to control overfitting, it is a highly popular machine learning algorithm.

4.5.2.4 Adaptive Boosting

Adaptive Boosting (AdaBoost) is a supervised learning technique that was created to increase the efficiency of binary classifiers. It is an ensemble learning method that combines various weak classifiers, usually decision trees with a single split known as decision stumps, in order to build one strong classifier. After each iteration, AdaBoost puts more weight into the data points that are harder to classify. In addition, classifiers are weighted based on their accuracy.

4.5.2.5 K-Nearest Neighbors

K-Nearest Neighbors (k-NN) algorithm is one of the simplest methods among the supervised machine learning algorithms. It was developed as a non-parametric classification method by Fix & Hodges (1951) , however, it can be used to overcome classification and regression problems. For the classification case, a data point is assigned to the class that it most common among its k nearest neighbors. For the regression case, value of the output is equal to the average of its k nearest neighbors. Since the most important hyperparameter is k, determining the optimal value for it may be a complex task. In addition, it does not work well with the large data due to computational costs and complexity in distance measurements. Similarly, as number of input variables increase, k-NN struggles with its predictions. This phenomenon is also known as the curse of dimensionality.

4.5.2.6 Artificial Neural Network

Artificial neural networks (ANNs) are computational models originating from an aspiration towards human brain. They were built to simulate the network of neurons in a human brain that processes information and perform computing activities.

Models used for the binary classification of survey results along with their fine-tuned hyperparameters can be found in Table 4.8. Performance metrics of these models can be found in Table 4.9.

Table 4.8 Machine Learning Algorithms for Binary Classification

Models	**Parameters**	**Value**
Logistic Regression	C	1
	penalty	l2
	solver	liblinear
	maximum number of iterations	200
k-Nearest Neighbor	K	9
	algorithm	brute
	weights	distance
Random Forest Classifier	minimum sample leaf	3
	minimum sample split	10
	maximum depth	80
	number of estimators	100
XGBoost Classifier	learning rate	0.01
	maximum depth	90
	number of estimators	100
AdaBoost Classifier	learning rate	1
	number of estimators	300
Artificial Neural Network	batch size	12
	epochs	100
	activation function	sigmoid
	optimizer	Adam

Table 4.9 Performance Metrics of the Models

Models	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
Logistic Regression	61.56	59.47	61.52
k-Nearest Neighbors	57.40	57.36	59.21
Random Forest Classifier	59.91	59.71	59.98
XGBoost Classifier	60.83	61.02	60.83
AdaBoost Classifier	61.41	61.27	61.42
Artificial Neural Network	64.06	62.19	62.03

Under normal circumstances, these accuracy values may not be satisfactory since they are less than the general threshold of %70. However, the primary focus of this research was to determine which attributes effect people's preferences while choosing a penalty shootout to watch. For random forest, XGBoost and AdaBoost models, feature importance was obtained by using fitted feature_importances attribute from the scikit-learn library. For logistic regression, feature importance was obtained by the best.coef_ attribute from the scikit-learn library. The most important features for each model can be found in Table 4.10. Using the estimated coefficients of the features, ranking of the feature importance as a combination of all models can be found in Figure 4.2.

Table 4.10 Feature Importance

Ranking	LR	RF	XGBoost	AdaBoost
1	Goal Difference (R)	Goal Difference (R)	Goal Difference (R)	Total Goal
2	Probability Difference (R)	Probability Difference (R)	Total Goal	Scoring Probability A (R)
3	Total Goal	Total Goal	Probability Difference (R)	Scoring Probability B
4	Scoring Probability A (R)	Scoring Probability B	Round	Round
5	Round	Scoring Probability A (R)	Scoring Probability A (R)	Probability Difference (R)
6	Penalty Turn	Round	Scoring Probability B	Goal Difference (R)
7	Scoring Probability B	Penalty Turn	Penalty Turn	Penalty Turn

Notes: (R) represents reversely proportional. It means, that feature negatively effects the response variable.

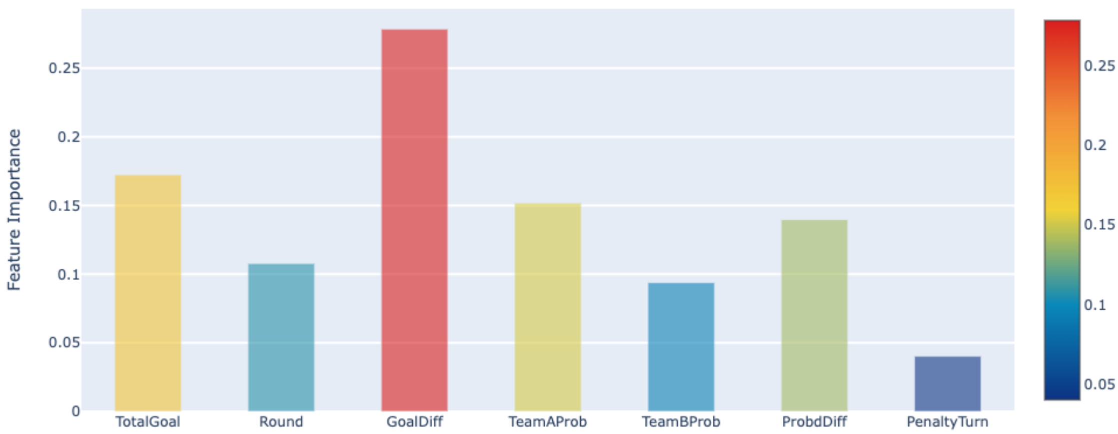

Figure 4.2 Combination of feature importance of all models

As it can be observed from table and graph shown above, the most important feature in distinguishing between two classes is the goal difference. Further analysis have shown that, as the goal difference between the two teams' penalty shootouts in a scenario increases, the participants' desire to watch that penalty shootout scenario decreases. Similarly, as the scoring probability difference between two teams in a scenario increases, the participant's desire to watch that penalty shootout scenario decreases. Thus, it can be concluded that people tend to watch penalty shootouts where the strengths of the two teams were similar. Another important factor affecting people's preference is the total number of total goals. As the total goal count in a penalty shootout increases, participants' desire to watch that penalty shootout also increases.

4.6 SHAP Values

Shapley additive explanations (SHAP) were developed by Lundberg & Lee (2017) in order to describe individual predictions and they are based on the shapley values from game theory. They aim to calculate the contribution of each feature in order to understand the prediction of each target data point. They can be considered as a measure of feature importance. SHAP feature importance suggests that the features with larger absolute shapley values are more important.

In order to find the total importance, we need to sum each absolute shapley value per feature in data. It is important to note that after determining feature importance in Section 4.5, we eliminated Penalty Turn feature from our list. That is why it is not included in graphs of SHAP.

4.6.1 Variable Importance Plots

SHAP has a function that creates summary plot of the feature contributions. The top features in the plot contribute more to the model than the bottom features which indicates higher predictive capability. In this section, we provide the summary plots for models created using Logistic Regression, Random Forest and XGBoost algorithms. Since SHAP library does not support AdaBoost, we could not include it to our study.

4.6.1.1 Logistic Regression Model

According to Figure 4.3, feature with the highest predictive capability is the goal difference, denoted as *GoalDiff* and feature with lowest predictive capability is the scoring probability of team B, denoted as *TeamBProb*. Figure 4.4 shows the effect of each feature in classifying a data point. For example, as the value of goal difference between two teams increases, impact on the model output becomes negative. This indicates that data points labeled as 0 have higher goal differences. In contrast, as the total number of goals increases, its impact on the model output becomes positive. This indicates that data points labeled as 1 have higher total number of goals.

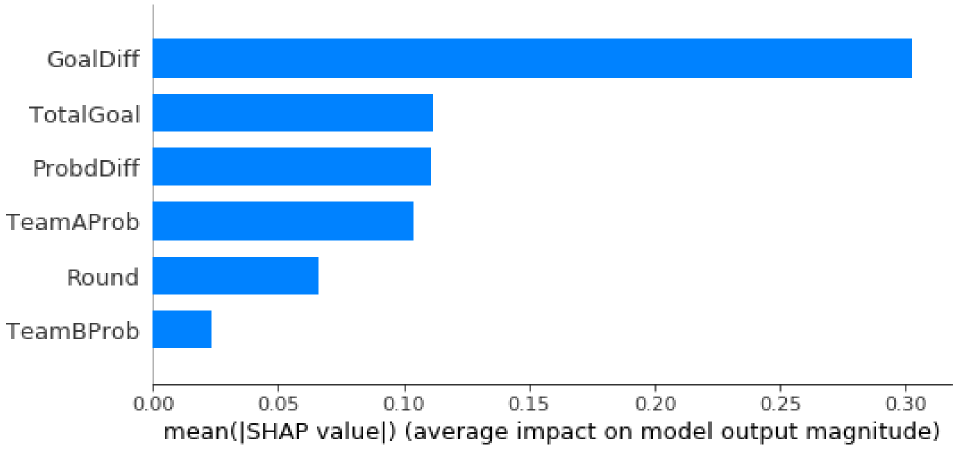

Figure 4.3 SHAP variable importance plot for logistic regression

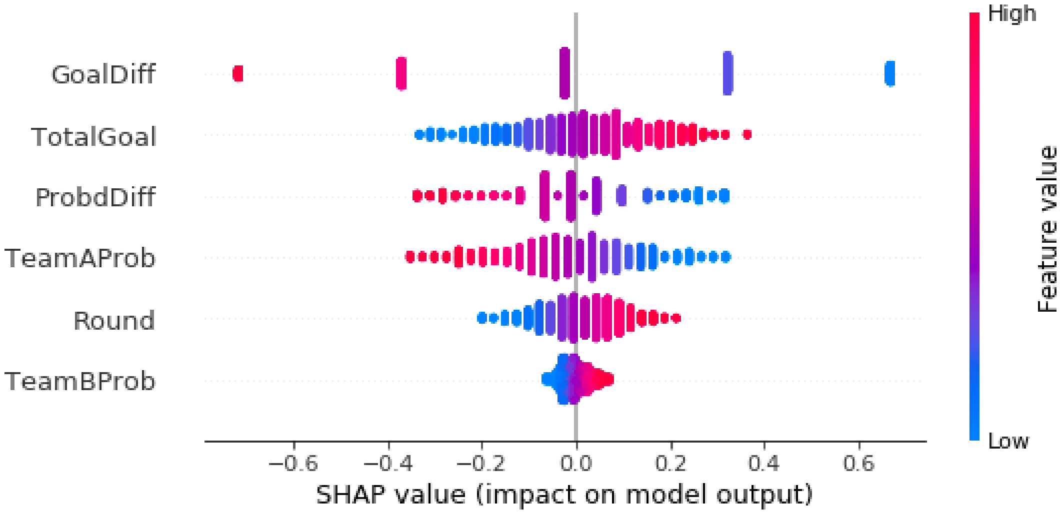

Figure 4.4 Detailed SHAP variable importance plot for logistic regression

4.6.1.2 Random Forest Model

According to Figure 4.5, feature with the highest predictive capability is the goal difference, denoted as *GoalDiff* and feature with lowest predictive capability is the scoring probability of team B, denoted as *TeamBProb*.

Figure 4.7 shows an individual feature importance detection case. Feature values in red causes an increase in the predicted value whereas feature values in blue causes a decrease in the predicted value. Size of the bar displays the magnitude of that feature's overall effect. The sum of these bars helps us to explain the difference between the predicted value and baseline value.

For this data point, model predicted 0.37 and the base value was equal to 0.4721.

Thus, since predicted value is smaller than the baseline value, this data point was labeled as 0. The most effective feature was goal difference and it dropped the chances of that scenario being selected. Similarly, the probability difference of 0.2 decreased chances of being selected along with the difference between scoring probability of team B of two scenarios being 0.3. On the other hand, total number of goals being 3 increased the chances of getting selected.

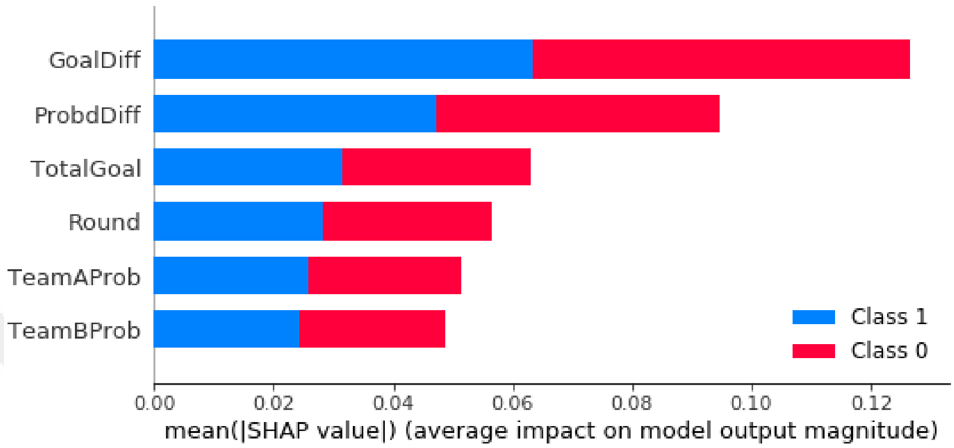

Figure 4.5 SHAP variable importance plot for random forest

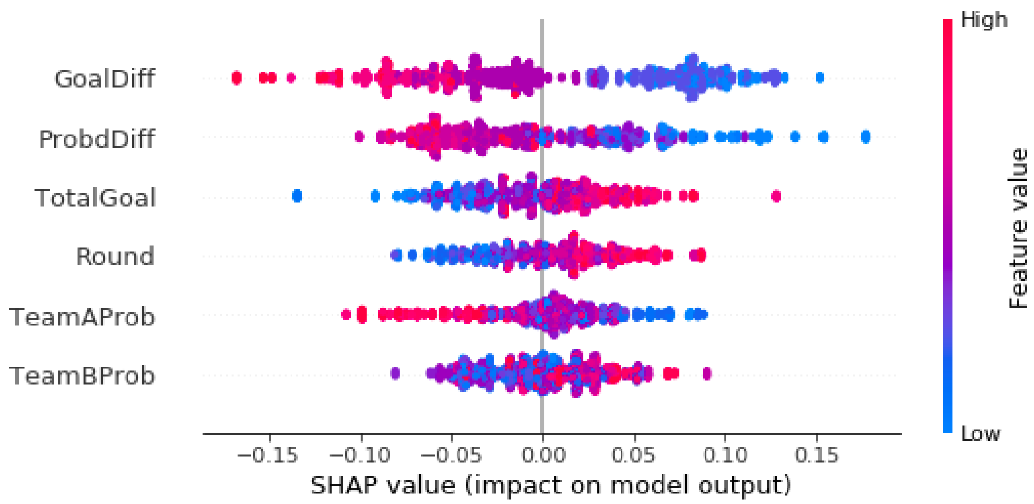

Figure 4.6 Detailed SHAP variable importance plot for random forest

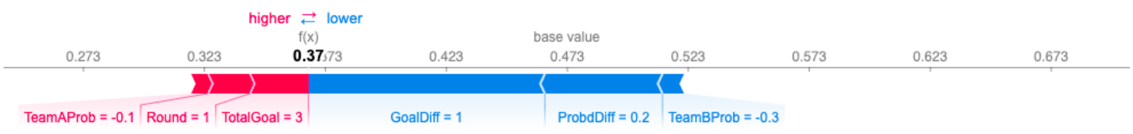

Figure 4.7 Individual feature importance example for random forest

4.6.1.3 XGBoost Model

According to Figure 4.8, feature with the highest predictive capability is the goal difference, denoted as *GoalDiff* and feature with lowest predictive capability is the scoring probability of team B, denoted as *TeamBProb*. Similarly, for Figure 4.9, the feature with the highest impact on the model output is *GoalDiff*. Higher goal difference effect the output negatively, which means a decrease in that scenarios' chances of being selected. *GoalDiff* feature is followed by the feature denoting the probability difference between two teams. According to Figure 4.9, similar to goal difference, as the probability difference between two teams increases, the chances of that scenario being selected decreases. Thus, *ProbDiff* has a negative effect on the output.

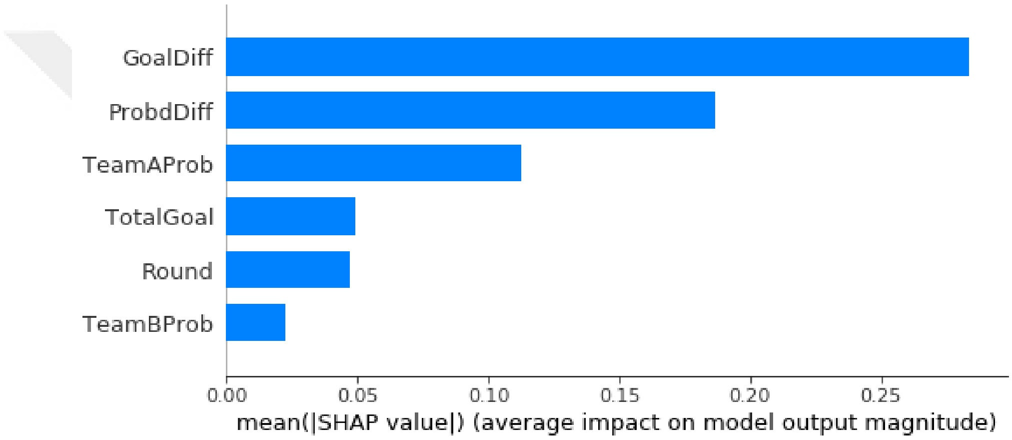

Figure 4.8 SHAP variable importance plot for XGBoost

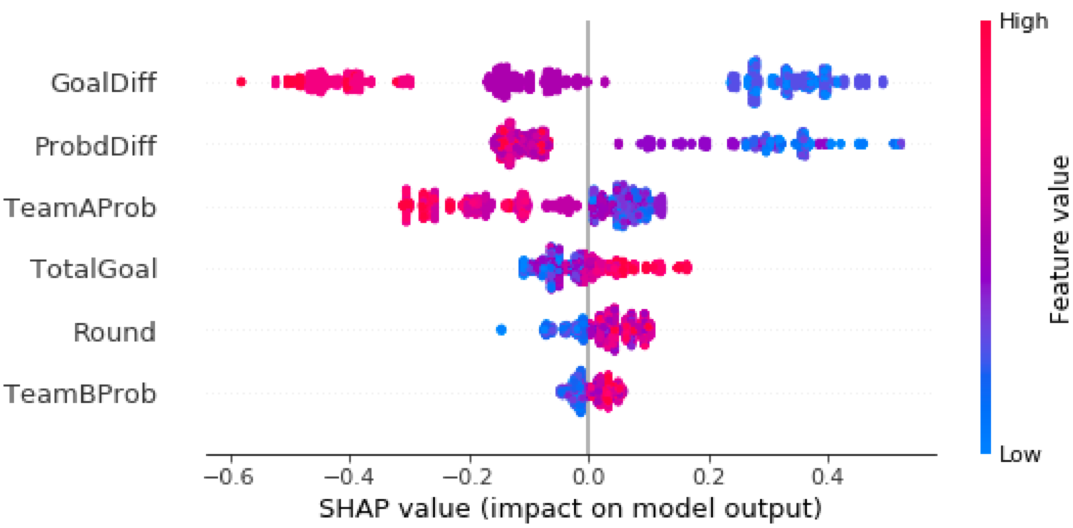

Figure 4.9 Detailed SHAP variable importance plot for XGBoost

4.7 Alternative Survey Structure and Feature Importance

In order to evaluate the survey from a different perspective, we formed our survey data with an alternative structure. This alternative structure consists of same features, however, their types are binary instead of numeric. This form of data can be considered as a summarized version of the previous one. As we have discussed in Section 4.1.2, the survey has 190 questions. Each of these questions compares two different scenarios and asks participants to prefer one of them. For the alternative structure, if, among these pairwise comparisons a scenario was selected more than its opponent, they will be labeled as 1. In addition, their features will be labeled as 0 or 1 based on the fact that whether they have higher feature values than their opponents or not. A portion of this data are shown in Figure 4.10.

	HigherTotalGoal	HigherRound	HigherGoalDiff	HigherTeamAProb	HigherTeamBProb	HigherProbDiff	Selected
0	1	1	1	1	0	1	0
1	0	0	0	0	0	0	1

Figure 4.10 Sample from alternative data

Figure 4.10 shows comparison of Scenario 1 and Scenario 2. Table 4.2 shows characteristics of each scenario. Accordingly, it can be seen that Scenario 1 has higher total number of goals, higher number of rounds, higher goal difference, higher team A probability and higher probability difference between its opposing teams. Team B probability is %50 in both scenarios, so it is labeled as 0. Finally, since Scenario 2 was selected more than Scenario 1, it is labeled as 1.

In total, this data consist of 380 rows and there are two rows per question. Similarly, classifier models were created for this data by utilising same machine learning algorithms except Artificial Neural Network. Along with their fine-tuned hyperparameters, classification models are shown in Table 4.11. Performance metrics of these models can be found in Table 4.12.

Table 4.11 Machine Learning Algorithms for Binary Classification with Alternative Data

Models	Parameters	Value
Logistic Regression	C	0.01
	penalty	l2
	solver	liblinear
	maximum number of iterations	100
k-Nearest Neighbor	K	7
	algorithm	auto
	weights	uniform
Random Forest Classifier	minimum sample leaf	4
	minimum sample split	12
	maximum depth	80
	number of estimators	100
XGBoost Classifier	learning rate	0.01
	maximum depth	3
	number of estimators	100
AdaBoost Classifier	learning rate	0.01
	number of estimators	1000

Table 4.12 Performance Metrics of the Models

Models	Accuracy (%)	Precision (%)	Recall (%)
Logistic Regression	72.63	73.58	73.26
k-Nearest Neighbor	73.08	76.94	73.68
Random Forest Classifier	73.68	74.19	73.95
XGBoost Classifier	71.05	71.44	71.00
AdaBoost Classifier	74.56	74.99	75.06
Naive Bayes Classifier	73.68	75.48	74.72

Without considering the numerical effects of the features, accuracy of the models are much higher. By creating a summarized version of the data, we had an opportunity to explore the feature importance more clearly. Again, the estimated coefficients of features can be used to find ranking of feature importance as a combination of all models. The result of this combination is demonstrated in Figure 4.11.

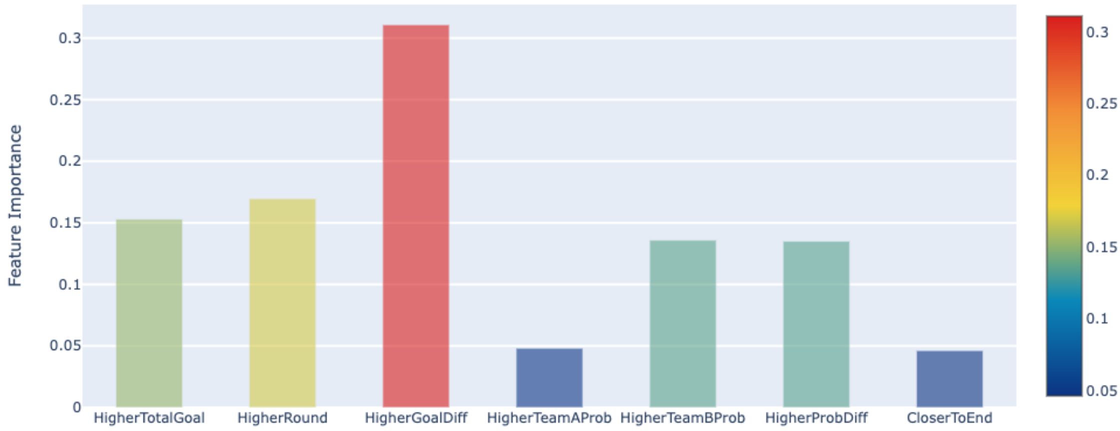

Figure 4.11 Combination of feature importance of all models using alternative data

According to Figure 4.11, higher goal difference is the main factor in the selection process. The total number of rounds and total number of goals can be considered as other crucial factors. Penalty turn, named as *CloserToEnd* for this case, is not relevant to the selection process. Thus, it is eliminated. It can be observed that people do not prefer to watch penalty shootouts with higher probability differences. In addition, they prefer penalty shootouts with higher total number of goals, i.e. higher number of successful penalties must be performed.

4.8 Results and Discussion

In this section, we display the final scenario rankings based on the voting results of survey participants, Bradley-Terry preference model and excitement score calculations in Table 4.13. The voting based ranking can be considered as a baseline and represents the total number of votes received for each scenario.

By comparing mathematical and empirical results, it can be concluded that the mathematical calculations were successful in determining the least exciting shootout scenarios. Conversely, while determining the most exciting shootout scenarios, mathematical results were quite different than the preferences of the viewers. Figure 4.12 shows the differences between four ranking models based on the excitement ranking. According to this figure, the most exciting scenario (i.e. Scenario 4) in excitement score model is the 9^{th} most exciting scenario for the remaining models. Similarly, the second most exciting scenario (i.e. Scenario 2) in excitement score model is the 13^{th} most exciting scenario for the alternative excitement model, 16^{th}

most exciting scenario for the Bradley-Terry's ability model and 17^{th} most exciting scenario for the voting based model.

In Figure 4.13, we display the differences between three ranking models based on the alternative excitement ranking. According to this figure, alternative excitement ranking is a lot more consistent with the voting based and Bradley-Terry preference rankings. Since alternative excitement ranking is based on the incremental excitement, it can be depicted that the viewers are more focused on the current round's excitement and they do not consider the remaining number of rounds during their selection process.

In general, the most exciting scenarios according to the viewers were Scenario 11, Scenario 17 and Scenario 19. Common features of these scenarios can be described as shootouts going into overtime, performing higher number of successful penalty shootouts and opposing teams with similar strengths. According to the viewers, excitement of a penalty shootout stems from a long rally among strong opponents. We further explain the similarities and differences between the mathematical and statistical analysis in the following paragraphs.

The mathematical calculations presented in this thesis to find the excitement score consider a penalty shootout sequence with higher variability in the winning probability as exciting. This means that recently started penalty shootouts tend to be more exciting, since winning probability of both teams will vary significantly during the course of shootouts. It is clear that each of the top 5 most exciting scenarios according to our excitement calculations were at the beginning of their shootouts as their current scores were in between 0-0 and 2-2 and their rounds had smaller values. In addition, if opposing teams had similar strengths with scoring probabilities that were close to 50%, that scenario will be considered as more exciting. Finally, similar to the preference of the viewers, shootout scenarios with less goal differences were found more exciting.

One of the main differences between viewers' ranking and mathematical ranking is that viewers perceive penalty shootouts with higher number of goals more exciting. They do not consider the remaining number of penalties left and the variability in winning probabilities of both teams. For viewers, a fierce competition among opposing teams finalising with a golden goal is more exciting than the suspenseful overall shootout experience. Additionally, competition among stronger teams is preferred rather than teams with average scoring probabilities.

Alternative excitement score formulation focuses on resolving the first difference and ignores the remaining number of penalty shootouts. As a result of that, ranking for the alternative excitement score calculation is very similar to the viewer preference rankings. Our first excitement score formulation predicts the remaining number of penalty shootouts by subtracting the number of penalties used by each team from $T = 5$, which is the regular number of rounds for a penalty shootout. However, from the survey results it can be understood that viewer's excitement accumulates each round and it may not be ideal to reduce it with an assumption. By omitting the remaining number of penalties, we were able to have results that are closer to the perception of the football viewers.

Table 4.13 Scenario Preferences

Rank	**Voting Based Ranking**	**Ability Based Ranking**	**Excitement Score Ranking**	**Alternative Excitement Score Ranking**
1	Scenario 17	Scenario 19	Scenario 4	Scenario 16
2	Scenario 11	Scenario 11	Scenario 2	Scenario 19
3	Scenario 19	Scenario 17	Scenario 8	Scenario 17
4	Scenario 15	Scenario 15	Scenario 6	Scenario 11
5	Scenario 3	Scenario 3	Scenario 14	Scenario 3
6	Scenario 16	Scenario 16	Scenario 19	Scenario 9
7	Scenario 9	Scenario 10	Scenario 17	Scenario 6
8	Scenario 10	Scenario 9	Scenario 18	Scenario 12
9	Scenario 4	Scenario 4	Scenario 11	Scenario 4
10	Scenario 6	Scenario 5	Scenario 3	Scenario 18
11	Scenario 7	Scenario 6	Scenario 7	Scenario 14
12	Scenario 18	Scenario 7	Scenario 12	Scenario 8
13	Scenario 8	Scenario 12	Scenario 16	Scenario 2
14	Scenario 12	Scenario 8	Scenario 9	Scenario 7
15	Scenario 5	Scenario 18	Scenario 10	Scenario 10
16	Scenario 2	Scenario 14	Scenario 15	Scenario 15
17	Scenario 14	Scenario 2	Scenario 5	Scenario 5
18	Scenario 13	Scenario 13	Scenario 13	Scenario 13
19	Scenario 20	Scenario 1	Scenario 1	Scenario 1
20	Scenario 1	Scenario 20	Scenario 20	Scenario 20

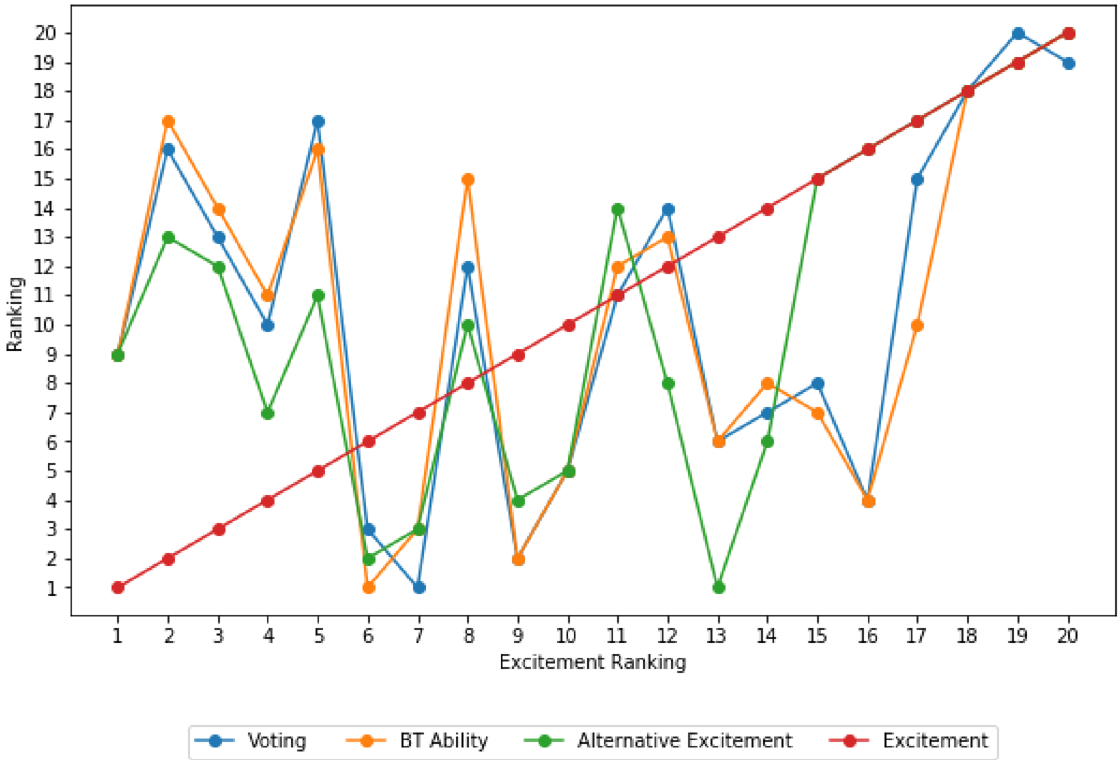

Figure 4.12 Comparison of the scenario rankings

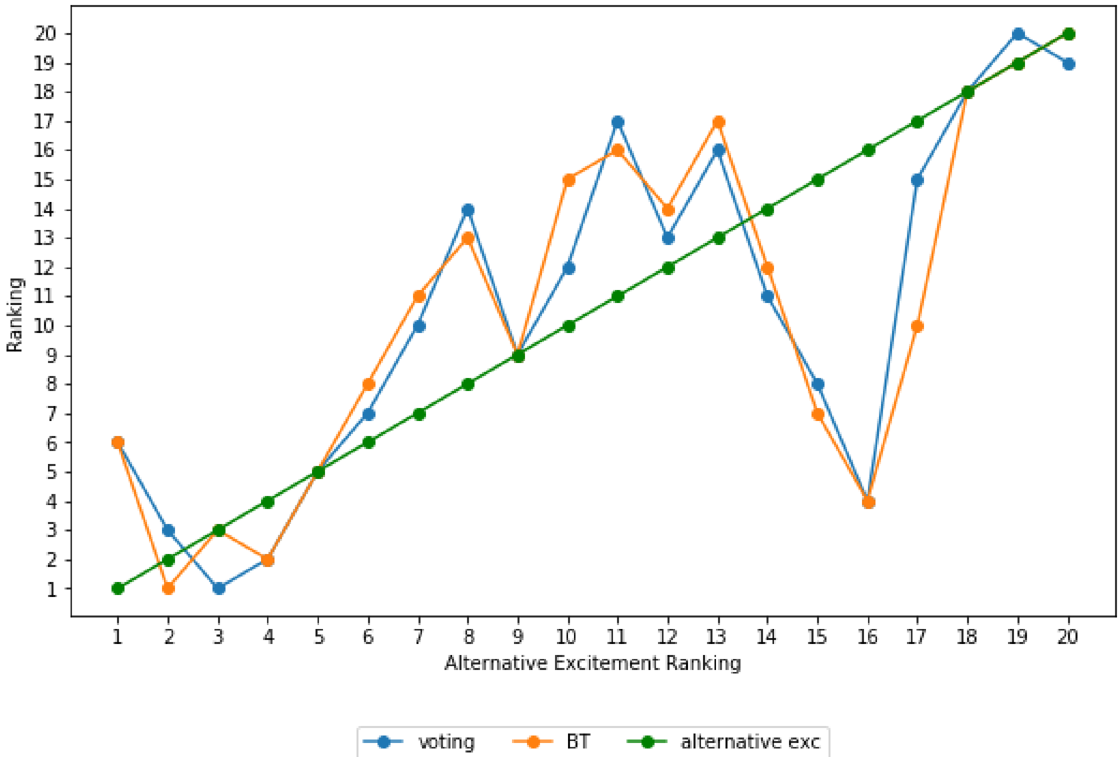

Figure 4.13 Comparison of the scenario rankings for alternative excitement

For the second part of this study, we focused on determining which factors were influencing the viewer's decision making process and compared them with the results of our mathematical analyses. Although we had some differences between mathematical calculations and statistical results, we were also able to find some similarities. Most of the time, features deemed as important by the viewers were also pivotal in ascertaining the excitement score.

Figure 4.14 represents the importance of each feature in each binary classification model by using their coefficients. The results for this study were obtained by using the survey data which consists of 4656 entries. As mentioned in Section 4.5, the most effective factor is the goal difference between two opposing teams and it is reversely proportional with the frequency of selecting a scenario. The second most important factor is the total number of goals, which is one of the main factors causing differences among the ranking results.

Correspondingly, the most influential factor in calculating the excitement score of a scenario is also the goal difference between two teams. From Table 4.13 it can be seen that least preferred scenarios and scenarios with the lowest excitement score are the same. The common feature of these scenarios (i.e. Scenario 1, Scenario 13 and Scenario 20) is that the goal difference between opposing teams is 2, which is the maximum allowable difference. From the mathematical point of view, higher goal difference suggests a lack of variability in winning probabilities which makes the final score of the shootouts very predictable. Thus, that penalty shootout scenario would be evaluated as not exciting. Similarly, it would be easier for the viewer to predict the outcome of a shootout with higher goal difference, which makes it less desirable to watch.

As a consequence of the structure of the survey, scoring probability of team A was usually higher than the scoring probability of team B. Since third most important factor is the scoring probability of team A and it is reversely proportional to the response variable, it can be observed that survey participants prefer to watch shootouts with teams that have similar strengths. Following very closely, fourth most effective factor is the scoring probability difference between opposing teams and it is reversely proportional to the response variable as well. Again, viewers prefer to watch shootout scenarios with teams that have a similar strength. Thus, increasing the scoring probability difference decreases the chances of that scenario being selected. This is another similarity with the mathematical excitement score calculation. As represented by Table 3.4 and Figure 3.2, increasing the scoring probability difference between teams causes excitement score to decrease significantly.

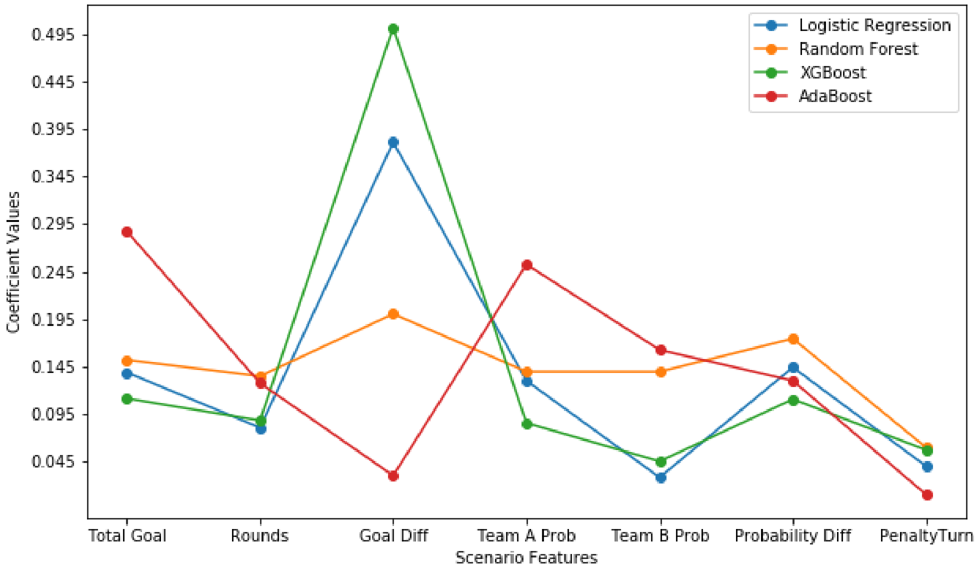

Figure 4.14 Feature importance of each classification model

Finally, Figure 4.15 shows the relationship between number of times a penalty shootout scenario is selected by the survey participants and its features as a heat map. By examining this heat map, we can observe a strong negative correlation between goal difference and frequency of choosing that scenario. We also notice a negative correlation between probability difference and frequency of choosing that scenario. On the contrary, as the total number of goals increase, the chance of that scenario being selected increases. Similarly, as the penalty shootout scenario goes into overtime, chances of that scenario being selected becomes higher. The heat map can be interpreted as a supplementary material and confirms the arguments that we have established in this section.

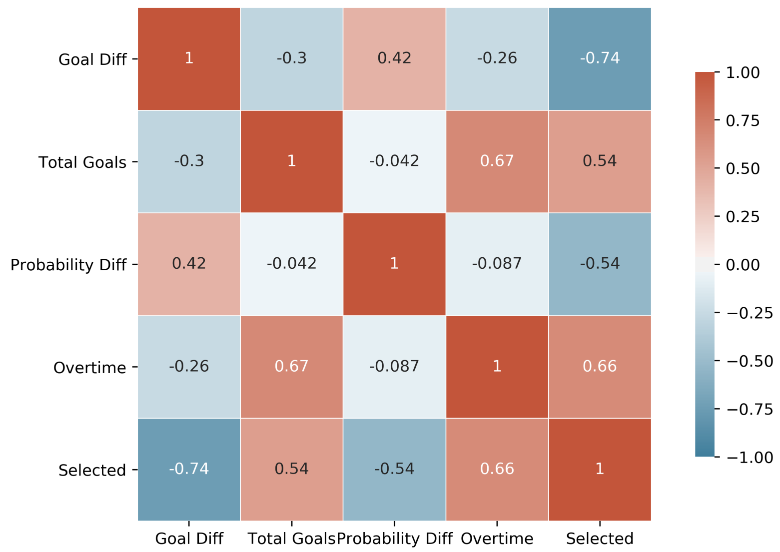

Figure 4.15 Correlation heat map representing relations between scenario features and selection frequency

5. CONCLUSION

The main objective of this thesis was to calculate the excitement score of penalty shootouts and explore an area of sports analytics that is often overlooked by the researchers. Furthermore, we aimed to understand which factors of a penalty shootout have a determinant role in viewer's selection process. To achieve these goals, we first calculated the expected excitement for penalty shootout scenarios in which the scoring probabilities at each around remains constant. In order to obtain more realistic results, we calculated the excitement for cases where scoring probabilities are unknown or they vary at each round. Then, we compared the mathematical results with the statistical analysis of a survey that we conducted with 310 participants. By using the results of this survey, we built several models utilising machine learning algorithms to find feature importance. Different survey structures and alternative mathematical excitement calculations were used to provide elaborated outcomes.

We display and explain the findings of our study in Section 4.8. We concluded that the excitement score calculations were consistent with the viewer's preferences while determining the least exciting penalty shootouts. For the most exciting shootouts, excitement score rankings differed from the viewer's preferences. We observed that this difference was due to the fact that viewers tend to focus on the current round of the shootouts, rather than considering the remaining number of rounds. To obtain similar results with the viewers' preference rankings, we altered the excitement score calculation. The results of the alternative excitement score calculations were more similar to the preference rankings since they were calculated by using the incremental excitement formulations. Finally, we found that the most important factor for the excitement score calculation and the viewers' was the goal difference between opposing teams.

For future work, the excitement score of an entire football match can be calculated and the methodology can be used to calculate the excitement of various sports. By understanding people's feelings towards sports games, we hope to improve the traditional game rules and make sports competitions more entertaining to the viewers.

BIBLIOGRAPHY

Abdulla, S. & Alashoor, A. (2020). An artificial deep neural network for the binary classification of network traffic. *International Journal of Advanced Computer Science and Applications*, *11*.

Bradley, R. A. & Terry, M. E. (1952). The rank analysis of incomplete block designs — I. The method of paired comparisons. *Biometrika*, *39*, 324–345.

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32.

Cattelan, M., Varin, C., & Firth, D. (2013). Dynamic bradley-terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(1), 135–150.

Chetprayoon, P., Iida, H., & Takahashi, R. (2017). An analysis of sports using game refinement measure. *ODISHA Journal of Social Science*, *4*, 4–48.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em - algorithm plus discussions on the paper.

Dittrich, R., Katzenbeisser, W., & Reisinger, H. (2000). The analysis of rank ordered preference data based on bradley-terry type modelsdie analyse von präferenzdaten mit hilfe von log-linearen bradley-terry modellen. *Or Spektrum*, *22*, 117–134.

Fix, E. & Hodges, J. (1951). *Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties*. USAF School of Aviation Medicine.

Guo, Z., Zhou, K., Zhang, X., & Yang, S. (2018). A deep learning model for short-term power load and probability density forecasting. *Energy*, *160*.

Hanjalic, A. (2005). Adaptive extraction of highlights from a sport video based on excitement modeling. *IEEE Transactions on Multimedia*, *7*(6), 1114–1122.

Ian, M. & Morton, A. (2011). A bradley-terry type model for forecasting tennis match results. *International Journal of Forecasting*, *27*(2), 619–630.

Iida, H., Takahara, K., Nagashima, J., Kajihara, Y., & Hashimoto, T. (2004). *An Application of Game-Refinement Theory to Mah Jong*, volume 3166, (pp. 333–338).

Iida, H., Takeshita, N., & Yoshimura, J. (2003). *A Metric for Entertainment of Boardgames: Its Implication for Evolution of Chess Variants*, (pp. 65–72).

Jia, H., Lin, J., & Liu, J. (2019). An earthquake fatalities assessment method based on feature importance with deep learning and random forest models. *Sustainability*, *11*, 2727.

Junki, T., Rido, R., & Hiroyuki, I. (2014). Game-refinement theory and its application to volleyball. Technical Report 3, School of Information Science, Japan Advanced Institute of Science and Technology, Department of Informatics, Institut Teknologi Bandung, School of Information Science, Japan Advanced Institute of Science and Technology.

Khullar, V., Salgotra, K., Singh, H., & Sharma, D. (2021). Deep learning-based binary classification of adhd using resting state mr images. *Augmented Human Research*, *6*.

Lee, G.-G., Kim, H.-k., & Kim, W.-Y. (2009). Highlight generation for basketball video using probabilistic excitement. In *2009 IEEE International Conference on Multimedia and Expo*, (pp. 318–321).

Lundberg, S. & Lee, S.-I. (2017). A unified approach to interpreting model predictions.

Nossal, N. & Iida, H. (2014). Game refinement theory and its application to score limit games. (pp. 1–3).

Pollard, G. (2016). Measuring excitement in sport. *Journal of Sports Analytics*, *3*, 1–7.

Sutiono, A., Purwarianti, A., & Iida, H. (2014). *A Mathematical Model of Game Refinement*, (pp. 148–151).

Vecer, J., Ichiba, T., & Laudanovic, M. (2007). On probabilistic excitement of sports games. *Journal of Quantitative Analysis in Sports*, *3*, 6–6.

Yamak, Z., Saunier, J., & Vercouter, L. (2015). Detection of multiple identity manipulation in collaborative projects.

APPENDIX A

Calculations for Equation 3.9

First component:

$$a(1-b)(1-p_A) = a(1-b)\left(1-\frac{(1-b)a}{a+b-2ab}\right)$$

$$(1-p_A) = \frac{a+b-2ab-a+ab}{a+b-2ab} = \frac{b-ab}{a+b-2ab} = \frac{b(1-a)}{a+b-2ab}$$

$$= \frac{ab(1-a)(1-b)}{a+b-2ab} \qquad (1)$$

Second component:

$$(1-a)bp_A = (1-a)b\left(\frac{(1-b)a}{a+b-2ab}\right)$$

$$= \frac{ab(1-a)(1-b)}{a+b-2ab} \qquad (2)$$

Third component:

$$2ab(1-p_A)(1-b) = 2ab(1-b)\frac{b(1-a)}{a+b-2ab}$$

$$= \frac{2ab^2(1-b)(1-a)}{a+b-2ab} \qquad (3)$$

Fourth component:

$$2(1-a)(1-b)bp_A = 2(1-a)(1-b)b\left(\frac{(1-b)a}{a+b-2ab}\right)$$

$$= \left(\frac{2ab(1-a)(1-b)^2}{a+b-2ab}\right) \qquad (4)$$

$(1)+(2)+(3)+(4) =$

$$\left(\frac{2ab(1-a)(1-b)+2ab^2(1-b)(1-a)+2ab(1-a)(1-b)^2}{a+b-2ab}\right)$$

$$= \left(\frac{2ab(1-a)(1-b)[(1+b+(1-b)]}{a+b-2ab}\right) = \frac{4ab(1-a)(1-b)}{a+b-2ab}$$

Raw Survey Data Sample for Chapter 4

TotalGoal	Round	GoalDiff	TeamAProb	TeamBProb	ProbDiff	PenaltyTurn	Selected
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	1
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0
4	4	2	0,7	0,5	0,2	0	0

Figure A.1 Raw survey data sample

Processed Survey Data Sample for Chapter 4

TotalGoal	Round	GoalDiff	TeamAProb	TeamBProb	ProbDiff	PenaltyTurn	Selected
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	1
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0
3	3	1	0,1	0	0,1	-1	0

Figure A.2 Processed survey data sample

Example Survey Questions

1. Which of these penalty shootout scenarios would you prefer to watch?

Table A.1 Example Survey Question 1

	Scenario 1	Scenario 2
Round	4	1
Score	3-1	1-0
Penalty kick in	Team A	Team B
Scoring Probability (Team A)	%70	%60
Scoring Probability (Team B)	%50	%50

2. Which of these penalty shootout scenarios would you prefer to watch?

Table A.2 Example Survey Question 2

	Scenario 3	Scenario 14
Round	5	1
Score	4-4	1-0
Penalty kick in	Team A	Team B
Scoring Probability (Team A)	%70	%80
Scoring Probability (Team B)	%60	%90

3. Which of these penalty shootout scenarios would you prefer to watch?

Table A.3 Example Survey Question 3

	Scenario 7	Scenario 19
Round	3	6
Score	0-1	5-5
Penalty kick in	Team A	Team A
Scoring Probability (Team A)	%80	%70
Scoring Probability (Team B)	%80	%70

Calculations for Example 3.1

For simplicity, superscripts are not written since they are the same as the example.

Table A.4 Calculations for Example 3.1

	Parameter - A	Value	Parameter - B	Value
t=1				
	$A_{0,0}(0,0)$	42.05%	$B_{0,0}(0,0)$	57.94%
	$A_{1,0}(0,0)$	26.61%	$B_{1,0}(0,0)$	34.60%
	$A_{1,0}(1,0)$	52.35%	$B_{1,0}(1,0)$	60.53%
	a_1	0.6	b_1	0.9
	$E_{1,1}$	0.1235	$E_{1,2}$	0.0466
t=2				
	$A_{0,0}(0,0)$	42.05%	$B_{0,0}(0,0)$	57.94%
	$A_{1,0}(0,0)$	24.56%	$B_{1,0}(0,0)$	40.65%
	$A_{1,0}(1,0)$	49.55%	$B_{1,0}(1,0)$	67.25%
	a_2	0.7	b_2	0.65
	$E_{2,1}$	0.1049	$E_{2,2}$	0.1210
t=3				
	$A_{0,0}(0,0)$	42.05%	$B_{0,0}(0,0)$	57.94%
	$A_{1,0}(0,0)$	27.73%	$B_{1,0}(0,0)$	39.25%
	$A_{1,0}(1,0)$	53.77%	$B_{1,0}(1,0)$	65.91%
	a_3	0.55	b_3	0.7
	$E_{3,1}$	0.1288	$E_{3,2}$	0.1119
t=4				
	$A_{0,0}(0,0)$	42.05%	$B_{0,0}(0,0)$	57.94%
	$A_{1,0}(0,0)$	22.75%	$B_{1,0}(0,0)$	44.63%
	$A_{1,0}(1,0)$	46.90%	$B_{1,0}(1,0)$	71.08%
	a_4	0.8	b_4	0.5
	$E_{4,1}$	0.0772	$E_{4,2}$	0.1322
t=5				
	$A_{0,0}(0,0)$	42.05%	$B_{0,0}(0,0)$	57.94%
	$A_{1,0}(0,0)$	30.14%	$B_{1,0}(0,0)$	41.87%
	$A_{1,0}(1,0)$	56.59%	$B_{1,0}(1,0)$	68.56%
	a_5	0.45	b_5	0.6
	$E_{5,1}$	0.1309	$E_{5,2}$	0.1281

$DE^{(0)} = 0.2612$

$$\begin{aligned}\mathcal{E} &= \sum_{t=1}^{5} [E_{t,1}(a_{1:5}, b_{1:5}) + E_{t,2}(a_{1:5}, b_{1:5})] + DE^{(0)} \\ &= 1.1041 + 0.2612 \\ &= 1.3653\end{aligned}$$