University of Cape Town

Statistics Honours

Honours Project

Fantasy Surfing Team Selection

Author:
Aiden Harris
Sebastian De Klerk

Student Number:
HRRAID003
DKLSEB001

Supervisor:
Neil Watson

October 22, 2024

Plagiarism Declaration

I, the undersigned, declare that:

1) I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2) Each significant contribution to, and quotation in, this project study from work of other people has been attributed, and has been cited and referenced.
3) This project is my own work.
4) I have not allowed and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

I understand that if evidence of plagiarism is found, appropriate actions will be taken in accordance with university policies.

Signature: A.Harris

Name: Aiden Harris

Signature: Sebastian.dk

Name: Sebastian de Klerk

Contents

1 Introduction **5**
- 1.1 Fantasy Sports . . . 5
- 1.2 Fantasy Surfing Rules Overview . . . 5
- 1.3 Championship Tour . . . 5
- 1.4 Wave Scoring . . . 6
- 1.5 Wave Breaks . . . 6
- 1.6 Objective of Research . . . 7

2 Literature Review **8**
- 2.1 Fantasy Sports . . . 8
- 2.2 Mixed Integer Linear Programming . . . 8
- 2.3 Professional Surfing . . . 8
- 2.4 Location and Environmental Variability in Competitive Surfing . . . 9
- 2.5 Player Selection in Team Sports . . . 9
- 2.6 Conclusion . . . 10

3 Exploratory Data Analysis **12**
- 3.1 Data Description . . . 12
- 3.2 Variable Description . . . 12
- 3.3 Ethical Clearance . . . 14
- 3.4 Data Manipulation . . . 14
- 3.5 Potential Variables of Interest . . . 14
  - 3.5.1 Surfer Stance . . . 14
  - 3.5.2 Number of Excellent Heats . . . 15
  - 3.5.3 Heats Surfed . . . 15
  - 3.5.4 Previous Event Fantasy Points . . . 15
  - 3.5.5 Maximum Heat Score . . . 15
  - 3.5.6 Championship Tour Experience . . . 15
  - 3.5.7 Event Placement . . . 15
  - 3.5.8 Nationality . . . 16
  - 3.5.9 Wave Break Type . . . 16
  - 3.5.10 Data for Rookies and Wildcard Surfers . . . 16
- 3.6 Variable Selection . . . 16
  - 3.6.1 Distribution of points across 2017 season . . . 16
  - 3.6.2 QQ plots and Shapiro-Wilk Test . . . 17
- 3.7 Variable Importance Plots . . . 18
- 3.8 Lasso Modelling . . . 21
  - 3.8.1 Generalised Linear Modelling . . . 21
  - 3.8.2 Cross Validation Curves . . . 21

4 Methodology **25**
- 4.1 Base Model Formulation (ML Approach) . . . 25
  - 4.1.1 Fundamental Decision Variables . . . 25
  - 4.1.2 Objective Function . . . 25
  - 4.1.3 Fundamental Constraints . . . 25
  - 4.1.4 Event 1 . . . 26
  - 4.1.5 Event 2 . . . 26
  - 4.1.6 Event 3 . . . 27
  - 4.1.7 Event 4 . . . 27

4.1.8 Event 5 . . . 28
4.1.9 Event 6 . . . 28
4.1.10 Event 7 . . . 29
4.1.11 Event 8 . . . 29
4.1.12 Event 9 . . . 30
4.1.13 Event 10 . . . 30
4.1.14 Event 11 . . . 31
4.2 Team Selection Strategy Model Formulation . . . 31
4.2.1 Optimal Team . . . 31
4.2.2 Random Team . . . 31
4.2.3 2016 Score . . . 32
4.2.4 Previous Event Fantasy Points . . . 32
4.2.5 Current Standings . . . 32
4.2.6 Wave Break Type . . . 32
4.3 Model Formulation based on Break Type . . . 33
4.3.1 Event 1 . . . 33
4.3.2 Event 2, 5, 7, 8 and 11 . . . 34
4.3.3 Event 3 . . . 34
4.3.4 Event 4, 9 and 10 . . . 35
4.3.5 Event 6 . . . 35
4.4 Top three statistics . . . 36
4.4.1 Event 1 . . . 37
4.4.2 Event 2 . . . 37
4.4.3 Event 3-11 . . . 38
4.5 Power Surfer Strategy Model Formulation . . . 38
4.5.1 Tier 1 Only . . . 38
4.5.2 Historical Winner . . . 39
4.5.3 Previous Event Optimal Power Surfer . . . 39
4.6 Alterations To Model Formulation . . . 40
4.7 Base Model Alterations for 2018 Season . . . 40
4.7.1 Event 3 Formulation . . . 40
4.7.2 Event 4 Formulation . . . 40
4.7.3 Event 5 Formulation . . . 41
4.7.4 Event 7 Formulation . . . 42
4.7.5 Event 8 Formulation . . . 42
4.7.6 Event 11 Formulation . . . 43
4.8 Reoccurring Stats Model Alterations . . . 43
4.8.1 Event 3, 6, 9 Model Formulation . . . 43
4.8.2 Event 4 Model Formulation . . . 44
4.8.3 Event 7 Model Formulation . . . 44
4.8.4 Event 8 Model Formulation . . . 45
4.8.5 Event 10 Model Formulation . . . 45
4.8.6 Event 11 Model Formulation . . . 46
4.9 Break Type Model Alterations . . . 46
4.9.1 Event 2 Model Formulation . . . 46
4.9.2 Event 3 Model Formulation . . . 47
4.9.3 Event 4 Model Formulation . . . 47
4.9.4 Event 5 Model Formulation . . . 48
4.9.5 Event 7, 11 Model Formulation . . . 48
4.9.6 Event 8 Formulation . . . 49

4.9.7 Event 9,10 Model Formulation 49

5 Results and Discussion **50**
5.1 2017 Championship Tour Season 50
5.1.1 Team Selection Strategies 50
5.1.2 Average fantasy points per tier 54
5.1.3 Power Surfer strategies 55
5.2 2018 Championship Tour Season 56
5.2.1 Team Selection Strategies 56
5.2.2 Power Surfer strategies 62
5.3 Strategy Ranking and Performance Comparison 63

6 Conclusion and Recommendations for further Research **65**

7 Appendix **68**
7.1 Appendix A1: World Surf League Fantasy and Public Github Page containing data for 2014 to 2016 WCT seasons 68
7.2 Appendix A2: General structure For Lasso Modelling 68
7.3 Appendix B: Tables of coefficient scores for lasso modelling 69
7.4 Appendix C: Nationality Table 82
7.5 Appendix D: University of Cape Town Research Ethics Committee Approval Letter 84

1 Introduction

1.1 Fantasy Sports

Fantasy sports have origins in North America dating back to the early 1960s when several friends sought fantasy baseball (Ploeg, 2021) as a diversion from the fear and isolation of World War 2 (Ruihley, 2021). The Fantasy Sports industry has grown substantially since and this has in-turn grown the sports economy, with increased viewership and net profit (Billings, 2013). Fantasy Sports are games where a player will construct a virtual team of professional players in a certain sport. Fantasy Sports have become a competition where players compete against each other and the winner is decided by their superior fantasy points tally. The team of professional players is chosen based on predictions in players' expected performance (Leishman, N.C., 2016). Each Fantasy Sport has its own points system and points are assigned based on each player's performance in the virtual team.

1.2 Fantasy Surfing Rules Overview

Competitive Surfing entered the world of fantasy sports for the World Championship Tour (WCT) in 2014. Fantasy Surfing is based on competitive surfer performance in real-life events on the World-Surf-League's World Championship Tour. The surfing fantasy competition is played on the World Surf League Fantasy Page (Appendix A). In every championship tour event, a player participating in the fantasy surf competition will select a team of surfers. Each player selects a team of 8 male surfers for every WCT event. Surfers are allocated a tier or ranking and these 8 surfers selected must adhere to tier constraints. A power surfer is selected, this surfer gets double their fantasy points. One of the eight surfers selected must be the power surfer. The player's fantasy points tally is the sum of all 8 surfers' fantasy points and the power surfer's fantasy points. The power surfer in fantasy surfing resembles the role of a captain or vice captain in fantasy rugby (Farley, 2014), fantasy football (Drayer, 2010) and other fantasy games. Such an athlete is chosen because the decision maker expects that player to earn their team more fantasy points. There are 11 events in the WCT season, therefore 11 fantasy teams are selected and the sum of all these teams' points contribute to the player's season fantasy score on the global leader board.

1.3 Championship Tour

The WCT is the pinnacle of competitive surfing, where the best 36 men and 24 women compete for a world surfing title. The WCT has 11 events in different locations across the world to make up a WCT season. In each event, every surfer will surf 6 to 8 rounds or heats. A heat is a 40 minute period of time where 2 to 4 surfers will compete in a pre-determined wave zone against other surfers they have been matched up against. There are 6 two-man heats and 2 three-man heats. The winner of every two-man-heat moves on to the next round, whilst the loser is eliminated from the event. The winner of the three-man round 1 and round 4 heats will skip round 2 and round 5 respectively. The second- and third-placed surfers of round 1 and round 4, are matched up in round 2 and round 5 respectively. Depending on a surfer's performance in a WCT event their number of heats surfed will differ. In each heat, a surfer can catch as many waves as they wish and every wave surfed is scored on a scale of 1 to 10 by a judging panel. The judging panel scores every wave subject to certain criteria: speed, power, flow and degree of difficulty. In every heat, a surfer's top two wave scores are summed for a heat score. The surfer with the higher heat score wins their heat and progresses to the next round. In the surfing fantasy sports game, a surfer's heat score for every round is summed for their fantasy points score in the particular event.

For every event in the WCT, the 36 male surfers competing in an event are known prior to the event starting. If one of the surfers is injured, there is an injury-replacement surfer that takes their spot. For every event, two surfers that are not full-time competitors will get an invitation to participate in a WCT event. Such surfers are known as wildcards and serve as a 'dark horse' since they have skill, knowledge and good past performances in the venue by winning the local trials competition a few days prior to the event starting. These wildcard surfers have a difficult run in the event since they get matched up against the highest ranked surfers on the WCT. In essence, for every event, there is a fixed number of 36 male surfer's competing. Fantasy Sports for competitive surfing differs from other sports. A line-up of 8 male surfers are chosen for each event, this entails that each event is a new slate where a new team can be selected. The ranking of surfers on the WCT changes after every event and thus one in general cannot select the same team for every event because the tier assignment changes after every event as the leaderboard changes. The Challenger Series is a season long competition where the top 80 ranked men and 48 ranked women compete for 10 qualification spots on the WCT for the following season. At the end of the WCT season, the top 22 men on WCT, two injury replacements, two event wildcards and the top 10 men on the Challenger Series are selected as the 36 surfers for the next season.

1.4 Wave Scoring

Waves are scored based on the combination of three types of maneuvers. Depending on the wave and the location, a surfer will perform a combination of certain maneuvers to get a good score. The three types are a tube ride, turn and an aerial maneuver. A tube ride section forms when a wave induces a siphon that creates a straight tube lying parallel to the incoming waves (Higgins, 1983). A surfer will attempt to fit their body inside this hollow tube-like section of the wave to obtain a good wave score. A tube ride section is dependent on the wave surfed and the location of the wave. A maneuver or turn is when a surfer shifts their weight onto a certain side of the surfboard causing a change in direction. The peel of the wave is the angle between the white water part of a wave and the unbroken part of a wave as it moves to shore (Scarfe, 2003). The turns a surfer performs have slight variations depending on the peel of the wave (Hutt, 2001). If the peel is very flat then the maneuver is a drawn out movement, known as a 'cutback', where the surfer returns to the power source of the wave. If the angle of the peel is steeper, a surfer tends to put more weight or power on the surfboard's rail with a more radical change of trajectory. Such maneuvers are termed "snap" and "carve".

An aerial maneuver is performed when a surfer generates speed and ramps off the wave. The surfer is launched into the air and lands back on the face of the same wave. Previous research into wave scores for the three types of maneuvers indicate that aerial maneuvers scored significantly higher than tube rides and other maneuvers (Forsyth, 2017). Forsyth et al. (2017) further concluded that aerial maneuvers had the lowest completion rate of the three types with a completion rate of 45%. This indicates that aerial maneuvers are the most risky type of maneuver, but if successful, score a surfer more points. An aerial maneuver is difficult to successfully land because a surfer needs to maintain contact with the surfboard and land in a stable position (Forsyth, 2017).

1.5 Wave Breaks

The 11 events on the WCT are surfed in three major wave-break types; Beach breaks, point breaks and reef breaks. Beach breaks break on shallow sand banks. These waves are powerful, hollow and fast. The maneuvers performed on the wave include tube rides, airs and turns. Three of the eleven events on the WCT are beach breaks, these include: Saquarema, Hossegor and Peniche. Point Breaks break on sand and rock. The waves are longer and offer sections for tube rides, turns and airs. Three of the eleven events on WCT are points breaks, these include: Jeffreys Bay, Bells Beach and Snapper Rocks. Reef breaks break on shallow coral reef. These waves offer large steep tube ride sections and on these waves tube rides and turns earn a surfer very high scores. Four of the eleven events on the WCT are reef breaks, these include: Pipeline, Cloudbreak, Margaret River, Lower Trestles and Tahiti.

1.6 Objective of Research

This study aims to build a model that selects the optimal team for every event on the WCT season. This model selects 8 surfers that form a line-up, including a power surfer for one's fantasy surfing team. The model selects the team which maximizes some objective subject to the constraints imposed. Additionally, this model serves as an analysis of surfer performance across the season and aids a decision maker in selecting an optimal team line-up by indicating which features are significant in determining surfer performance in a particular event. Various team selection and power surfer strategies will be evaluated to aid the decision maker in selecting a team of surfers to represent one's fantasy team.

2 Literature Review

2.1 Fantasy Sports

Fantasy Sports have become a large industry with large amounts of money being generated in revenue every year (Conley, 2018). Fantasy Sport is important to sport leagues, as it allows the participants of these games to interact with these sports, promoting engagement and developing deep relationships with their fans (Buser and Woratchek, 2021). In turn, these fans tend to be more loyal to their sports, due to the amount of engagement they have with the sport compared to regular fans (Buser and Woratcheck, 2021). The reason these fantasy games are able to do this is due to the pure commitment to games. Participants have to be consistent with their involvement, having to play every game week, match or event in the respective sport. The objective in fantasy sports is to maximize gain (Hunter, 2019). Gain can imply points, revenue or a form of resource with value in the game. Entries in fantasy sports are constructed based on constraints in resources available. A lineup is constructed based on metrics used to predict player performance. Many methods or indicators for analyzing player performance exist. Logistic regression and decision support analysis are approaches that support sports performance analysts (Wedding, 2022). Fantasy sports are very important for improving understanding of the system of interest and by using Fantasy Sports betting, one can better understand factors associated with the sport of interest. Predictions in fantasy sports make use of mathematical optimisation techniques such as mixed integer linear programming (MILP) to solve the problem of predicting a lineup of players to maximize one's objective.

2.2 Mixed Integer Linear Programming

MILP model formulations have been used in fantasy sports like NFL, football and basketball. The structure of MILP models are broken down into three major features; decision variables, constraints and the objective of interest. Decision variables are chosen by the decision maker and can take on continuous, integer or binary values. The constraints are a function of the decision variables and provide limitations to which decision variables or combinations of decision variables can be selected. The objective is expressed as a function of the decision variables to minimize costs or maximise gain (Karloff, 1991). Mixed integer linear programming is an approach where some objective function is maximized or minimized relative to certain constraints placed on the decision variables of interest which are constrained to be integer values. Previous literature that used MILP followed two major constraints. Firstly, lineups must be constructed within a fixed budget. Secondly we assume that each athlete's fantasy points are independent (Hunter, 2019). If these constraints can be met, then mixed integer programming assists in finding the optimal solution for a complex problem with various constraints. Other methods do exist. One can use trial and error, logistic regression, gradient boosting and other analytical methods. Literature on Fantasy Sports have shown that mixed integer linear programming is an effective and successful approach.

2.3 Professional Surfing

Professional surfing is a very under-researched sport. Professional surfing began in 1976 and prior to the 21st century, did not capture the public's interest (Booth, 1995). This is because of a struggle to find sponsors and the limited popularity of the sport (Booth, 2024). Women's competitive surfing also had very low prize money and so many athletes could not make a living from surfing alone. The prize money and viewership of professional surfing has increased post 2010 with technological and surfing equipment advancements. Professional surfing has thus become a sport for which fantasy sports games and betting have become associated.

Fantasy sports involves selecting a line-up of athletes that have the highest chances of winning. There is a lack of research on fantasy sports games for professional surfing and the features that affect how well a surfer will perform in a surf competition. Previous literature indicates an influence of the variables swell size, swell period, currents, wave frequency or length, beach type and break, geography, and environmental conditions (Farley, 2011). Surfing performance can be highly variable in nature and thus it is difficult to predict an individual's success from one event to another since each event can vary in the features mentioned. Farley et al. (2011) concluded that in competitive surfing athletes, there is a significant relationship between seasonal rank and their maximal power exerted. Farley et al. (2011) further concluded that this in turn increased surfer performance because increased power improves paddling and surfing on a wave that others lacking this power would miss. The extent to which these features influence a surfer's performance is unknown. There is huge importance in finding measures for these factors on selection of a surfer into one's fantasy line-up.

Previous literature on professional surfing is physiological. Studies of male surfers have found that increased muscularity and lower levels of body fat are associated with improvements in competitive ranking (Barlow et al., 2016). The number of waves ridden in a heat, the length of the rides, and activity levels were significantly related to heat placement and competition progress (Barlow et al., 2016). This result suggests that surfers should aim to maintain competitive pressure through actively seeking scoring waves over their opponents. This result also concludes that wave count in a heat is a significant feature to achieve success. There is no literature on Fantasy Surfing, so this paper will be a first in the world of fantasy sports. This coincides with the objectives of this study.

2.4 Location and Environmental Variability in Competitive Surfing

Previous literature has indicated that wave characteristics are highly variable. Waves break with varying intensities (Hutt, 2001). Wave peel rate describes the speed at which a wave breaks or formally known as the lip of the wave approaches. The angle at which a wave peels affects the maneuver type that a surfer does (Hutt, 2001). In professional surfing, surfers will perform in a wide variety of countries with different wave breaks. These different wave breaks will have different peeling rates and peeling angles because they differ in the material that the waves break upon (Hutt, 2001). The level of a surfer's skill is a determining variable required for surfing waves of different heights, peeling rates and peeling angles (Hutt, 2001). Hutt (2001), studied the effect of surf breaks in Hawaii and observed that since the waves break on shallow coral reefs, there exists greater potential for fast rides.

Mendez-Villanueva et al. (2010) studied the consistency of competitive surfing performance for 11 events in the World Championship Tour (WCT). The study looked at wave scores for 46 male surfers on the WCT and the within-surfer variability (termed Cohen Effect Size) differed, ranging from 0.72 to 1.01. These results indicated large variability in competitive surfing and the conclusion that the competition outcomes are largely unpredictable. The variability in surfer performance between successive events on the WCT were found to be moderate.

2.5 Player Selection in Team Sports

Player quality in team sports has evolved through training programs and the mastering of tactical techniques or skills (Trninic, 2008). It follows that athletes that have had longer training experience, tend to be more mature and will perform better in team plays. Trninic et al. (2008) created a player selection model where player performance in practice sessions and matches are compared to assess performance consistency. The model is a continuous measure of player performance by comparing the player's potential with their actual quality observed. This model facilitates decision making and can be generalized to assess a player's quality to different plays. Mendez-Villanueva et al. (2010) supported the statement that players evolve from training and learning new skills, however in the WCT, small to moderate changes in player performance are only observed between several events. This indicates that the improved performance in competitive surfing as a result of training is a slow, gradual process.

Tavana et al. (2013) created a Fuzzy Inference System (FIS) for team selection in multi-player sports. The model derived from Lang et al. (1992), converts input data into performance metrics like skill level and team work. The metrics are termed "fuzzy" since they are broken down into high, medium and low. Tavana et al. applied this model to soccer and used FIS to select the best team formulation combination. Al Shaboul et al. (2017) created a Competitive Neural Network that selects the best possible team of 11 players in a football or soccer match from a pool of available players. The model had 60% accuracy and motivated a supervised learning approach to predict and measure player performance. Al Shaboul et al. (2017) used player scores and ratings that were composed of player features including: age, wins, goals scores, goals conceded and red cards. The 60% accuracy score was calculated by $\sum \frac{(TP+FP)}{Total}$, where TP denotes the True Positive Rate and FP denotes the False Positive Rate. These two studies indicate two alternatives one can take in creating a model that predicts player performance and both can ultimately be used for fantasy sports. Becker et al. (2013) built a mixed-integer optimization model to predict team and player performance on a football pitch. The MILP model's objective maximized the total points scored and games won that relate to winning a fantasy football season based on historical data for the 2007 and 2008 football season. In fantasy football, each player has a starting line-up and four reserve players. The reserve players do not contribute to one's fantasy points and Beck et al. (2013) used a greedy approach to select the starting line-up for each week of fantasy football. The model was assessed against real-life opponents and achieved an average rank of 5.14 with the model improving draft decisions and draft performance.

Robinson (2020) predicted NFL player performance. The fantasy points for 2020 NFL fantasy football season were calculated using an Autoregressive Integrated Moving Average model.(ARIMA model). Players selected required a minimum of 16 games and the algebraic expression calculated a fantasy score for every player. An error measurement (Mean Absolute Percentage Error) of 4.56% was obtained, indicating success in player selection. The models in Robinson (2020) and Becker et al. (2013) were both negatively affected by player injuries since injuries occur throughout the season and result in players missing several games, these players then earn one's fantasy team 0 points.

Beal et al. (2020) used mixed-integer programming and artificial intelligence (AI) on NFL 2014-2017 seasons to predict player performance and form an optimal fantasy team. Several models were presented and MILP solutions outperformed the other models for 81.3% of game-weeks over the season. Beal et al. (2020) concluded that MILP team formulation models outperform simulated human results and MILP in conjunction with machine-based methods increase prediction accuracy by a further 12.2% with increased profits outputted.

2.6 Conclusion

In conclusion, several literature sources have indicated that MILP is an efficient model type for predicting player performance and maximizing one's fantasy points scores. Several other approaches and models have been conducted previously in fantasy team sports like soccer, NFL and cricket. No models have been conducted for fantasy surfing and this study aims to be the first in that field. Previous literature on competitive surfing on the WCT has been psychological and indicates location/environmental variability in the sport. The MILP model will incorporate all rules and constraints in fantasy surfing and exploratory data analysis will assist in exploring what features are important in predicting surfer performance on the WCT. two previous studies (Farley et al. 2011 and Barlow et al. 2016) have found several important features that affect surfer performance. This study aims to build on the work done previously to find further features that are important in predicting surfer performance. The studies mentioned provide useful insight into the world of competitive surfing and fantasy sports and assist this study in developing a model for fantasy surfing.

3 Exploratory Data Analysis

3.1 Data Description

The data used in this study was obtained from a public GitHub (Appendix A1: Rayblick, 2017). The data included heat and wave scores from the 2015-2017 World Surf League Championship Tour seasons. The 2018 and 2014 Championship Tour season wave and heat scores were data-scraped from the official Championship Tour World Surf League website (Appendix A1). The heat scores were manipulated into two separate excel files. One file was named 'leaderboard', and had 22 sheets, each event had a pre- and post-event sheet to track tier rankings for the surfers. The second excel document, named 'Historical Event surfer stats', had 11 sheets, one per event, containing surfer statistics obtained from 2014-2016 WCT seasons. Exploratory data analysis was conducted on the 2017 season to build MILP models and develop team-selection strategies. These models were tested on the 2018 WCT season. The 2014-2016 WCT seasons were used as historical data in determining the important variables that predict surfer performance in the 2017 WCT season.

3.2 Variable Description

There are 20 variables in the data and a description of each variable is included in Table 1 and Table 2. These variables were selected based on previous literature, surfer statistics on World Surf League website and ethical constraints.

Table 1: Description of Surfing Data Attributes (Part 1)

Attribute	**Description**
Year Average	The average heat score of the surfer in that year for the specified event. This is calculated as the total score over the number of heats surfed.
Score	The total fantasy/event score for that event for the specified year.
Max Heat Score	The maximum heat score that the surfer achieved for that event over the previous 3 years.
Previous Heat Score	The fantasy/event points accumulated over the previous event in the same season.
Average Heat Score	The average event/fantasy points over the previous 3 events (or the previous 2 events for event 3).
Championship Score	The surfer's championship score prior to the event taking place. This denotes that surfer's ranking in the Championship Tour leader board.
Years Surfing	The number of years that a surfer has been surfing in the championship tour.
Experience	Categorical variable: "Rookie" (first year in the championship tour), "Experienced" (less than 5 years), "Veteran" (more than 5 years).
Actual Points	The fantasy points the surfer achieved for that event. This is the sum of all heat scores that surfer has in the particular event.
Tier	A surfer's ranking based on the Championship Tour leader board.

Table 2: Description of Surfing Data Attributes (Part 2)

Attribute	**Description**
Surfing Stance	How a surfer positions themselves on the surfboard. 'Goofy' footers ride waves with their right foot forward while 'Regular' surfers ride waves with their left foot forward.
Heat Winning Percentage (HWP)	This is the surfer's heat winning percentage for the specified event. Calculated as the number of heats won over the total number of heats entered. If a year was specified, it would reflect the heat winning percentage for that year. If overall was specified, it is the average heat winning percentage over the 3 years.
Heats Surfed 2016	This is the total number of heats that a surfer surfed in the 2016 season of the Championship Tour.
Local Surfer	Binary variable that determines if a surfer is local to the area in the country where the event is taking place. 1 if local, 0 if not.
Nationality	The country that a surfer represents. Appendix C contains all nationalities in this study.
National Advantage	Binary variable that determines if a surfer is from the country where the event is taking place. 1 if a citizen, 0 otherwise. The following countries had potential national advantages since an event was held there: France (FRA), United States (USA), Australia (AUS), South Africa (ZAF) and Brazil (BRA).
Placement	This is the surfer's placement in the event of the specified year. Available placements were 1, 2, 3, 5, 9, 13, and 25. Surfers were allocated a placement of 99 if no data was available for the event.
Number of Excellent Heats	This is the number of heats in which a surfer achieved a total heat score over 16 points. The World Surf League classifies these as "Excellent Heats."
Average Heat	This is the surfer's average heat score across previous years of the same event.
Total Heats Surfed	This is the total number of heats that a surfer has surfed at the event over the past 3 years.
Total Heats Surfed (Per Year)	Total number of heats surfed in that event in the specified year.

3.3 Ethical Clearance

Ethical Clearance was granted by the University of Cape Town Research Ethics Committee (SCI/ 00938/ 2024), (Appendix D). All surfer names were anonymous prior to exploratory data analysis and modelling in this study. The surfer names were replaced with 36 fictional names. No personal information on the surfers in this study were explored or shared.

3.4 Data Manipulation

Tiers were assigned based on each surfer's position in a WCT event. 10000 points for 1st, 8000 points for 2nd, 6500 points for 3rd, 5200 points for 5th, 4000 points for round 5, 1750 points for round 3 and 500 points for round 2. Every surfer's Championship Score points tally was found post and prior for all 11 events in 2017 and 2018 WCT season and the leader board was constantly changing after each event. Tiers 1,2 and 3 were assigned based on the ranking of every surfer's Championship Score. The top 8 surfers were assigned tier 1, 9th to 24th ranked were assigned tier 2 and 25th-36th ranked were assigned tier 3. After every event, the leader board was updated and the tiers were found for the next event. 36 surfers participate in each event, this is a hard constraint. If a surfer is injured, there is an injury replacement surfer that will participate. An injured surfer withdraws prior to the event starting and so this does not impact one's fantasy team. If two surfers are injured then there is a back-up injury replacement surfer. The MILP model will only consider surfers participating in the event and so injuries or surfer withdrawals are not impacting surfer selection. All surfers that participate are known before the event begins. If a surfer is injured and misses an event, their Championship Score ranking increases by 500 points. This is equivalent to a 25th placement or 2nd round exit.

3.5 Potential Variables of Interest

The potential variables of interest (Table 1 and 2) were further evaluated based on the results in 2014-2016 WCT seasons. Exploratory Data Analysis was conducted on these variables to justify their inclusion in this study.

3.5.1 Surfer Stance

Surfer stance is a binary variable with two outcomes: Regular or goofy. Regular-footed surfers surf with their left foot forward. Goofy-footed surfers surf with their right foot forward. These types of stances provide an advantage in some cases because goofy-footers surf with their back facing the wave on right hand wave-breaks: J-bay, Bells Beach, Snapper Rocks. Regular-footers surf with their back facing the wave on left hand breaks: Fiji, Tahiti, Pipeline. Facing towards the wave gives an advantage in speed. Facing ones back to a wave gives an advantage in doing very steep, powerful and high-risk maneuvers. These maneuvers are very hard to do and can put surfers with this stance in a disadvantage.

In 2014 and 2016, three of 11 events in the WCT season were won by a goofy-footed surfer (Furley, 2018). In the 2015 WCT season, two of 11 events were won by goofy footed surfers. The magnitude of the disadvantage for a certain surfer stance is not apparent since surfers participating on the WCT are of the most skilled and technically gifted (Furley, 2018). Another key consideration is that on average 12 of the 34 surfers are goofy (excluding wildcards and injury replacement surfers), so in general, less surfers are goofy-footed. Surfer stance has the potential to be a key variable of interest in this study and we aim to observe and reflect on the relationship between surfer stance and fantasy points obtained.

3.5.2 Number of Excellent Heats

An excellent heat is a heat score $\geq$ 16 points. Such a heat score entails that a surfer has caught at least 1, in most cases 2, waves of excellence and have surfed the wave very well. A surfer in one's fantasy team that scores an excellent heat rewards the team with lots of points. Surfers with many excellent heats mean that in the same event in 2014-2016, the surfer performed very well and would potentially have been in the optimal team for that particular event historically. These surfers serve as potentially good candidates for selection in the 2017 event because they have surfed very well in the same event in the past. Surfers with a large number of excellent heats also are good candidates for the power surfer selection because they have obtained a large number of fantasy points historically.

3.5.3 Heats Surfed

This variable represents the number of heats surfed in the 2014, 2015 and 2016 WCT seasons. More heats surfed entails that the surfer has progressed far into the event in the three previous events. A total of 8 rounds (heats) in an event are surfed so if historically the surfer has surfed many heats then they potentially have progressed to the final/semi-final. 2 heats surfed indicates that the surfer went out in the second round and thus did not perform well in the event historically. A larger number of heats surfed entails that historically the surfer potentially was in the optimal team for that event and are a reliable pick for the 2017 event.

3.5.4 Previous Event Fantasy Points

Previous Event Fantasy Points are used in events 2-11. This variable comprises of the surfer's fantasy points accumulated so far prior to the event. A large value indicates which surfers performed well in the event prior and are thus in good form coming into the event. Each event is a new slate where an optimal team can be selected, so the results of the previous event are assumed known to the decision maker prior to team selection.

3.5.5 Maximum Heat Score

Maximum heat score in the event historically. A large maximum heat score indicates that the surfer is in tune with the waves in that particular event in the past. A very large heat score usually wins a surfer a heat and so if they're maximum heat score is very large then in the past the surfer has performed very well in the event historically

3.5.6 Championship Tour Experience

Experience is a factor variable with three levels: Rookie, Experienced, Veteran. Surfers are assigned a category based on the number of years surfing in the WCT. Rookies have surfed 0 years on tour and so that particular season represents their first season competing on WCT. Surfers that have been on the WCT for 1-4 years are experienced and $\geq$ 5 years are veteran. A veteran has surfed this event in the past multiple times and thus has the experience of surfing that event in the past on the WCT level. A rookie is surfing the event for the first time on the WCT level. More data exists for experienced and veteran surfers than rookies.

3.5.7 Event Placement

The placement of each surfer in the event from years 2014-2016. Placements are awarded by, if the surfer was the winner of the event = 1st, runner-up = 2nd, Semi-final = 3rd, Quarter-final = 5th. Round 5 = 9th, round 3 = 13th, round 2 = 25th. Round 4 is a three-man heat where no-one is eliminated, the winner of round 4 skips round 5. The winner of round 4 moves straight to the quarterfinals whilst the losers go to round 5. A low event placement indicates that the surfer has performed well in that event in the last three seasons.

3.5.8 Nationality

Nationality was considered but we decided to not include it to due to ethical reasons as well as bias to selct those surfers. There are multiple surfers where they are the only representative of their country so the model would select these surfers if they performed well or poorly.

3.5.9 Wave Break Type

All events were categorised by three wave types: beach break, point break and reef break. The results of surfers in the same wave break were compared across events. This was done to observe the influence of wave break type on a team's fantasy points.

3.5.10 Data for Rookies and Wildcard Surfers

The data obtained for the WCT seasons 2014-2016 lacked data for many rookie and wildcard surfers. Many surfers that had WCT experience greater than two years had historical data. Several Rookie and Wildcard surfers were surfing the event for the first time, so depending on the surfer's experience, we either used their Challenger Series results from previous years in the same event or used the average rookie scores for that specific year on the WCT.

3.6 Variable Selection

The features mentioned are selected based on the relative influence of said features on the fantasy scores a surfer obtains. Several statistical methods will be examined. The distribution of the response variable, Actual Points, must be evaluated to determine if linear or non-linear models must be implemented. We will test if our data is normally distributed by producing a histogram of all data points throughout the 2017 season, Quantile-Quantile (QQ) plots and the Shapiro-Wilk test.

3.6.1 Distribution of points across 2017 season

The distribution of fantasy scores for all surfers in the 2017 WCT season were plotted to assess normality of the variable Actual Points.

The data does not resemble a normal distribution, since it is heavily skewed to the right or positively skewed (Figure 1). Most surfers appear to get between 20-30 points. This makes sense as over half of the surfers are eliminated by the 3rd round meaning that most of these surfers achieve points in this range. The large tail to the right is for surfers who have surfed multiple rounds by making it far into the event or participating in many elimination rounds where they would receive points. The distribution of fantasy points is subjective, so various plots and stern statistical methods will be evaluated to determine if it is normally distributed.

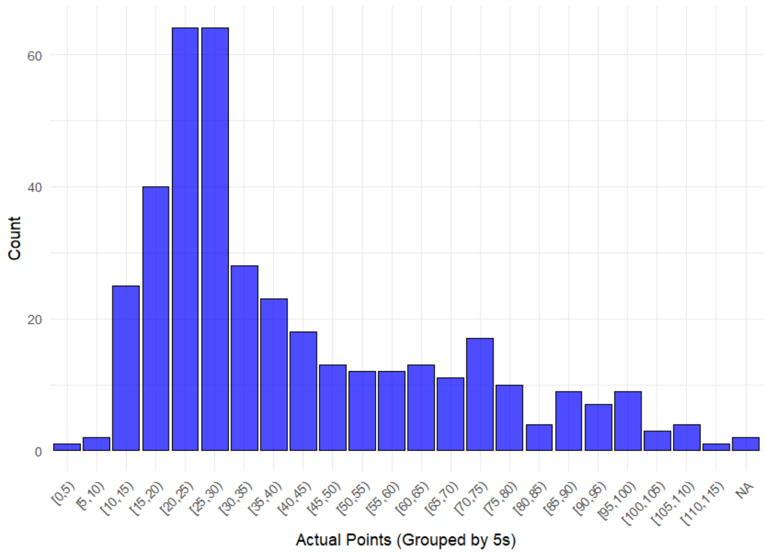


Figure 1: Distribution of Surfer Points for 2017 Season

3.6.2 QQ plots and Shapiro-Wilk Test

The Quantile-Quantile plot displays the quantiles of the distribution of the data against quantiles of a normal distribution (Das et al., 2016). We expect our data points to fall on the normal distribution $y = x$ line.

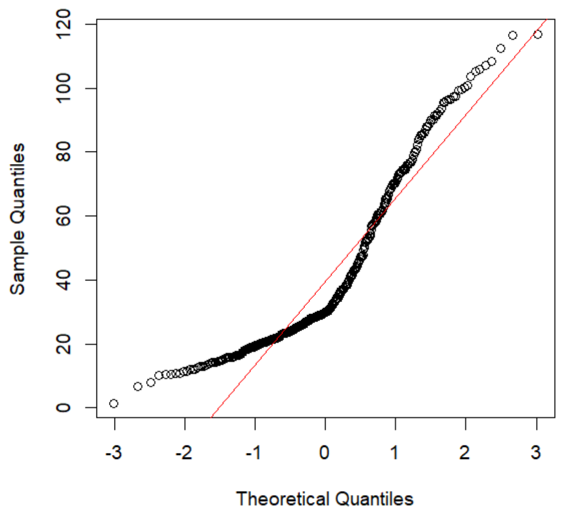


Figure 2: Quantile-Quantile Plot

Our data points deviate significantly from the normal distribution (Figure 2: The red line) indicating that our data is not normally distributed. The Shapiro-Wilk test will be used in conjunction with the QQ-plot to test for normality. The Shapiro-Wilk test compares the expected values of our data to our actual data points (Das et al., 2016). This tests the correlation between our expected value and our actual value and expresses this as a correlation metric. We then conduct a null hypothesis test to see if our data is normally distributed, with the null hypothesis being that it is normally distributed. If our test statistic, w_{test}, is significantly less than 1, we assume that it is not normally distributed.

Statistic	**Value**
Test Type	Shapiro-Wilk Normality Test
w_{test}	0.878
p-value	$\leq$ 2.2e-16

Table 3: The Shapiro-Wilk Normality Test on the Actual Points observed.

Our test for normality returns a small p value of approximately 0 and our test statistic w_{test} is significantly smaller than 1 (Table 3, w_{test}=0.878). We conclude that our data is not normally distributed.

The response variable Actual Points is non-normally distributed and therefore to perform variable selection, a set of non-linear models are required. Gradient boosted Trees (GBM) are one such set of non-linear models where decision trees are sequentially constructed to correct errors made in previous trees (Natekin, 2013). Such a model is suited for non-liner relationships between features and the response (Bentejac, 2020). The algorithm can further handle data distributions, where the data is skewed and is not normally-distributed.

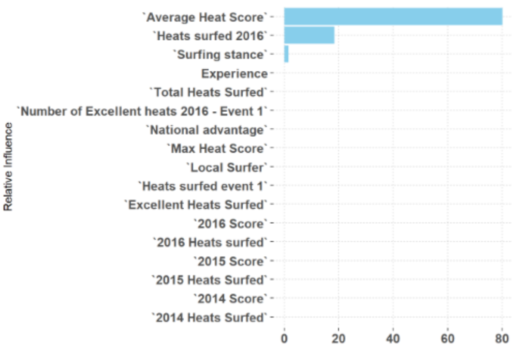
GBM's were fit to every event and variable importance plots were extracted to determine which variables are important in determining the amount of fantasy points a surfer obtains. These plots aided the decision maker in determining which variables and data are to be used in the objective function and the constraints of every event in the MILP model formulations. The most important variable was selected and used in the objective function.

3.7 Variable Importance Plots

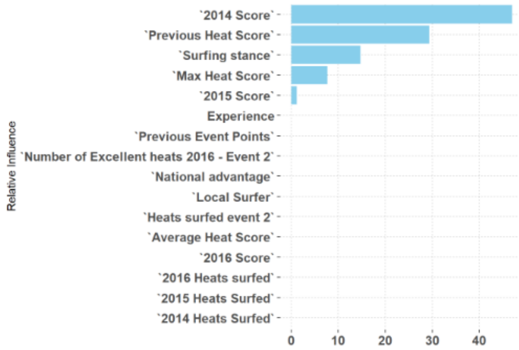
Eleven variable importance plots were obtained for every event in the 2017 WCT season. All variables were incorporated in the gbm model with one exception. Previous event fantasy points were not possible for event 1 since it's the first event of the season. Event 1 thus lacked the variable Previous Event Fantasy Points, which serves as a measure of form coming into the event. Subsequent events had data on surfer performance in prior events of the same season which can be used as a measure of form coming into the event.

The most important variable in event 1, 3, 4, 5, 7 and 11 is average heat score. This indicates that surfers obtain a large number of fantasy points in these events if they on average have a large heat score in their rounds surfed on the championship tour prior. Event 2, 3 and 6 had previous heat score being an important variable. This indicates that form and heat scores in prior events in the 2017 season were important in determining the fantasy points for these events. The most important variables in events 6, 8 and 9 were historical heat scores in the 2014-2016 seasons. this indicates that historically high heat scores in these events are an indication that a surfer will perform well in these events in the 2017 season.

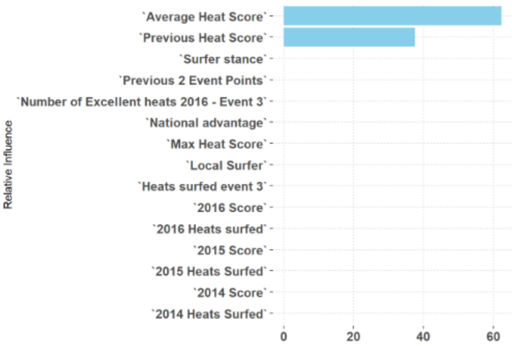
The variables Local Surfer and Nationality were not important for any of the 11 events in the 2017 season. These features were not considered for defining strategies for selection. Surfer Stance was the third most important feature in determining a surfer's fantasy score in event 1 and 2 but was not important for events 3 to 11 in the 2017 WCT season. Surfer stance was therefore not considered in determining team-selection strategies.



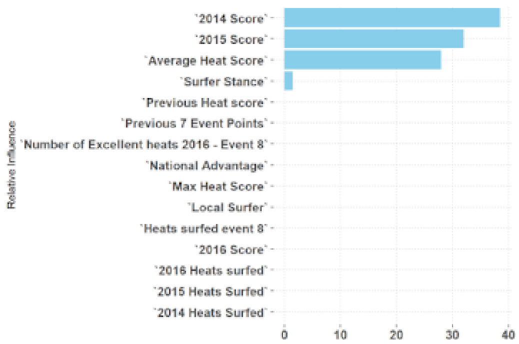
(a) Event 1



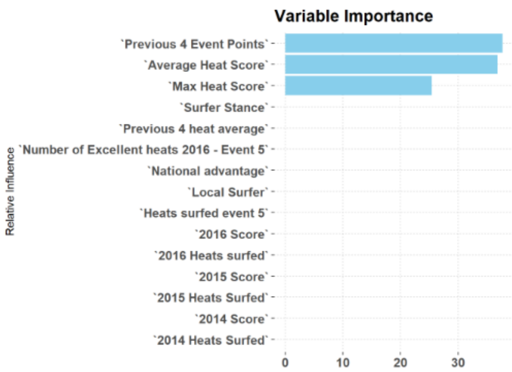
(b) Event 2



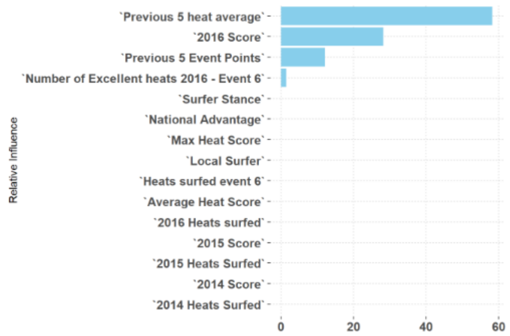
(c) Event 3



(d) Event 4

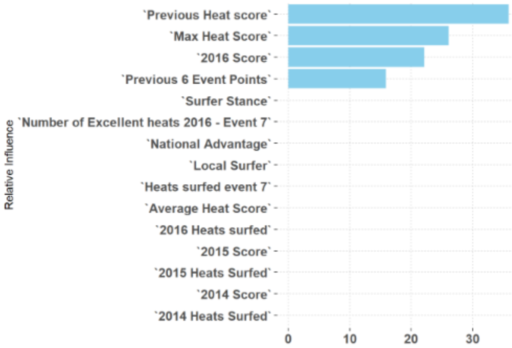


(e) Event 5

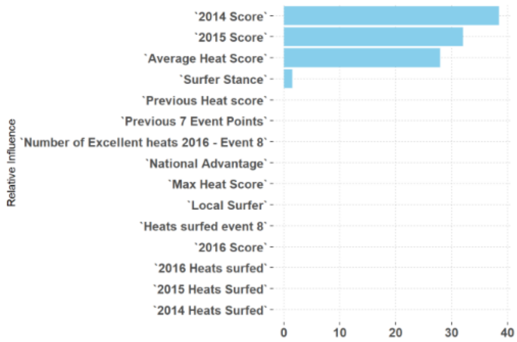


(f) Event 6

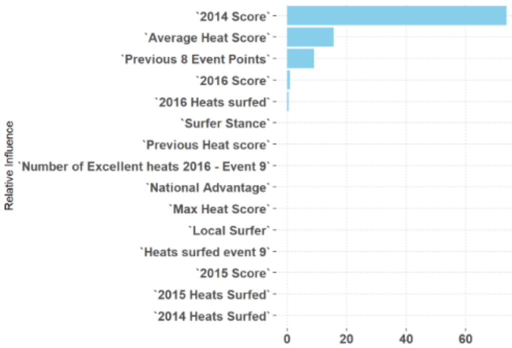
Figure 3: Variable Importance plots for events 1 to 6



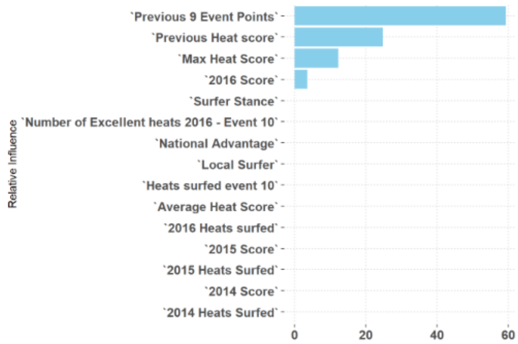
(a) Event 7



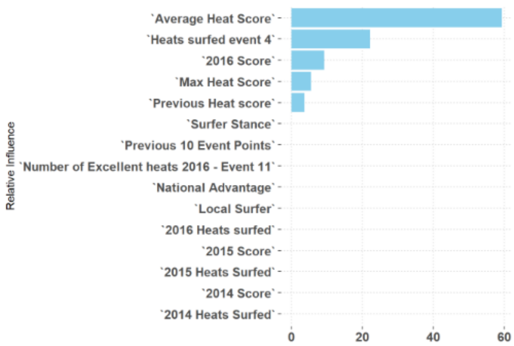
(b) Event 8



(c) Event 9



(d) Event 10



(e) Event 11

Figure 4: Variable Importance plots for events 7 to 11

3.8 Lasso Modelling

Another method used for variable selection, in conjunction with the variable importance plots, was using Linear Modelling with Lasso (L_1) regularisation.

3.8.1 Generalised Linear Modelling

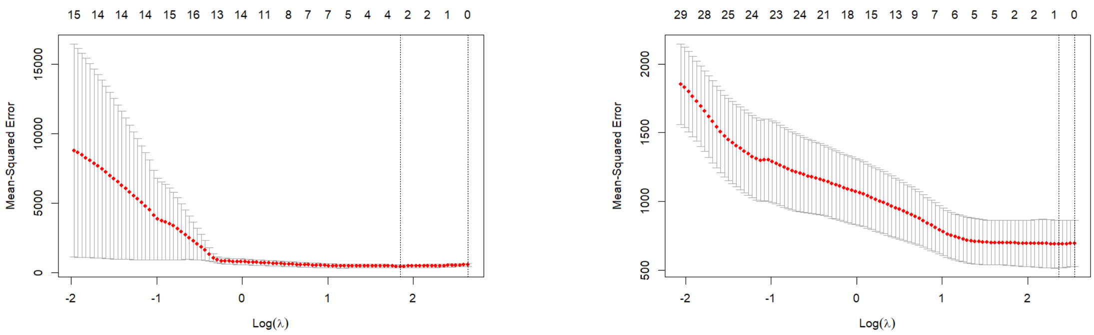
As our dataset was concluded to not be normally distributed, we used Generalized Linear Models to model our variable, Actual Points. To perform variable selection, the R package glmnet is used per event and per break type. Lasso Regression is particularly helpful in datasets where there are a large amount of variables compared to the observations which will be the case in our datasets (Fonti, 2017).

The data was imported and we conducted Lasso modelling with Actual Points as our response variable and all other variables as our explanatory variables. The R function cv.glmnet was used to choose the most appropriate value of λ for our model, where λ determines how much we weight our penalty term for our coefficients (Fonti, 2017). We will select the model with the λ value that gives the lowest cross-validated error across all values, this corresponds to the model with the lowest cross-validated error for all events. This model indicates which variables are left in the model and how many variables in total are present in the model.

After obtaining the variables and coefficients for the model, we will select the 4 or 5 largest coefficients (depending on size) to consider for our MILP problem along with the results from the Gradient Boosted Models.

3.8.2 Cross Validation Curves

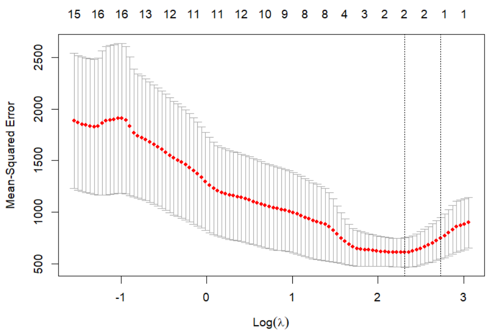
The cross validation curves for all 11 events and all 3 break types were plotted to assess the number of variables obtained in the Lasso (L_1) Regularisation models.



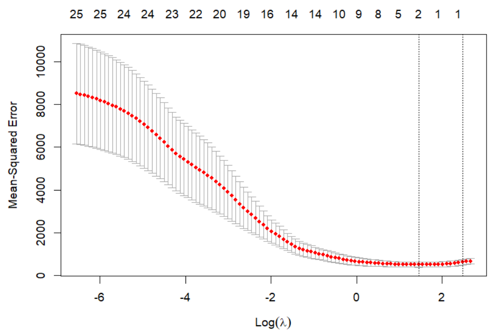
(a) Event 1

(b) Event 2

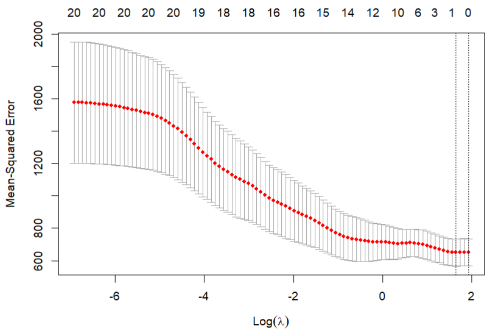
Figure 5: Cross Validated Curves for Events 1 and 2



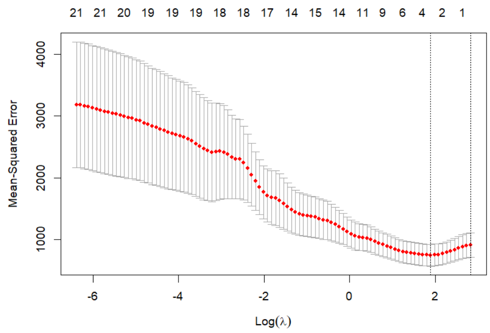
(a) Event 3



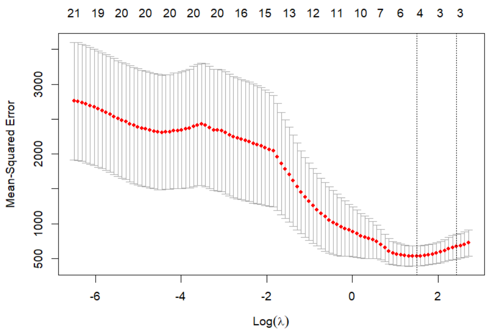
(b) Event 4



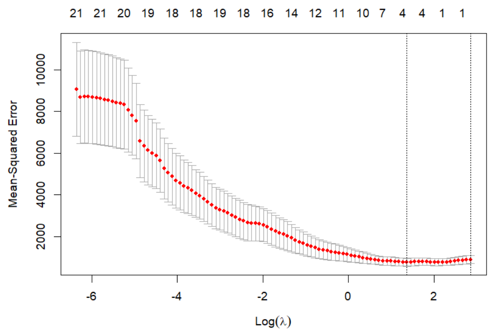
(c) Event 5



(d) Event 6

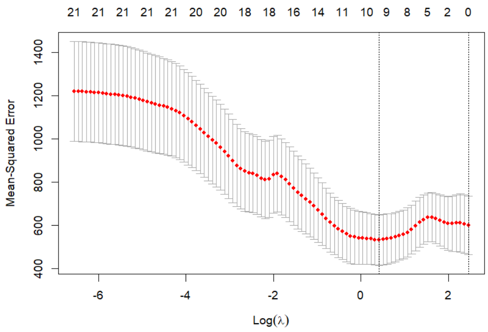


(e) Event 7

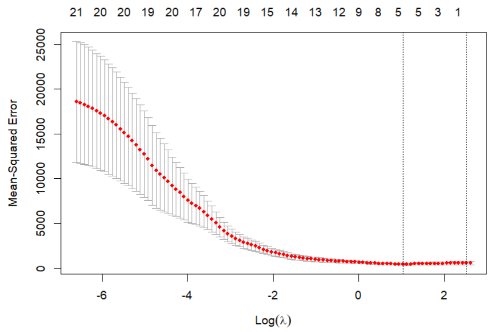


(f) Event 8

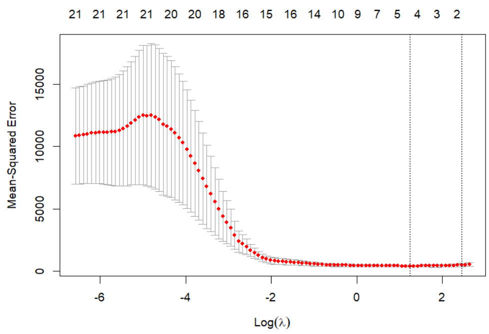
Figure 6: Cross Validated Curves for events 3 to 8



(a) Event 9

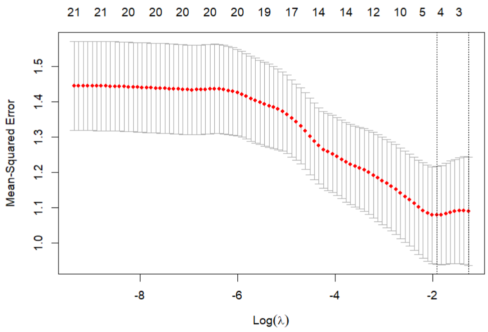


(b) Event 10

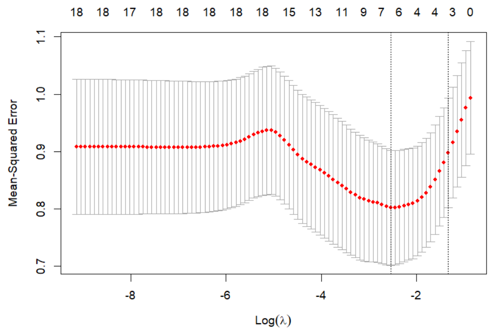


(c) Event 11

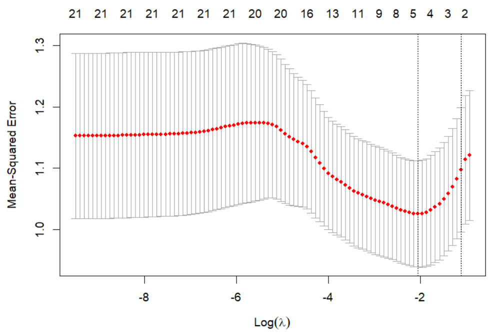
Figure 7: Cross Validated Curves for events 9 to 11



(a) Beach Break Cross Validated Curve



(b) Point Break Cross Validated Curve



(c) Reef Break Cross Validated Curve

Figure 8: Cross Validated Curves for the three wave-break types

For 10 of 11 events and 2 of 3 break types, the cross validation curves select less than 5 variables. It is only in a few cases where more variables were recommended in the final model such as the Event 9 Cross validation curve and Point Break Cross Validation curve that selected more than 5 variables. From the selected variables, the 3-5 highest variables depending on their coefficients are considered and used in our model formulations. These coefficients are found in Appendix B.

4 Methodology

4.1 Base Model Formulation (ML Approach)

Eleven mixed integer linear programming models were constructed for every event in the WCT season. The model formulation differed between events since the most important variable in determining a surfer's fantasy points was found to be different between each event.

4.1.1 Fundamental Decision Variables

The decision variables x are binary variables consisting of a 1 if the surfer is selected and 0 if not. A total of eight surfers are selected. A second binary variable y is a binary decision variable denoting a 1 if that surfer is selected as the power surfer. The power surfer gets double points and must be selected as 1 of the 8 surfer's selected in x selected.

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{36} \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{36} \end{pmatrix}$$

4.1.2 Objective Function

The objective is to maximize the number of fantasy points obtained by a selected team line-up. The model will be a function of the decision variables with some performance measure for each surfer. The team selected will be comprised of surfers that the model predicts will perform well, which in turn results in the team's fantasy points which the decision maker aims to maximise for each event.

The objective function for each MILP model depends on the most important variable for each event. In event 1 (Figure 3A and Appendix B Table 16), event 5 (Figure 3E) and event 11 ((Figure 4E and Appendix B Table 26), the most important variable was deemed Maximum Heat Score, hence this was maximized in the objective function. If the strategy is to select a surfer based on form, then previous heat score is maximized. If the strategy is to select a surfer based on surfer results in the previous season, then 2016 heat score is maximized. The strategies have the same constraints and fundamental decision variables, but differ in their objective functions.

4.1.3 Fundamental Constraints

There are several factors constraining the selection of a decision maker's line-up. Every event must adhere to the tier constraints. 2 surfers from tier 1, 4 surfers from tier 2 and 2 surfers from tier 3. Eight surfers in total must be in a player's line-up and one power surfer must be selected from one of the 8 surfers selected.

4.1.4 Event 1

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Max heat Score}^T \cdot \mathbf{X} + \textbf{Max heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Heats Surfed Previously}^T \cdot \mathbf{X} &\geq 100 \\
\textbf{Excellent Heats}^T \cdot \mathbf{X} &\geq 20 \\
\textbf{Years Surfing}^T \cdot \mathbf{X} &\geq 20 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.5 Event 2

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Previous Event Ranking}^T \cdot \mathbf{X} + \textbf{Previous Event Ranking}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2015 Average Heat Score}^T \cdot \mathbf{X} &\geq 80 \\
\textbf{2015 Average Heat Score}^T \cdot \mathbf{X} &\geq 100 \\
\textbf{Excellent Heats}^T \cdot \mathbf{X} &\geq 15 \\
\textbf{Heats Surfed}^T \cdot \mathbf{X} &\geq 30 \\
\textbf{Max Heat Score}^T \cdot \mathbf{X} &\geq 90 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.6 Event 3

$$\text{Minimise} \quad \mathbf{Z} = \mathbf{2016\,Placement}^T \cdot \mathbf{X} + \mathbf{Previous\,Heat\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Previous\,Heat\,Score}^T \cdot \mathbf{X} &\geq 310 \\
\mathbf{Max\,Heat\,Score}^T \cdot \mathbf{X} &\geq 130 \\
\mathbf{Previous\,Average}^T \cdot \mathbf{X} &\geq 150 \\
\mathbf{Excellent\,Heats}^T \cdot \mathbf{X} &\geq 35 \\
\mathbf{Regular-footed\,Surfers}^T \cdot \mathbf{X} &\geq 2 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.7 Event 4

$$\text{Maximise} \quad \mathbf{Z} = \mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Excellent\,Heats}^T \cdot \mathbf{X} &\geq 5 \\
\mathbf{Regular-footed\,Surfers}^T \cdot \mathbf{X} &\geq 2 \\
\mathbf{Previous\,Event\,Fantasy\,Points}^T \cdot \mathbf{X} &\geq 350 \\
\mathbf{Max\,Heat\,Score}^T \cdot \mathbf{X} &\geq 130 \\
\mathbf{Past\,three\,Event\,Score}^T \cdot \mathbf{X} &\geq 400 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.8 Event 5

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Max Heat Score}^T \cdot \mathbf{X} + \textbf{Max Heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Previous Event Fantasy Points}^T \cdot \mathbf{X} + \textbf{Previous Event Fantasy Points}^T \cdot \mathbf{Y} &\geq 200 \\
\textbf{Championship Score}^T \cdot \mathbf{X} + \textbf{Championship Score}^T \cdot \mathbf{Y} &\geq 100000 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.9 Event 6

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{2016 Score}^T \cdot \mathbf{X} + \textbf{2016 Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Excellent Heats}^T \cdot \mathbf{X} + \textbf{Excellent Heats}^T \cdot \mathbf{Y} &\geq 30 \\
\textbf{Championship Score}^T \cdot \mathbf{X} &\geq 130000 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.10 Event 7

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Excellent\,Heats}^T \cdot \mathbf{X} + \mathbf{Excellent\,Heats}^T \cdot \mathbf{Y} &\geq 55 \\
\mathbf{Previous\,Event\,Fantasy\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Fantasy\,Points}^T \cdot \mathbf{Y} &\geq 440 \\
\mathbf{Average\,Fantasy\,Points}^T \cdot \mathbf{X} + \mathbf{Average\,Fantasy\,Points}^T \cdot \mathbf{Y} &\geq 450 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.11 Event 8

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Average\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Average\,Heat\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y} &\geq 500 \\
\mathbf{Previous\,Event\,Fantasy\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Fantasy\,Points}^T \cdot \mathbf{Y} &\geq 350 \\
\mathbf{Nationality\,USA}^T \cdot \mathbf{X} + \mathbf{Nationality\,USA}^T \cdot \mathbf{Y} &\leq 0 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.12 Event 9

$$\textbf{Maximize} \quad \mathbf{Z = Average\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Average\,Heat\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Goofy-footed\,Surfers}^T \cdot \mathbf{X} &\geq 4 \\
\mathbf{Championship\,Score}^T \cdot \mathbf{X} + \mathbf{Championship\,Score}^T \cdot \mathbf{Y} &\geq 200000 \\
\mathbf{Nationality\,FRA}^T + \mathbf{Nationality\,FRA}^T \cdot \mathbf{Y} \cdot \mathbf{X} &\geq 2 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary}& \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.13 Event 10

$$\textbf{Minimise} \quad \mathbf{Z = 2017\,Event\,Placement}^T \cdot \mathbf{X} + \mathbf{2017\,Event\,Placement}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Max\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Max\,Heat\,Score}^T \cdot \mathbf{Y} &\geq 150 \\
\mathbf{Previous\,Average\,Fantasy\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Average\,Fantasy\,Points}^T \cdot \mathbf{Y} &\geq 450 \\
\mathbf{Previous\,Event\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Heat\,Score}^T \cdot \mathbf{Y} &\geq 550 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary}& \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.1.14 Event 11

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Max Heat Score}^T \cdot \mathbf{X} + \textbf{Max Heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Previous Average Fantasy Points}^T \cdot \mathbf{X} + \textbf{Previous Average Fantasy Points}^T \cdot \mathbf{Y} &\geq 100 \\
\textbf{Previous Event Heat Score}^T \cdot \mathbf{X} + \textbf{Previous Event Heat Score}^T \cdot \mathbf{Y} &\geq 450 \\
\textbf{Nationality USA}^T + \textbf{Nationality USA}^T \cdot \mathbf{Y} \cdot \mathbf{X} &\geq 3 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} & \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.2 Team Selection Strategy Model Formulation

The base model was used to construct several model strategies that were assessed for both the 2017 and 2018 season. These model strategies were compared against each other, the base model and the optimal model. Each model strategy had the same base model constraints and decision variables, just differed in its objective function. The base model in essence determined the important features for each event by looking at the fantasy results obtained post-event. This approach is effective in building an efficient model and observing how different features are important for different events. The process also highlighted the variability in surfer performance between consecutive events. This approach however is expected to perform well for the 2017 season since it is modeling for results that it has already seen. The other strategies do not use such an approach and rather use historical or seasonal data prior to the event. The base model will be more fairly assessed in conjunction with the strategies in the 2018 season.

4.2.1 Optimal Team

In each individual event, a surfer's fantasy points is the sum of the heat scores for every round surfed in the event. For every event, fantasy points are found for all the surfers that participated. For every event, a post-event optimal and random team were selected based on the fantasy points in the 2017 and 2018 WCT season. In the optimal team, the optimal power surfer for each event was the surfer that had the largest fantasy points. The team was selected adhering to tier constraints; surfers with the top 2 largest fantasy points scoring surfers from tier 1 were selected. The top 4 from tier 2 were selected and the top 2 from tier 3 were selected. The fantasy points of all surfers selected are summed and equate to an optimal model fantasy points total.

4.2.2 Random Team

The random model assigned random numbers to all surfers in the respective three tiers. Adhering to the tier constraints, random surfers in the respective tiers were selected. 2 surfers from tier 1, 4 surfers from tier 2 and 2 surfers from tier 3. The power surfer was selected randomly by choosing a random surfer of the 8 selected to be the power surfer. The fantasy points of all surfers selected are summed and equate to a random model fantasy points total. The random model process is repeated ten times and the ten fantasy points totals are averaged. This process was done for all 11 events in the 2017 and 2018 WCT season.

4.2.3 2016 Score

A surfer's wave scores for the same event in 2016 are termed '2016 score'. 2016 score was maximized in the objective function for every event. Surfers that performed well in the 2016 season for that particular event are favored by the model. This strategy helps observe how consistent surfers are in the same event in consecutive seasons.

4.2.4 Previous Event Fantasy Points

The features previous heat score and previous event fantasy points are maximized. In this strategy, surfers that performed well in previous events of the same season were favored. This strategy is a measure of surfer consistency between successive WCT events.

4.2.5 Current Standings

The variable Championship Score is maximised. Championship Score is a measure of ranking on the WCT leader board. This strategy selects surfers that are performing well that season and are highly ranked on the WCT leader board.

4.2.6 Wave Break Type

Wave break types are grouped into Reef Break, Point Break and Beach Break. For the 2017 Events, they were allocated the following type of breaks:

Break Type	**Event**
Beach Break	Oi Rio Pro – Saquarema, Brazil (Event 4) Quiksilver Pro France – Hossegor, France (Event 9) Meo Rip Curl Pro Portugal – Peniche, Portugal (Event 10)
Point Break	Quiksilver Pro Gold Coast – Snapper Rocks, Australia (Event 1) Rip Curl Pro Bells Beach – Bells Beach, Australia (Event 3) Corona Open J-Bay – Jeffreys Bay, South Africa (Event 6)
Reef Break	Margaret River Pro – Margaret River, Australia (Event 2) Fiji Pro – Cloudbreak, Fiji (Event 5) Tahiti Pro Teahupo'o – Teahupo'o, Tahiti (Event 7) Hurley Pro at Trestles - Lower Trestles, United States (Event 8) Billabong Pipe Masters – Banzai Pipeline, Oahu, Hawaii (Event 11)

Table 4: Allocation of events into the three wave break types in the 2017 WCT season

For this strategy, we take a Mixed-Integer Linear Programming approach such as our baseline model approach. The main difference would be the amount of data used to train the model. For each break, we would model the fantasy points based off all available variables. However, the data for all events of that break type would be pooled together and we would then select the 3 top variables and use these in our MILP problem. The variables to be selected were found using linear modelling L_1 regularisation.

Break Type	Variables used
Beach Break	2015 Score Average Previous Heat Score 2016 Score
Point Break	2016 Heat Score Maximum Heat Score Championship Points/Score
Reef Break	Maximum Heat Score Number of Excellent Heats Previous Event Score

Table 5: Top three important variables for each Wave Break Type

The important features differ for different wave breaks. Surfers that performed well in 2015 and/or 2016, tend to be consistent performers in the 2017 season in beach breaks and points breaks. Surfers in reef breaks tend to do well if historically they have a large number of excellent heats and high heat score totals (Table 5).

4.3 Model Formulation based on Break Type

For each of the events, the constraints were kept constant except for championship score. As this value increases throughout the year for majority of the surfers, these values were scaled for events 1,3 and 6 in order to constrain these values more.

4.3.1 Event 1

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Max heat Score}^T \cdot \mathbf{X} + \textbf{Max heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2016 Score}^T \cdot \mathbf{X} + \textbf{2016 Score}^T \cdot \mathbf{Y} &\geq 500 \\
\textbf{Championship Score}^T \cdot \mathbf{X} + \textbf{Championship Score}^T \cdot \mathbf{Y} &\geq 29000 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} & \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.3.2 Event 2, 5, 7, 8 and 11

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Excellent\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Excellent\,Heat\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Average\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Average\,Heat\,Score}^T \cdot \mathbf{Y} &\geq 500 \\
\mathbf{Max\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Max\,Heat\,Score}^T \cdot \mathbf{Y} &\geq 150 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.3.3 Event 3

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Max\,heat\,Score}^T \cdot \mathbf{X} + \mathbf{Max\,heat\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y} &\geq 500 \\
\mathbf{Championship\,Score}^T \cdot \mathbf{X} + \mathbf{Championship\,Score}^T \cdot \mathbf{Y} &\geq 60000 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.3.4 Event 4, 9 and 10

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{2015\,Score}^T \cdot \mathbf{X} + \mathbf{2015\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y} \geq 500$$
$$\mathbf{Previous\,Event\,Average\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Average\,Heat\,Score}^T \cdot \mathbf{Y} \geq 500$$
$$\mathbf{Tier\,1}^T \cdot \mathbf{X} = 2$$
$$\mathbf{Tier\,2}^T \cdot \mathbf{X} = 4$$
$$\mathbf{Tier\,3}^T \cdot \mathbf{X} = 2$$
$$\sum_{i=1}^{36} \mathbf{y}_i = 1$$
$$\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary}$$
$$i = 1, 2, \ldots, 36$$

4.3.5 Event 6

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Max\,heat\,Score}^T \cdot \mathbf{X} + \mathbf{Max\,heat\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y} \geq 500$$
$$\mathbf{Championship\,Score}^T \cdot \mathbf{X} + \mathbf{Championship\,Score}^T \cdot \mathbf{Y} \geq 115000$$
$$\mathbf{Tier\,1}^T \cdot \mathbf{X} = 2$$
$$\mathbf{Tier\,2}^T \cdot \mathbf{X} = 4$$
$$\mathbf{Tier\,3}^T \cdot \mathbf{X} = 2$$
$$\sum_{i=1}^{36} \mathbf{y}_i = 1$$
$$\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary}$$
$$i = 1, 2, \ldots, 36$$

4.4 Top three statistics

A strategy termed 'Reoccurring Stats' built an MILP model on the top three variables that were deemed important in determining surfer performance. The variables 2016 score, previous event fantasy points and average heat score were used for every event. In essence, the same three statistically important variables were used to model all 11 events. These top 3 variables were obtained by taking the top 3 variables that were used across all 11 events in our base model approach.

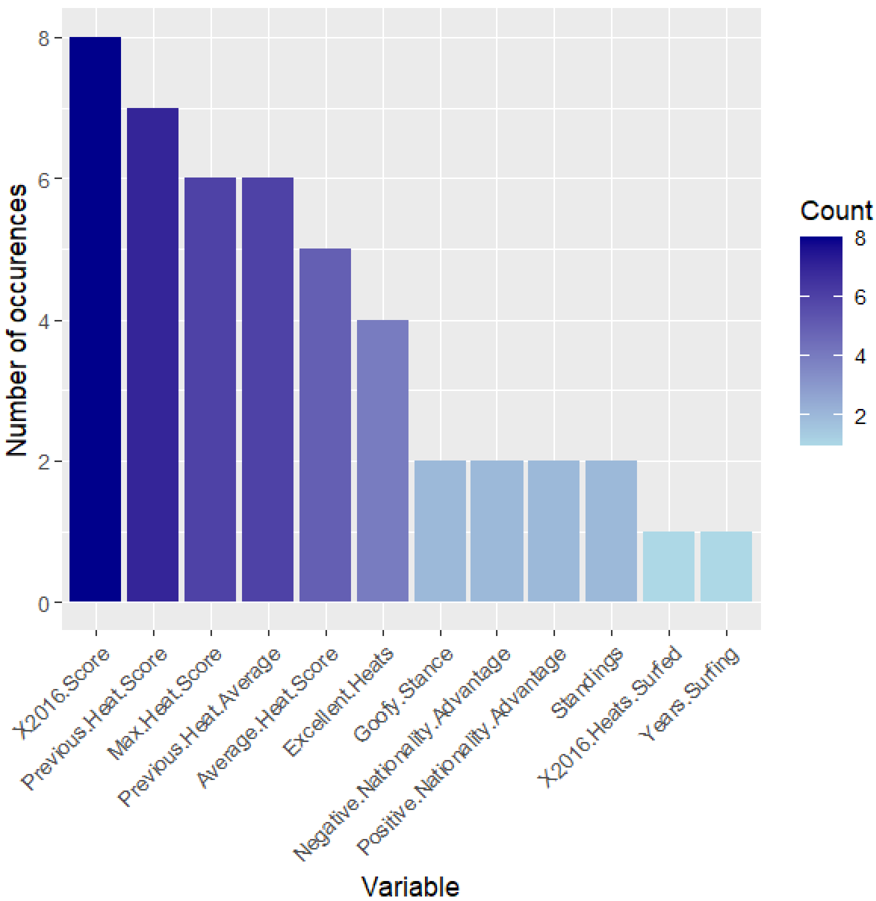


Figure 9: Variable count across 11 events

Variables regarding the 2016 season were grouped into one count as it would be misleading to have multiple variables for 2016 where it would seem unimportant compared to other variables. The top 4 variables were the previous event score, maximum heat score, previous event averages and the variables regarding the 2016 season (Figure 9). The variables chosen were the 2016 total score, previous heat score and the previous event average. Maximum Heat Score was omitted in favour of the Average Heat Score as Average Heat Score could have been used in the first two events and appeared in 6 out of the next 9 events.

Model formulation for the top 3 variables

4.4.1 Event 1

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{2016\,Points}^T \cdot \mathbf{X} + \mathbf{2016\,Points}^T \cdot \mathbf{Y}$$

$$\textbf{Subject to:}$$

$$\begin{aligned}
\mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y} &\geq 550 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.4.2 Event 2

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Previous\,Event\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Points}^T \cdot \mathbf{Y}$$

$$\textbf{Subject to:}$$

$$\begin{aligned}
\mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y} &\geq 550 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.4.3 Event 3-11

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Previous\,Event\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Points}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y} &\geq 500\\
\mathbf{Previous\,Event\,Average}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Average}^T \cdot \mathbf{Y} &\geq 450\\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2\\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4\\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2\\
\sum_{i=1}^{36} \mathbf{y}_i &= 1\\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary}\\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.5 Power Surfer Strategy Model Formulation

Several Power Surfer Strategies were implemented to assess the best approach or strategy a player can implement when selecting a power surfer.

4.5.1 Tier 1 Only

The power surfer was firstly restricted to being only in tier 1. This implies that only highly ranked surfers coming into an event are eligible to be the power surfer. The objective function and constraints were unchanged, with the addition of one further constraint:

$$\sum_{i=1}^{8} \mathbf{y}_i = 1$$

The power surfer essentially is constrained to be one of the top 8 ranked surfers. This is a measure of seasonal form since the top 8 ranked surfers coming into an event have good form and have performed well in prior events in the same season.

4.5.2 Historical Winner

The power surfer is constrained to being a surfer that won the event the previous season. The constraints are modified and the 2016 event placement is used as a constraint. If the 2016 event placement has a score of 1, then it implies that the surfer won the event the previous season.

The following additional constraint was added:

$$\textbf{2016 Event Placement}^T \cdot \mathbf{X} = 1$$

4.5.3 Previous Event Optimal Power Surfer

The power surfer is constrained to being a surfer that obtained the most fantasy points in the previous event that same season. This surfer was essentially the optimal power surfer in the previous event. This strategy is a measure of surfer consistency between events. The constraints are modified and the previous event fantasy points are modified into a binary variable z that can take on either 0 or 1.

$$\mathbf{Z} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \vdots \\ z_{36} \end{pmatrix}$$

$$z_i = \begin{cases} 1 & \text{if the surfer has the largest previous event fantasy points} \\ 0 & \text{otherwise} \end{cases}$$

The following additional constraint was added:

$$\mathbf{Y}^T \cdot \mathbf{Z} = 1$$

4.6 Alterations To Model Formulation

Several alterations were made to model formulations in the 2018 season since the points system changed in 2018 compared to the previous years with 1 round fewer. Three new events were added in the 2018 season compared to 2017. We had to adjust some of the constraints to make the model feasible. The models below show how the models altered for our 3 Machine Learning MILP problems.

The main variables that were altered were those relating to form as surfers achieved less points than the 2018 season due to the fact there was one fewer round in all events in the 2018 season. Other variables that had slight alterations was excellent heats achieved in some events over the past three years. 2014 Appeared to have a good amount of excellent heats in that season, so once we removed the 2014 season from consideration for modelling the 2018 season, there were fewer excellent heats for that event. There were no major rule changes from 2014 to 2015 so this is largely due to seasonal variability in surfer performance between events in consecutive seasons. All other variables stayed the same between the 2017 and 2018 WCT seasons so event 3, 4, 5, 7, 8 and 11 had minor alterations in constraints of Excellent Heats, Previous Event Fantasy Points and Average Fantasy Points.

4.7 Base Model Alterations for 2018 Season

4.7.1 Event 3 Formulation

$$\textbf{Minimise} \quad \mathbf{Z = 2016\,Placement}^T \cdot \mathbf{X} + \mathbf{Previous\,Heat\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Previous\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Previous\,Heat\,Score}^T \cdot \mathbf{Y} &\geq 310 \\
\mathbf{Max\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Max\,Heat\,Score}^T \cdot \mathbf{Y} &\geq 130 \\
\mathbf{Previous\,Average}^T \cdot \mathbf{X} + \mathbf{Previous\,Average}^T \cdot \mathbf{Y} &\geq 150 \\
\mathbf{Excellent\,Heats}^T \cdot \mathbf{X} + \mathbf{Excellent\,Heats}^T \cdot \mathbf{Y} &\geq 35 \\
\mathbf{Regular-footed\,Surfers}^T \cdot \mathbf{X} &\geq 2 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} & \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.7.2 Event 4 Formulation

$$\textbf{Maximise} \quad \mathbf{Z = Previous\,Event\,Fantasy\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Excellent Heats}^T \cdot \mathbf{X} + \textbf{Excellent Heats}^T \cdot \mathbf{Y} &\geq 27 \\
\textbf{Max Heat}^T \cdot \mathbf{X} + \textbf{Max Heat}^T \cdot \mathbf{Y} &\geq 320 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.7.3 Event 5 Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Previous Event Ranking}^T \cdot \mathbf{X} + \textbf{Previous Event Ranking}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2016 Average Heat Score}^T \cdot \mathbf{X} + \textbf{2016 Average Heat Score}^T \cdot \mathbf{Y} &\geq 80 \\
\textbf{2017 Average Heat Score}^T \cdot \mathbf{X} + \textbf{2017 Average Heat Score}^T \cdot \mathbf{Y} &\geq 100 \\
\textbf{Excellent Heats}^T \cdot \mathbf{X} + \textbf{Excellent Heats}^T \cdot \mathbf{Y} &\geq 15 \\
\textbf{Heats Surfed}^T \cdot \mathbf{X} + \textbf{Heats Surfed}^T \cdot \mathbf{Y} &\geq 30 \\
\textbf{Best Heat Score}^T \cdot \mathbf{X} + \textbf{Best Heat Score}^T \cdot \mathbf{Y} &\geq 90 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.7.4 Event 7 Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Excellent\,Heats}^T \cdot \mathbf{X} + \mathbf{Excellent\,Heats}^T \cdot \mathbf{Y} &\geq 20 \\
\mathbf{Previous\,Event\,Fantasy\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Fantasy\,Points}^T \cdot \mathbf{Y} &\geq 400 \\
\mathbf{Average\,Fantasy\,Points}^T \cdot \mathbf{X} + \mathbf{Average\,Fantasy\,Points}^T \cdot \mathbf{Y} &\geq 380 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.7.5 Event 8 Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Max\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Max\,Heat\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{Championship\,Standings}^T \cdot \mathbf{X} + \mathbf{Championship\,Standings}^T \cdot \mathbf{Y} &\geq 200000 \\
\mathbf{Excellent\,Heats}^T \cdot \mathbf{X} + \mathbf{Excellent\,Heats}^T \cdot \mathbf{Y} &\geq 30 \\
\mathbf{Previous\,Fantasy\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Fantasy\,Points}^T \cdot \mathbf{Y} &\geq 400 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.7.6 Event 11 Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Max Heat Score}^T \cdot \mathbf{X} + \textbf{Max Heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Previous Average Fantasy Points}^T \cdot \mathbf{X} + \textbf{Previous Average Fantasy Points}^T \cdot \mathbf{Y} &\geq 90 \\
\textbf{Previous Event Heat Score}^T \cdot \mathbf{X} + \textbf{Previous Event Heat Score}^T \cdot \mathbf{Y} &\geq 450 \\
\textbf{Nationality HAW}^T \cdot \mathbf{X} &\geq 3 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i & \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.8 Reoccurring Stats Model Alterations

For the Reoccurring Stats Strategy minor alterations in the constraints of Excellent Heats, Previous Event Fantasy Points and Average Fantasy Points were made.

4.8.1 Event 3, 6, 9 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Previous Event Points}^T \cdot \mathbf{X} + \textbf{Previous Event Points}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2017 Score}^T \cdot \mathbf{X} + \textbf{2017 Score}^T \cdot \mathbf{Y} &\geq 500 \\
\textbf{Previous Event Average}^T \cdot \mathbf{X} + \textbf{Previous Event Average}^T \cdot \mathbf{Y} &\geq 380 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i & \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.8.2 Event 4 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Previous\,Event\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Points}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{2017\,Score}^T \cdot \mathbf{X} + \mathbf{2017\,Score}^T \cdot \mathbf{Y} &\geq 480 \\
\mathbf{Previous\,Event\,Average}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Average}^T \cdot \mathbf{Y} &\geq 350 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} & \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.8.3 Event 7 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{Previous\,Event\,Points}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Points}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\mathbf{2017\,Score}^T \cdot \mathbf{X} + \mathbf{2017\,Score}^T \cdot \mathbf{Y} &\geq 550 \\
\mathbf{Previous\,Event\,Average}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Average}^T \cdot \mathbf{Y} &\geq 400 \\
\mathbf{Tier\,1}^T \cdot \mathbf{X} &= 2 \\
\mathbf{Tier\,2}^T \cdot \mathbf{X} &= 4 \\
\mathbf{Tier\,3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} & \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.8.4 Event 8 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Previous Event Points}^T \cdot \mathbf{X} + \textbf{Previous Event Points}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2017 Score}^T \cdot \mathbf{X} + \textbf{2017 Score}^T \cdot \mathbf{Y} &\geq 450 \\
\textbf{Previous Event Average}^T \cdot \mathbf{Y} &\geq 350 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.8.5 Event 10 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Previous Event Points}^T \cdot \mathbf{X} + \textbf{Previous Event Points}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2017 Score}^T \cdot \mathbf{X} + \textbf{2017 Score}^T \cdot \mathbf{Y} &\geq 480 \\
\textbf{Previous Event Average}^T \cdot \mathbf{X} + \textbf{Previous Event Average}^T \cdot \mathbf{Y} &\geq 370 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.8.6 Event 11 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Previous Event Points}^T \cdot \mathbf{X} + \textbf{Previous Event Points}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2017 Score}^T \cdot \mathbf{X} + \textbf{2017 Score}^T \cdot \mathbf{Y} &\geq 450 \\
\textbf{Previous Event Average}^T \cdot \mathbf{X} + \textbf{Previous Event Average}^T \cdot \mathbf{Y} &\geq 300 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.9 Break Type Model Alterations

For the Break Type Strategy in the 2018 WCT season, minor alterations were made to the constraints of the features Excellent Heats, Previous Event Average Fantasy Heat Score and Previous Fantasy Points.

4.9.1 Event 2 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Max heat Score}^T \cdot \mathbf{X} + \textbf{Max heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2016 Score}^T \cdot \mathbf{X} + \textbf{2016 Score}^T \cdot \mathbf{Y} &\geq 500 \\
\textbf{Championship Score}^T \cdot \mathbf{X} + \textbf{Championship Score}^T \cdot \mathbf{Y} &\geq 30000 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.9.2 Event 3 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\mathbf{2017\,Score}^T \cdot \mathbf{X} + \mathbf{2017\,Score}^T \cdot \mathbf{Y} \geq 400$$
$$\mathbf{Previous\,Event\,Average\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Average\,Heat\,Score}^T \cdot \mathbf{Y} \geq 400$$
$$\mathbf{Tier\,1}^T \cdot \mathbf{X} = 2$$
$$\mathbf{Tier\,2}^T \cdot \mathbf{X} = 4$$
$$\mathbf{Tier\,3}^T \cdot \mathbf{X} = 2$$
$$\sum_{i=1}^{36} \mathbf{y}_i = 1$$
$$\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary}$$
$$i = 1, 2, \ldots, 36$$

4.9.3 Event 4 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \mathbf{2016\,Score}^T \cdot \mathbf{X} + \mathbf{2016\,Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\mathbf{2017\,Score}^T \cdot \mathbf{X} + \mathbf{2017\,Score}^T \cdot \mathbf{Y} \geq 400$$
$$\mathbf{Previous\,Event\,Average\,Heat\,Score}^T \cdot \mathbf{X} + \mathbf{Previous\,Event\,Average\,Heat\,Score}^T \cdot \mathbf{Y} \geq 380$$
$$\mathbf{Tier\,1}^T \cdot \mathbf{X} = 2$$
$$\mathbf{Tier\,2}^T \cdot \mathbf{X} = 4$$
$$\mathbf{Tier\,3}^T \cdot \mathbf{X} = 2$$
$$\sum_{i=1}^{36} \mathbf{y}_i = 1$$
$$\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary}$$
$$i = 1, 2, \ldots, 36$$

4.9.4 Event 5 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Excellent Heat Score}^T \cdot \mathbf{X} + \textbf{Excellent Heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Average Heat Score}^T \cdot \mathbf{X} + \textbf{Average Heat Score}^T \cdot \mathbf{Y} &\geq 460 \\
\textbf{Max Heat Score}^T \cdot \mathbf{X} + \textbf{Max Heat Score}^T \cdot \mathbf{Y} &\geq 130 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} & \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.9.5 Event 7, 11 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Excellent Heat Score}^T \cdot \mathbf{X} + \textbf{Excellent Heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Average Heat Score}^T \cdot \mathbf{X} + \textbf{Average Heat Score}^T \cdot \mathbf{Y} &\geq 470 \\
\textbf{Max Heat Score}^T \cdot \mathbf{X} + \textbf{Max Heat Score}^T \cdot \mathbf{Y} &\geq 130 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i \text{ binary} & \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.9.6 Event 8 Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{Max Heat Score}^T \cdot \mathbf{X} + \textbf{Max Heat Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{Championship Standings}^T \cdot \mathbf{X} + \textbf{Championship Standings}^T \cdot \mathbf{Y} &\geq 200000 \\
\textbf{Excellent Heats}^T \cdot \mathbf{X} + \textbf{Excellent Heats}^T \cdot \mathbf{Y} &\geq 30 \\
\textbf{Previous Fantasy Points}^T \cdot \mathbf{X} + \textbf{Previous Fantasy Points}^T \cdot \mathbf{Y} &\geq 400 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

4.9.7 Event 9,10 Model Formulation

$$\textbf{Maximize} \quad \mathbf{Z} = \textbf{2016 Score}^T \cdot \mathbf{X} + \textbf{2016 Score}^T \cdot \mathbf{Y}$$

Subject to:

$$\begin{aligned}
\textbf{2017 Score}^T \cdot \mathbf{X} + \textbf{2017 Score}^T \cdot \mathbf{Y} &\geq 450 \\
\textbf{Previous Event Average Heat Score}^T \cdot \mathbf{X} + \textbf{Previous Event Average Heat Score}^T \cdot \mathbf{Y} &\geq 340 \\
\textbf{Tier 1}^T \cdot \mathbf{X} &= 2 \\
\textbf{Tier 2}^T \cdot \mathbf{X} &= 4 \\
\textbf{Tier 3}^T \cdot \mathbf{X} &= 2 \\
\sum_{i=1}^{36} \mathbf{y}_i &= 1 \\
\forall i, \mathbf{x}_i, \mathbf{y}_i &\text{ binary} \\
i &= 1, 2, \ldots, 36
\end{aligned}$$

5 Results and Discussion

The MILP models for the base model, team selection strategies and power surfer strategies were solved using the OpenSolver add-in in Microsoft Excel. The objective function, constraints and decision variables were inputted and OpenSolver produced the optimal (maximised) objective function value and the optimal combination of decision variables. This optimal configuration of decision variables determined the 8 surfers and one power surfer selected for that particular event. The true fantasy points were obtained and were cross-producted with the decision variables to obtain the fantasy points for the event. The fantasy points for every event were summed to obtain a cumulative seasonal fantasy points score for both the 2017 and 2018 seasons respectively.

5.1 2017 Championship Tour Season

The MILP models generated a wide range of results for various team selection strategies for the 2017 season. We compare how the various strategies performed with the caveat that the machine learning models were trained on the 2017 WCT season.

5.1.1 Team Selection Strategies

For the 2017 season, the results obtained were by means of various machine learning methodologies, such as linear modelling and gradient boosted trees. The purpose of these methods were to obtain the important variables for predicting performance at a specific event or type of break that the event has. After obtaining these variables, we build a MILP model which would be used to select teams. These strategies were compared to other team selection strategies to see the performance across the various events in the 2017 season. The strategies performance were assessed with regards to their average round points, total variance across the season, standard deviation from the mean, their minimum and maximum value as well as the total points that the strategy obtained (Table 6).

Table 6: Summary Statistics of Strategies for 2017 Season (Organized by Total Points)

Strategy	**Mean**	**Variance**	**Std_Dev**	**Min**	**Max**	**Total**
Optimal	688.65	5811.79	76.24	587.26	838.28	7575.17
ML approach	499.93	5425.00	73.65	385.05	611.36	5499.27
Reoccuring stats	448.73	7362.67	85.81	334.87	612.97	4936.06
Break Type	445.56	10790.11	103.88	253.43	610.06	4901.14
Current Standings	430.84	3352.36	57.90	341.79	512.34	4739.21
2016 Score	416.51	9363.03	96.76	226.71	555.03	4581.56
Previous Event	401.37	6127.49	78.28	254.13	514.32	4415.05
Random Guess	383.78	6403.10	80.02	208.50	460.91	4221.63

By means of comparison for the models are the optimal team selection as well as an average of 10 random guesses. All 3 models that incorporated machine learning methods (ML approach, Reoccurring Stats and Break Type) placed in the top 3 of all strategies for the 2017 season. The total fantasy points achieved by these three strategies outperform the other strategies. The top three strategies ML Approach, Reoccurring Stats and Break Type achieve a higher maximum fantasy score score than the other strategies, by achieving a maximum score of over 600 fantasy points, whereas the other novel strategies achieved a max score of 550.

The standard deviation across all strategies are similar, ranging from 70-80 points. The notable outliers are the strategy where you favour surfers who did well the previous year (2016 Score) and our "Break type" model which is the most variable. The Break Type approach had the second lowest minimum fantasy score. The "Previous Event" method and Last year winner method performed poorly over the season, with it ranking last and second last of the traditional strategies, with only the random guess method being worse. This shows that a surfer's performance varies between events and it is vary rare that a surfer tends to do well in consecutive events. Current Standings is the best strategy for the 2017 season out of the traditional strategies. The Current Standings strategy has the lowest variance of all strategies, including the optimal and the random guess strategies.

A comparison of all strategies against the random guess and optimal strategy was obtained.

Table 7: Strategy Comparison with Mean and Standard Deviation Against Optimal and Random Guess

Strategy	**Mean**	**Std_Dev**	**Optimal Mean Delta**	**Optimal Std_Dev Delta**	**Random Guess Mean Delta**	**Random Guess Std_Dev Delta**
ML approach	499.93	73.65	-188.72	-2.59	116.15	-6.37
Previous Event	401.37	78.28	-287.28	2.04	17.59	-1.74
2016 Score	416.51	96.76	-272.14	20.52	32.73	16.74
Reoccuring stats	448.73	85.81	-239.92	9.57	64.95	5.79
Current Standings	430.84	57.90	-257.81	-18.34	47.06	-22.12
Break Type	445.56	103.88	-243.09	27.64	61.78	23.86

The closest strategy compared to the optimal strategy is our ML approach. Break Type and the Reoccurring Stats strategy which perform better than the rest, once again showing that a more generalised model can be used to predict performance. Although one would expect the optimal team to vary between events, most strategies were more variable with only the ML approach and Current Standings being less variable than the optimal model.

Compared to the Random guess strategy, all models perform better by means of averages, with the three Machine Learning approaches outperforming this strategy by at least 60 points per event. The previous event only performs marginally better than the random guess, receiving 10 more points per round. We would expect the standard deviation and variability of random guesses to be large due to the significant element of randomness in the strategy. However, two strategies, 2016 Score and Break Type, have a larger standard deviation than the random guess. Reoccurring Stats and the Previous Event strategy have similar standard deviations to the random guess model, only differing by less than 3 points.

The performance of the different strategies are compared per event in the 2017 WCT season to better assess the fantasy team-selection strategies.

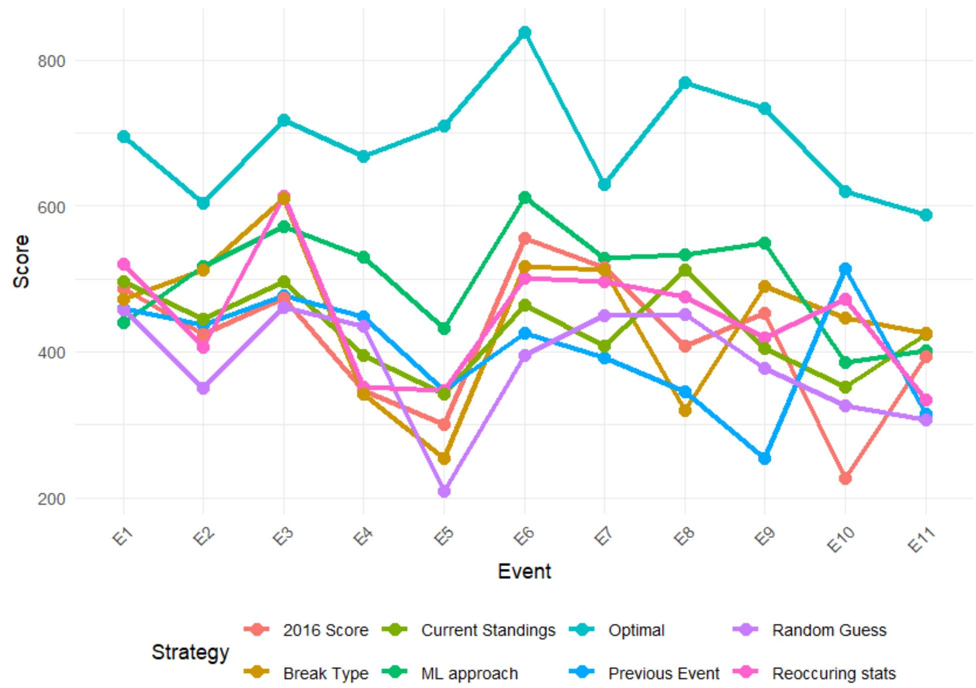
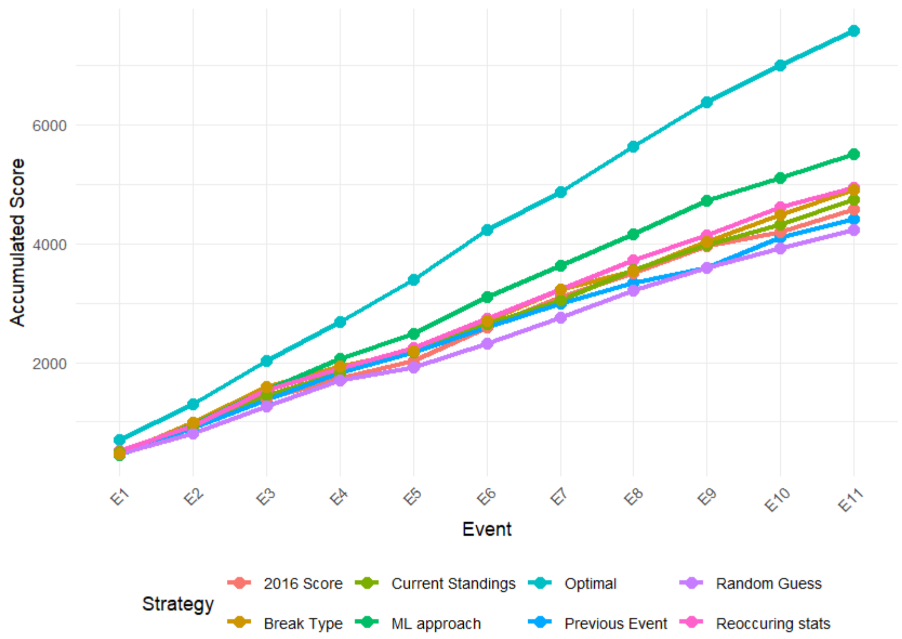


Figure 10: Fantasy Scores per Event for Each Strategy in 2017

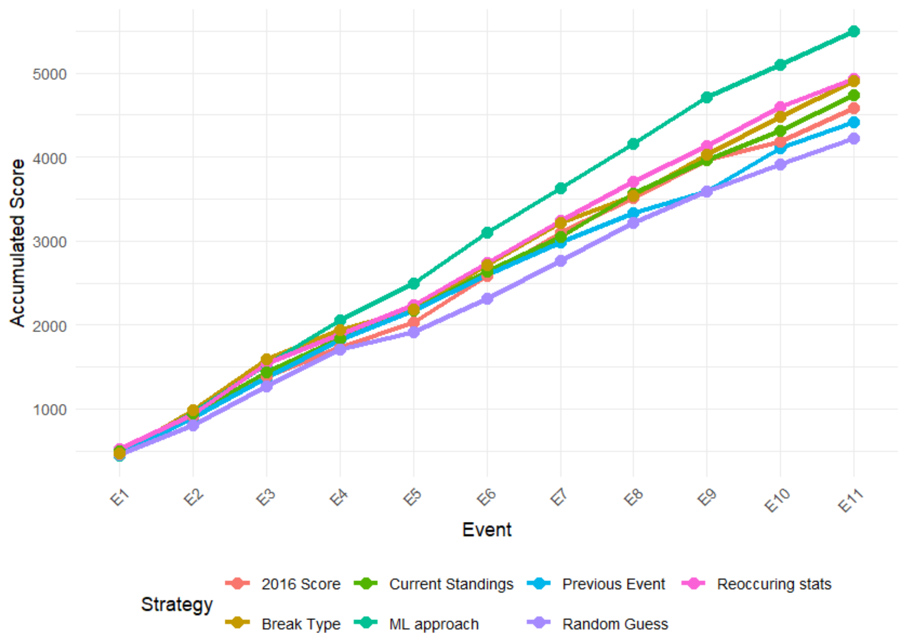
There are large fluctuations in fantasy scores across the 2017 WCT season indicating that no 2 events are the same. The optimal points achieved per round varies throughout the season, once again showing that events can be treated as independent. Surfer performance is also inconsistent, as seen in event 5 where several of the expected performers based on historical data of the same event in previous seasons were knocked out early, allowing for the unfavoured/unexpected rookie and wildcard surfers to perform well (Figure 10). As a result, all strategies performed poorly in event 5. In event 6, all surfers who were expected to do well based on historical data of event 6 in previous seasons performed well resulting in all strategies performing well for that particular event. Event 6 was also a very high scoring event with all surfers scoring very high points throughout the event.

The ML model was better than all strategies in event 4-9 (Figure 10). In the first three events, the strategies break type and Reoccurring Stats performed better than ML approach. This is possibly due to the limited amount of data that the model had available in the 2017 season. It is also important to note that the ML approach performed poorly in the first event, ranking last out of all available strategies. Post event 3, there were several events worth of data on all surfers in the 2017 season and so there was a more accurate measure of form on all surfers in the 2017 season. This is observed since break type was the best strategy from the first 3 events but was overtaken after event 3 by the ML approach.

The cumulative fantasy scores that denote a player's seasonal fantasy score total were obtained for every team-selection strategy in the 2017 WCT season.



(a) With Optimal team



(b) Without Optimal team

Figure 11: Cumulative Fantasy Scores for Each Strategy in 2017

The strategy of selecting a team based on maximising the previous event fantasy points has the lowest accumulated fantasy score of all strategies excluding the random guess. This indicates that surfers are not consistent between events since a surfer that obtains a large number of fantasy points in the previous event does not tend to perform well in the event after. In the first three events, the strategies break type and Reoccurring Stats performed better than ML approach. This is possibly due to the limited amount of data that the model had available in the 2017 season. Post event 3, there were several events worth of data on all surfers in the 2017 season and so there was a more accurate measure of form on all surfers in the 2017 season.

The ML approach can clearly be seen as the best model out of all strategies since there is clear separation between the ML approach and the other strategies (Figure 11B). The strategies Break Type and Reoccurring Stats perform second best. It is also clear that no strategy comes close to the optimal team model, including the ML approach, which is over 1000 fantasy points below the optimal team model and closer to the other team-selection strategies than the optimal (Figure 10A).

5.1.2 Average fantasy points per tier

The average fantasy points in each event were obtained across the three tiers. An average fantasy points obtained for all surfers participating in each event were obtained.

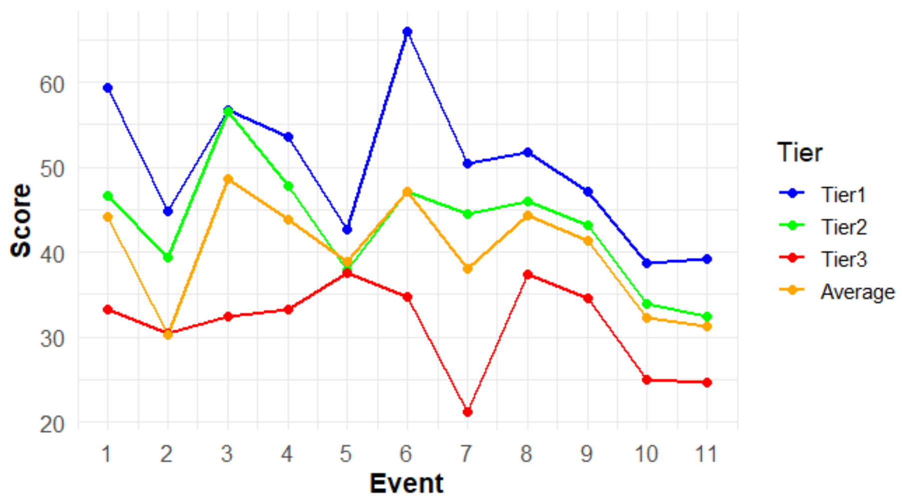


Figure 12: Fantasy Points obtained per event for the three tiers and on average across all surfers in the 2017 season

Tier 1 surfers have higher fantasy scores on average for every event in the 2017 WCT season (Figure 12). Tier 1 and 2 surfers had equal fantasy points for event 3. Tier 3 surfers have the lowest fantasy scores on average for every event in the 2017 WCT season of the three tiers. Tier 1 surfers score 15-20 fantasy points on average more than tier 3 surfers. Event 5 was the only exception where tier 2 and 3 surfers had similar average fantasy points. Overall, event 6 was a very high scoring event where tier 1 surfers had over 30 fantasy points more than tier 3 surfers on average. Event 1, 3 and event 6 have the largest average fantasy scores among tier 1, tier 2 and on average, these three events are points breaks. It is evident that point breaks are the highest fantasy scoring wave break type. Based on these findings, tier 1 is the best tier for power surfer selection.

5.1.3 Power Surfer strategies

The fantasy scores obtained for each power surfer strategy were compared across all events in the 2017 WCT season.

Figure 13: Fantasy Scores for Each Power Surfer Strategy Across Events in 2017

There are differences in fantasy scores for events in the 2017 WCT season, this was observed with several strategies. There are clear differences in fantasy scores depending on the power surfer strategy used (Figure 13). In several cases, the base MILP model which selects a power surfer not dependent on a certain strategy is out-performed by tier 1 only. Restricting the power surfer to tier 1 only out-performs the base MILP approach in 3 of 11 events, with a further 5 events having the same fantasy points score since the base MILP and the tier 1 only strategy chose the same surfer. The tier 1 only strategy in comparison to the other power surfer strategies, performed better than the 2016 event winner strategy in 8 of 11 events and performed equal to or better than the previous event optimal power surfer strategy in 8 of 11 events in the 2017 WCT season. The strategy of choosing the power surfer in one's fantasy team to be a tier 1 surfer is a consistent and top three performing strategy in 7 of 11 events.

Event 5 has a major decrease in fantasy score relative to all other events in the 2017 season. Event 5 is an anomaly since all highly ranked surfers prior to the event performed poorly. Rookies and wildcards in event 5 performed well. Such an event is difficult to predict since all features in historical data were not an indication of surfer performance. Therefore no power surfer strategies were effective in increasing fantasy score since the 2016 winner, tier 1 only and previous event optimal all performed poorly in event 5. Event 4 was an anomaly whereby the random guess chose several high scoring surfers, with one of the power surfers chosen, being a semi-finalist in that event. This random selection was highly effective and performed better than the other power surfer strategies.

5.2 2018 Championship Tour Season

The 2018 WCT season had three different events that were not present in 2017. Event 2 (Margaret River Pro) in Australia was moved to Uluwatu in Indonesia. Event 7 in 2018 was in an artificial wave pool (surf ranch), such an event was the first in competitive surfing and replaced the original event 7 held at Cloudbreak in Fiji in 2017. Event 8 in 2017 was held at Lower Trestles, in 2018 the event was replaced with Keramas in Indonesia. We acknowledge the potential for location variability for these two events when analysing the MILP model on the 2018 WCT season.

Two further key considerations are that the 2018 WCT season had a difference in the number of rounds surfed in each event. In the 2017 WCT season, there were a total of 8 rounds with a round 5 where the losers of the three-man round 4 heats went head-to-head. In 2018, there were only 7 rounds and no longer a round 5. The total fantasy points are therefore expected to be lower in the 2018 season compared to the 2017 season. Event 8 in 2018 is also expected to have a lower event fantasy points tally since only two rounds are surfed with one overall leader board.

5.2.1 Team Selection Strategies

Table 8: Summary of Strategies for 2018 Season

Strategy	**Mean**	**Variance**	**Std_Dev**	**Min**	**Max**	**Total**
Optimal	563.28	12975.69	113.91	256.92	663.62	6196.04
ML approach	348.07	7221.18	84.98	199.83	505.11	3828.80
Current Standings	343.70	4879.09	69.85	182.96	435.10	3780.73
Break Type	342.03	9876.32	99.38	199.83	550.98	3762.30
2017 Score	332.61	5337.70	73.06	205.03	434.51	3658.69
Previous Event	307.44	5688.92	75.42	194.81	460.82	3381.79
Reoccuring stats	307.24	6623.92	81.39	216.54	491.24	3379.69
Random Guess	299.11	3686.65	60.72	151.91	372.33	3290.17

The total seasonal fantasy points decrease significantly across all team-selection steartegies in the 2018 WCT season compared to the 2017 WCT season (Table 8). The optimal team model in the 2018 season has 1500 fantasy points less than the optimal model compared in the 2017 season. We expect all strategies to perform worse in 2018 than 2017. Our ML approach still performs the best out of all Machine Learning strategies and the Break Type method is still the third best team selection strategy overall. The gap between the optimal model and the next best strategy (ML Approach) is 2200 fantasy points, 100 fantasy points larger than in 2017. ML Approach accumulated 48 points more across the season in 2018 than current standings indicating that the gap between the ML Approach and the other team-selection strategies is less in the 2018 season than the 2017 season. The Reoccurring Stats strategy went from the second best strategy to the worst, only marginally edging out the random guess strategy by 90 points in this regard.

Previous Event and the 2017 Score (the updated version of the 2016 Score strategy) still performs quite poorly, ranking 2nd and 3rd last from the viable strategies. 2017 Score has a larger ranking in the 2018 season than in the 2017 season. The biggest improvement across all strategies is the strategy using the current WCT standings at that event, Current Standings. It improved from 4th to 2nd over the season, marginally losing out to the ML approach by 48 points.

As noted, there were new events that were added to the 2018 season compared to the 2017 season. The breakdown of how each strategy performed will be broken down into an overall breakdown, a breakdown for new events and a breakdown for known/previous events.

Table 9: Summary of Strategies for New Events

Strategy	**Mean**	**Variance**	**Std_Dev**	**Min**	**Max**	**Total**
Optimal	491.50	44276.54	210.42	256.92	663.62	1474.49
ML approach	312.01	9526.53	97.60	199.83	377.52	936.02
Current Standings	309.41	14247.18	119.36	182.96	420.12	928.24
2017 Score	301.11	8428.28	91.81	205.03	387.94	903.34
Reoccuring stats	291.86	4750.33	68.92	216.54	351.78	875.58
Previous Event	290.30	7325.44	85.59	194.81	360.10	870.91
Break Type	285.84	6170.54	78.55	199.83	353.79	857.52
Random Guess	260.79	8960.70	94.66	151.91	323.57	782.37

Before discussing results for the new events, it is important to note that for the 2017 Score strategy, aggregated data for the break type of the 2017 season was used in our model formulation for event 4 and 5. For event 8, we used a mixture of the two break types that were most similar for the artificial wave pool. Similar methods were adopted for the ML and Break Type approach for event 4 and 8 and event 5 for the Break Type model, where aggregated data was used for those events. For event 5 of the ML approach, as the event was initially in Margaret river (Event 2 for other seasons) and Uluwatu was also a reef break such as Margaret river, we decided to use 2017 Event 2's formulation for the Event 5 of 2018.

For the events that were new for the 2018 season, we see that Current Standings was the second best strategy for team selection across these events (Table 9), indicating that for unknown events, it may be better to select surfers based on their ranking in the WCT prior to the event. The ML approach was the best strategy across the new events but only marginally. This was unexpected since we only had data for event 5 to train on for these new events. Events 4 and 8 used the same model formulation as the break type method, so the point differential between the two strategies was due to event 5. Break Type performs the worst, but this is due to the wave pool event being a new break type and a complete unknown based on what the strategy aims to use to predict. Reoccurring Stats does not perform as badly in these new events compared to its performance over the whole season, indicating that a mixture of form and how they do in similar breaks to the previous year performs well. Reoccurring Stats performed the best out of all strategies in event 8. Current Standings achieved the highest maximum score out of all strategies across the events but also had the lowest minimum score, indicating that the it's performance could potentially be misleading and highly variable in this aspect.

The known events table is examined to see how the models perform.

Table 10: Summary of Strategies for Known Events

Strategy	Mean	Variance	Std_Dev	Min	Max	Total
Strategy	**Mean**	**Variance**	**Std_Dev**	**Min**	**Max**	**Total**
Optimal	590.19	2850.07	53.39	528.10	659.05	4721.55
Break Type	363.10	10485.64	102.40	226.42	550.98	2904.78
ML approach	361.60	6827.58	82.63	240.13	505.11	2892.78
Current Standings	356.56	2206.64	46.97	276.92	435.10	2852.49
2017 Score	344.42	4632.68	68.06	262.73	434.51	2755.35
Previous Event	313.86	5861.08	76.56	208.04	460.82	2510.88
Random Guess	313.47	1841.29	42.91	240.53	372.33	2507.80
Reoccuring stats	313.01	7966.03	89.25	235.12	491.24	2504.11

The Break Type strategy and ML approach both outperformed the Current Standings approach across the remaining 8 events (Table 10). This makes sense as both these models had historical data on those events and could better predict the performance. The Reoccurring Stats strategy performed worse than the Random guess model indicating that the three most important variables in predicting a surfer's fantasy points in the 2017 WCT season was not a useful strategy for fantasy team-selection in the known events of the 2018 WCT season. The 2017 Score and previous event strategies also performed poorly in the known events of the 2018 WCT season. All 3 Machine Learning approaches did achieve the 3 highest scores for the known events however, once again showing that there is reason to believe events have some underlying characteristics that can help us predict surfer performance and fantasy scores as a result. The strategy Break Type has the largest variance comparative to the others.

Current Standings has the highest minimum score, indicating that it is the safest strategy in terms of consistency, despite not yielding the best average fantasy score. It also has the second lowest variance, only having a larger variance than the random guess strategy. The team-selection strategies were compared to the random guess and optimal team strategies.

Table 11: Strategy Comparison with Mean and Standard Deviation Against Optimal and Random Guess

Strategy	Mean	Std_Dev	Optimal Mean Delta	Optimal Std_Dev Delta	Random Guess Mean Delta	Random Guess Std_Dev Delta
ML approach	348.07	84.98	-215.21	-28.93	48.96	24.26
Previous Event	307.44	75.42	-255.84	-38.49	8.33	14.70
2017 Score	332.61	73.06	-230.67	-40.85	33.50	12.34
Reoccurring Stats	307.24	81.39	-256.04	-32.52	8.13	20.67
Current Standings	343.70	69.85	-219.58	-44.06	44.59	9.13
Break Type	342.03	99.38	-221.25	-14.53	42.92	38.66

In compared to the optimal mean points per round, each strategy ranges from being 200 to 260 points worse off per round (Table 11). All models were less variable than the optimal team, indicating that the optimal fantasy points a team can obtain changes between events. This is linked to the inconsistency in surfer performance between successive WCT events. Every WCT event is in a different location and is subject to different environmental conditions, thus a player's fantasy team differs in performance between events.

Compared to the random guess models, 2 strategies are clearly not as good as the others, with Previous event and Reoccurring Stats attaining a 8 more fantasy points per event than the random guess. The other strategies achieve north of 30 points more per event. The main caveat to take away is that all models were more variable than the random guess model. This was not expected as random guesses are assumed to be the most variable out of all the strategies based off intuition that it is the most random and in theory, the most variable. The Break Type was by far the most variable out off all the strategies, achieving a standard deviation of 20 points more than the second highest, Reoccurring Stats.

The strategy breakdown for each event in the 2018 WCT season will be assessed to see how the models compare.

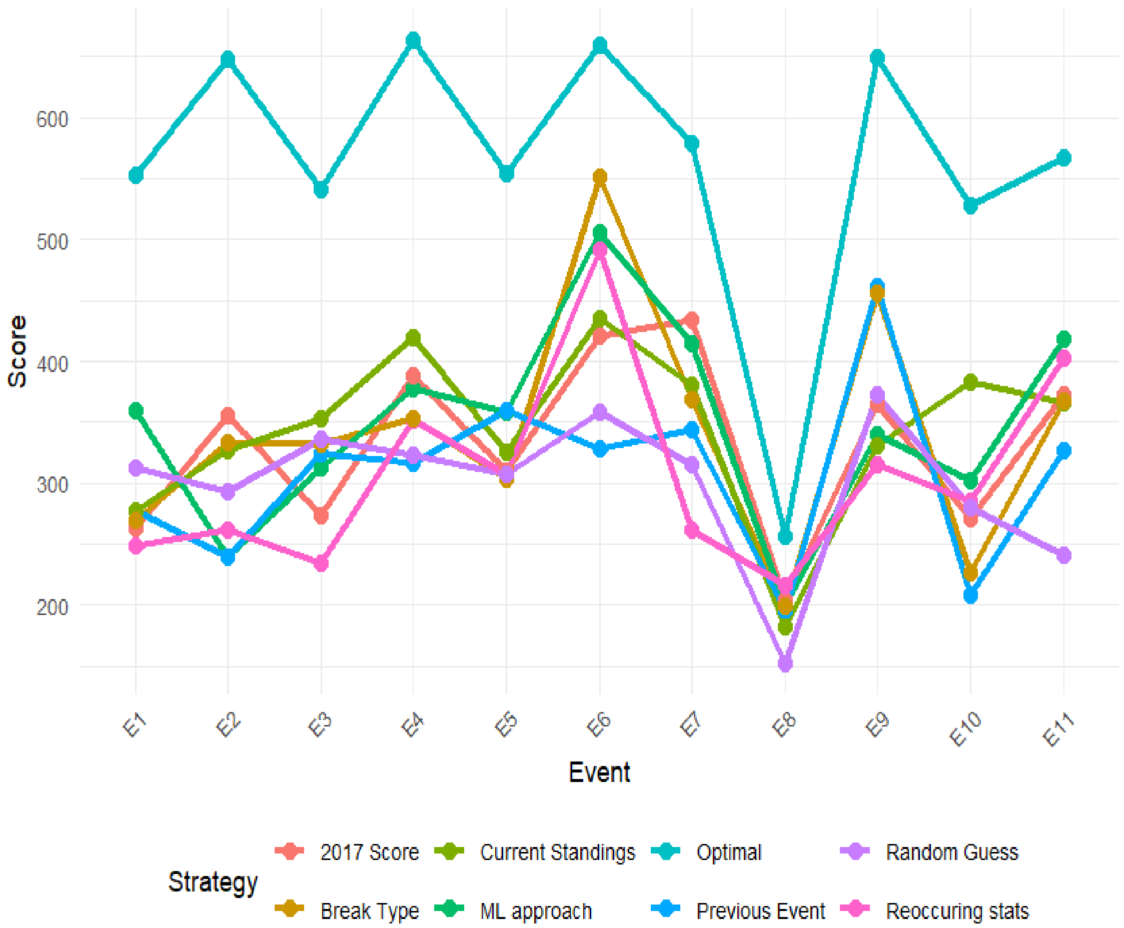


Figure 14: Fantasy Scores per Event for Each Strategy in 2018

There is a major dip of points in event 8 and this is due to there only 2 rounds of surfing in that event. Most strategies again perform well in event 6 just as it was in 2017. This shows that Event 6 is generally an easier event to predict in this regard, with all three machine learning models performing well. Unlike in 2017, the ML approach comfortably performed the best in event 1, whereas in 2017, it performed the worst. Events 10 and 11 were the most variable events in terms of points scored between the different strategies. This was the case in 2017 as well, where event 10, a beach break, was variable in both the 2017 and 2018 WCT seasons (Figure 10). This shows that these events have certain characteristics that makes the event more unpredictable as no strategies dominated in both years at this event. The major takeaway is that the strategies are close throughout the season, where there does not appear to be a strategy that dominates over the course of the year as was the case in 2017.

The strategies were ranked between events, where their placements across the season relative to each other were kept track of (Figure 15).

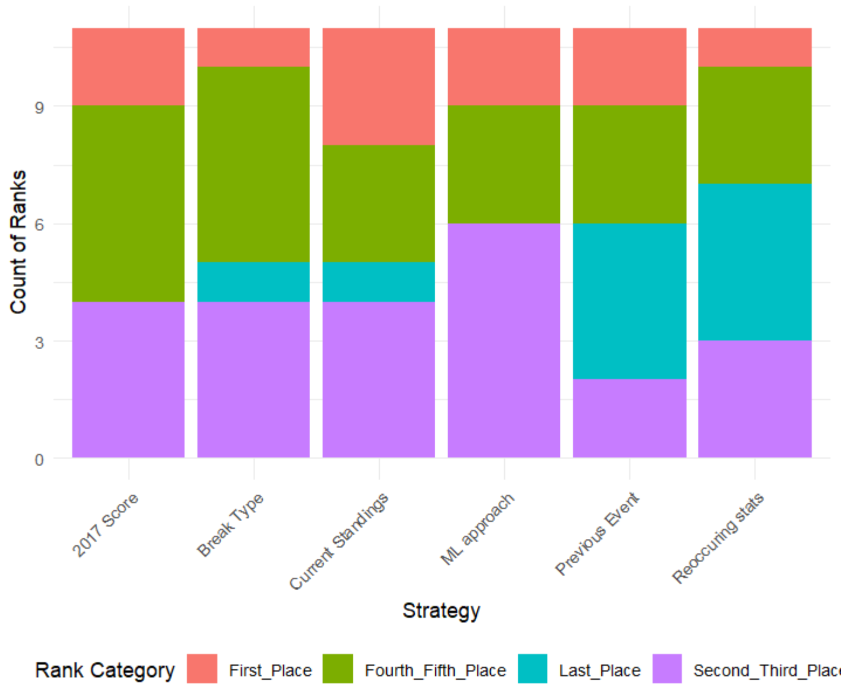
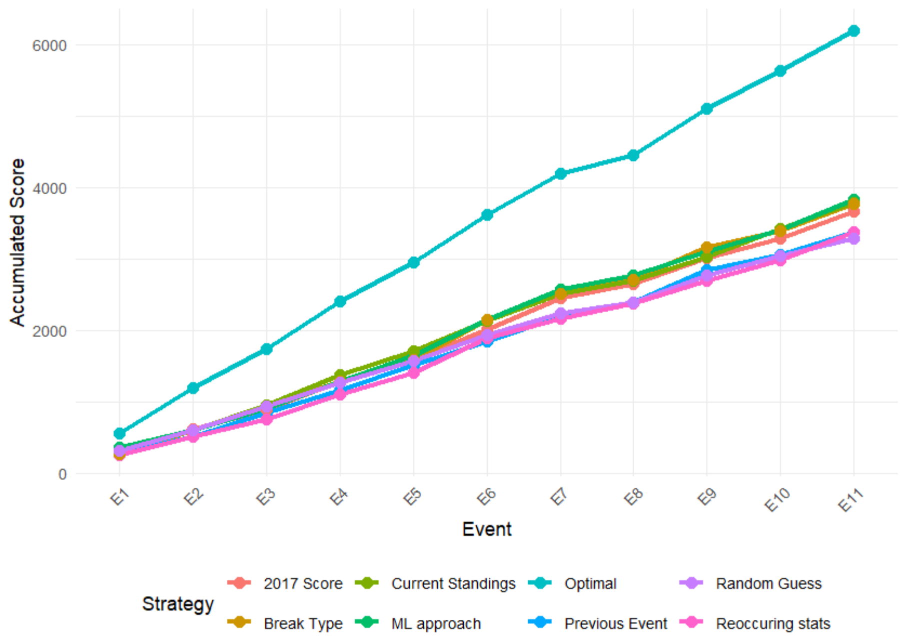
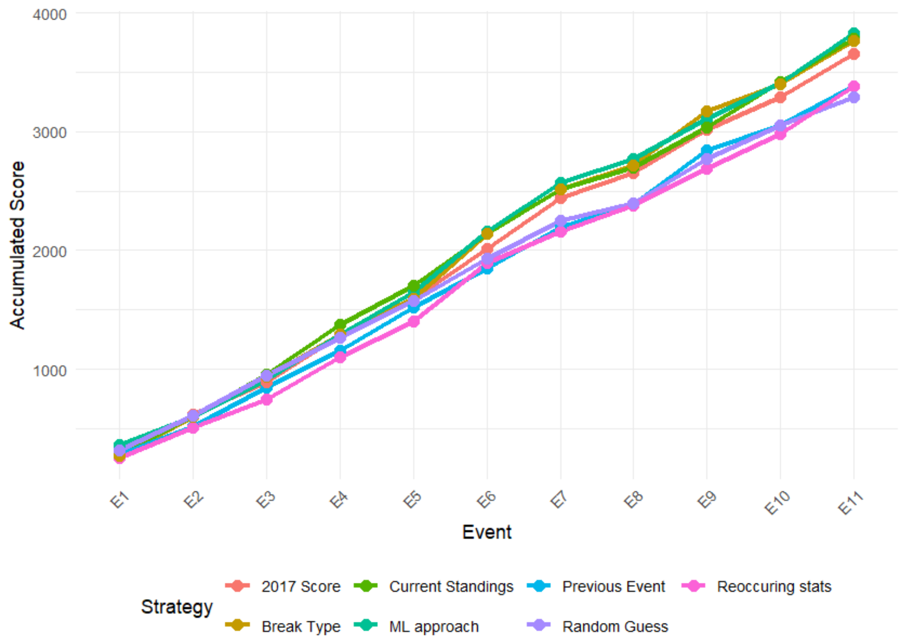


Figure 15: Strategy ranking based on a count of ranks for the various team-selection strategies

Although the ML approach performed the best overall, it only ranked first out of all the strategies twice over the course of the season, with Current Standings ranking first the most (Figure 15). The ML approach however ranked 2nd or third the most often as well as never finishing in last throughout the season, showing that although not having as low standard deviation as Current Standings, it was more consistent compared to the other strategies. Previous event came last place the most, tied with Reoccurring Stats, showing that picking a surfer based on their previous event performance is not the best strategy. Break type also performs poorly in the sense of occurring in the top 3 strategies in 5 of 11 events. 2017 Score performed well in this sense as it never finishing last while also finishing in the top 3 strategies. A cumulative fantasy scores line graph (Figure 16) was obtained to assess how the strategies performed in the 2018 WCT season.



(a) With Optimal team



(b) Without Optimal team

Figure 16: Cumulative Fantasy Scores for Each Strategy in 2018

Throughout the season it is very close between the all strategies with there being two groups of strategies with clear separation in their accumulated fantasy score (Figure 16). The strategies 2017 Score, Break Type and Current Standings perform better than Reoccurring Stats and Previous Event since these three strategies clearly pull away from the rest post-event 6. Current Standings is the preferred strategy in the first 5 events. The ML approach and Break Type strategy hustle for 1st and 2nd spot from event 6 to 11 along with Current Standings. At the end of the season, Break Type and ML Approach have a very similar seasonal fantasy score total indicating that these two strategies have very similar predictive power.

What is clear from this graph is that no strategy comes close to the optimal model while our ML approach is not as clear a winner as it was in the 2017 season.

5.2.2 Power Surfer strategies

The power surfer strategies were assessed per event for the 2018 WCT season.

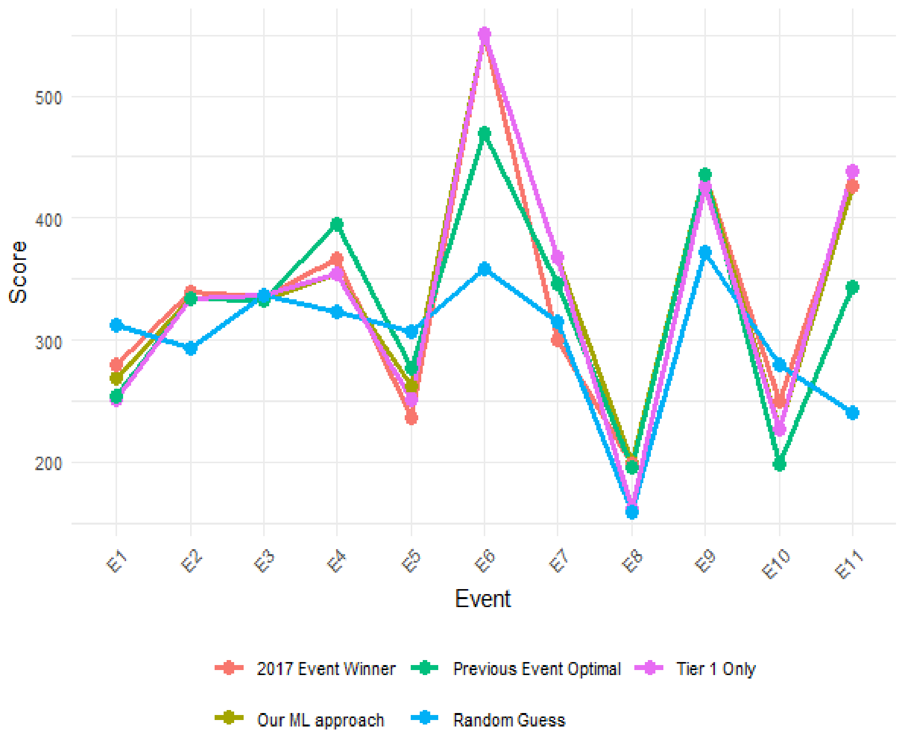


Figure 17: Fantasy Scores for Each Power Surfer Strategy in 2018

Tier 1 only and 2017 event winner perform joint best in 5 of 11 events in 2018 WCT season (Figure 17). Random Guess and Previous Event optimal perform the worst with large discrepancies in fantasy scores for event 6 and event 11. An important distinction is that in event 2, 3, 6 and 9, the 2017 event winner was in tier 1 and the above strategies opted for the same power surfer. The ML approach, 2017 event winner and tier 1 are the best strategies, however these strategies were further compared with an average ranking to better assess and contrast them.

5.3 Strategy Ranking and Performance Comparison

All strategies were further compared with an average ranking and an associated standard deviation. The process was done for both the team selection and power surfer selection strategies in the 2017 and 2018 WCT seasons.

Table 12: Average Ranking and Standard Deviation for the Team Selection Strategies and Random Guess for the 2017 WCT season

Strategy	Avg Ranking	Std Dev
ML Approach	2.18	1.89
Reoccurring Stats	3.14	1.42
Break Type	3.73	2.15
2016 Score	4.18	1.72
Current Standings	4.23	1.54
Previous Event	4.64	2.05
Random Guess	5.91	1.38

Table 13: Average Ranking and Standard Deviation for the Team Selection Strategies and Random Guess for 2018

Strategy	Avg Ranking	Std Dev
ML Approach	3.23	1.60
Current Standings	3.23	1.85
Break Type	3.41	1.96
2017 Score	3.64	1.86
Reoccurring Stats	4.64	2.20
Random Guess	4.82	1.89
Previous Event	4.95	2.26

The Random guess is the least variable of all strategies in the 2017 WCT season, however it had the lowest ranking of all strategies in both the 2017 and 2018 seasons (Table 12 and Table 13). The team-selection strategy of using a surfer's Current Standings, a measure of form coming into the event, performed better in the 2018 season than the 2017 season. The ML approach had a lower average ranking in the 2018 season compared to the 2017 season, indicating it is less effective in the 2018 season. We believe the ML approach performed worse in the 2018 season relative to the 2017 season largely due to variability in surfer performance between the seasons. Rookie surfers performed better in the 2018 season relative to the 2017 season. Several top ranked surfers that performed consistently and were selected by the ML approach in 2017 were less consistent in their performances in 2018. The ML approach was likely also affected by the changes in the WCT event schedule from 2017 to 2018. The ML approach is the best ranked strategy on average for both seasons, however it is less effective in 2018 and the random guess is the worst ranked strategy for both seasons. Break type has similar average ranking in both seasons and performs consistently in the top three of all strategies. The strategy of selecting a team based on the performance of surfers in the previous event performed second to last above the random guess in the 2017 seasons and the worst in the 2018 season. Team selection in fantasy surfing is not effective when selecting a team based on the previous event.

The 2017 score had a higher ranking in the 2018 season than the 2016 score did in the 2017 season, this indicates that there is more consistency or similarity between how surfers performed in 2018 relative to 2017 than in 2017 relative to 2016. This further supports literature on seasonal variability and how every season, surfer performance fluctuates. Reoccurring Stats performed second-best in the 2017 season, in 2018 the average ranking of Reoccurring Stats dropped by 1.5 units. The top three variables of importance in determining surfer performance in the 2017 season are therefore not as effective in fantasy team selection in 2018. This further supports the argument of seasonal variability in professional surfing performance and the difficulty in predicting surfer performance. The average ranking and standard deviation of the power surfer strategies were obtained for the 2017 and 2018 WCT season.

Table 14: Average Ranking and Standard Deviation for Power Surfer Strategies in 2017

Strategy	**Avg Ranking**	**Std Dev**
ML Approach	2.23	1.01
Tier 1 Only	2.41	1.14
Previous Event Optimal	2.91	1.14
2016 Event Winner	3.18	1.25
Random Guess	4.27	1.42

Table 15: Average Ranking and Standard Deviation for Power Surfer Strategies in 2018

Strategy	**Avg Ranking**	**Std Dev**
2017 Event Winner	2.59	1.28
ML Approach	2.73	1.10
Tier 1 Only	2.95	1.37
Previous Event Optimal	3.18	1.17
Random Guess	3.55	1.86

The ML approach and Tier 1 only are the best power surfer strategies in the 2017 season (Table 14). The 2017 Event Winner, ML approach and Tier 1 only are the best power surfer strategies in the 2018 season (Table 15). The Previous year Event Winner strategy performs worse in 2017 (2016 winner) than 2018 (2017 winner). This implies that the 2017 event winners are more consistent in that same event in the 2018 season than the season prior. The ML approach is the least variable of all power surfer strategies across both seasons, however in 2018 the ML approach has a lower average ranking. The previous event optimal power surfer strategy has a lower standard deviation than the other strategies, excluding the ML Approach, however performs more poorly and has a lower average ranking in 2018 relative to 2017. Tier 1 only performs worse in 2018 relative to 2017. This is partly due to the fact that it is more variable in the 2018 season and fluctuates more significantly in 2018 relative to 2017. The variability in strategy ranking between the 2017 and 2018 season support previous literature stating that surfing performance is difficult to predict since every wave is different and there exists more individual variability in competitive surfing compared to other sports (Mendez-Villanueva, 2010).

6 Conclusion and Recommendations for further Research

In conclusion, several MILP models were constructed to aid players in the decision making process of selecting a fantasy team line-up in the 2017 and 2018 WCT seasons. This study was effective in finding important features for predicting surfer performance. We are confident that further features exist that were not examined and tested in this study since the features investigated in this study were not exhaustive. This study aims to be used in conjunction with future work in improving the literature on competitive surfing in the context of fantasy sports. Several features not investigated in this study that have potential to be influential in predicting surfer performance on the WCT are age, wave size, wave period and surfboard manufacturer. This study created MILP models for every event since there was large variability observed within each event and between different break types, this was observed since the important features in determining a surfer's fantasy score differed between events and break types. One such recommendation is for future studies to view event number and break type as a mixed effects variable, this serves as an alternative method in assessing variable importance on the WCT. Other recommendations are to analyse individual surfers and their performance across the season as we believe this would improve model performance, this study was unable to make such analyses due to ethical constraints.

This study examined various strategies that a decision maker can implement when selecting a line-up in fantasy surfing. The outcome of all strategies deemed that certain strategies worked better than others. This study implemented said strategies for two respective seasons. The strategies break type and ML approach performed far better than a random guess however there are improvements required to reach the fantasy scores obtained by the optimal team. One such line of work that would be very useful is examining female competitive surfers on the WCT and studying differences between features of importance in male and female competitive surfing.

The findings of this study supported the argument that similar features are important for different wave break types. Surfer performance can loosely be broken down to the wave break type that the event is being surfed at, this was found for both the 2017 and 2018 seasons. When selecting the power surfer, this study concludes that two useful strategies are to select the power surfer as a tier 1 surfer or a past event winner the previous season.

We acknowledge that seasonal variability exists in competitive surfing and this study supported past literature since the important features in determining surfer performance across a WCT season differed between 2017 and 2018. Further recommendations are for one to implement said strategies on a large number of seasons and observe the influence of seasonal fluctuations in team selection strategies and surfer performance. In particular, the strategy of selecting a team line-up based on the previous events' fantasy points performed very poorly. The strategy of selecting the power surfer for an event as the optimal power surfer of the event prior performed badly as well. This supports the findings of Mendez-Villanueva et al. (2010) that found surfer performance in WCT events to be inconsistent in consecutive events. This study further supports and concludes that Mixed Integer Linear Programming can be used in a wide range of Fantasy Sports and highlights its usefulness and the success of these models in various fields of research.

References

[1] Al-Shboul, R., Syed, T., Memon, J., & Khan, F. (2017). Automated player selection for a sports team using competitive neural networks. *International Journal of Advanced Computer Science and Applications*, 8(8).

[2] Barlow, M.J., Rowe, J., Ruffle, O., Davidson, M. and O'hara, J., 2016. Anthropometric and performance perspectives of female competitive surfing. *Human Movement*, 17(3), pp.154-161.

[3] Beal, R., Norman, T.J., & Ramchurn, S.D. (2020). Optimising daily fantasy sports teams with artificial intelligence. *International Journal of Computer Science in Sport*, 19(2), 21-35.

[4] Becker, A., & Sun, X.A. (2016). An analytical approach for fantasy football draft and lineup management. *Journal of Quantitative Analysis in Sports*, 12(1), 17-30.

[5] Bentéjac, C., Csörgő, A. and Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54, pp.1937-1967.

[6] Billings, A., & Ruihley, B. (2013). *The fantasy sport industry: Games within games.*

[7] Booth, D., 1995. Ambiguities in pleasure and discipline: the development of competitive surfing. *Journal of Sport History*, 22(3), pp.189-206.

[8] Conley, B.F., 2018. How the rise of the daily fantasy sports industry can catalyze the liberalization of sports betting policies in the United States. *Buff. L. Rev.*, 66, p.715.

[9] Das, K.R. and Imon, A.H.M.R., 2016. A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics*, 5(1), pp.5-12.

[10] Drayer, J., Shapiro, S.L., Dwyer, B., Morse, A.L. and White, J., 2010. The effects of fantasy football participation on NFL consumption: A qualitative analysis. *Sport management review*, 13(2), pp.129-141.

[11] Farley, O.R., Abbiss, C.R. and Sheppard, J.M., 2017. Performance analysis of surfing: a review. *The Journal of Strength Conditioning Research*, 31(1), pp.260-271.

[12] Fonti, V. and Belitser, E., 2017. Feature selection using lasso. *VU Amsterdam research paper in business analytics*, 30, pp.1-25.

[13] Forsyth, J.R., De La Harpe, R., Riddiford-Harland, D.L., Whitting, J.W., & Steele, J.R. (2017). Analysis of scoring of maneuvers performed in elite men's professional surfing competitions. *International Journal of Sports Physiology and Performance*, 12(9), 1243-1248.

[14] Hutt, J.A., Black, K.P. and Mead, S.T., 2001. Classification of surf breaks in relation to surfing skill. *Journal of coastal research*, pp.66-81.

[15] Karloff, H. and Karloff, H., 1991. The simplex algorithm. *Linear programming*, pp.23-47.

[16] Longuet-Higgins, M.S. (1983). Wave set-up, percolation and undertow in the surf zone. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 390(1799), 283-291.

[17] Leishman, N.C. (2016). Daily fantasy sports as game of chance: Distinction without a meaningful difference. *BYU L. Rev.*, 1043.

[18] Mendez-Villanueva, A., Mujika, I., & Bishop, D. (2010). Variability of competitive performance assessment of elite surfboard riders. *The Journal of Strength & Conditioning Research*, 24(1), 135-139.

[19] Natekin, A. and Knoll, A., 2013. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7, p.21.

[20] Ploeg, A.J. and Ruihley, B.J., 2021. Genealogies of Fantasy Sport. *The International Journal of the History of Sport*, 38(1), pp.1-6.

[21] Robinson, C. (2020). The prediction of fantasy football.

[22] Ruihley, B.J. and Chamberlin, J., 2021. The history and evolution of the fantasy sport voice: An oral account of the major aspects forming the fantasy sports and gaming association. *The International Journal of the History of Sport*, 38(1), pp.135-151.

[23] Scarfe, B.E., Elwany, M.H.S., Mead, S.T. and Black, K.P., 2003. *The science of surfing waves and surfing breaks-a review.*

[24] Tavana, M., Azizi, F., Azizi, F., & Behzadian, M. (2013). A fuzzy inference system with application to player selection and team formation in multi-player sports. *Sport Management Review*, 16(1), 97-110.

[25] Trninić, S., Papić, V., Trninić, V., & Vukičević, D. (2008). Player selection procedures in team sports games. *Acta Kinesiologica*, 2(1), 24-28.

[26] Wedding, C.J., Woods, C.T., Sinclair, W.H. and Leicht, A.S., 2022. Operational insights into analysing team and player performance in elite Rugby League: a narrative review with case examples. *Sports Medicine-Open*, 8(1), p.140.

7 Appendix

7.1 Appendix A1: World Surf League Fantasy and Public Github Page containing data for 2014 to 2016 WCT seasons

The link to the World Surf League Championship Tour Fantasy Page: World Surf League Fantasy Team.

The link to the Public Github (Rayblick, 2017) Page: Fantasy-Surf-League.

All data and R-files can be found in the following Github Repository Fantasy-Surfing-Project

7.2 Appendix A2: General structure For Lasso Modelling

The code to conduct each of these events/break types after reading in the data and making alterations are listed below:

```
library(glmnet)
x <- model.matrix(`Actual Points` ~ ., Data)[,-1]
dim(x)
y <- Data$`Actual Points`

lassoMod <- cv.glmnet(x, y, alpha=1)
coef(lassoMod, s="lambda.min")
plot(lassoMod)
```

7.3 Appendix B: Tables of coefficient scores for lasso modelling

Table 16: Event 1 Coefficients

Variable	Coefficient
(Intercept)	-68.0831254
HWP Event 1 2014	0
HWP Event 1	0
Heats Surfed Event 1	0
HWP Event 1 2015	0
HWP Event 1 2016	0
Heats Surfed 2016	0.8119053
Local Surfer	0
National Advantage: 1	0
2014 Placement	0
2016 Placement	0
2015 Placement	0
Number of Excellent Heats 2016 - Event 1	0
Number of Excellent Waves 2016 - Event 1	0
Average Heat Score	0
Total Heats Surfed	0
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	1.3013433
2016 Score	0
2016 Heats Surfed	0
Max Heat Score	3.3996906
Excellent Heats Surfed	0
Years Surfing	0.5940917
Experience: Rookie	0
Experience: Veteran	0

Table 17: Event 2 Coefficients

Variable	**Coefficient**
(Intercept)	33.464
HWP Event 2 2014	0
HWP Event 2	0
Heats surfed event 2	0
HWP Event 2 2015	0
HWP Event 2 2016	0
Local Surfer: 1	0
National Advantage: 1	0
2014 Placement	0
2016 Placement	0
2015 Placement	0
Number of Excellent heats 2016 - Event 2	0.381
Average Heat Score	0
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0.166
2015 Heats Surfed	0
2016 Heat Score	0
2016 Score	0
2016 Heats surfed	0
Max Heat Score	0
Previous Event Points	0
Previous Heat Score	0
Experience: Rookie	0
Experience: Veteran	0
Years Surfing	0

Table 18: Event 3 Coefficients

Variable	**Coefficient**
(Intercept)	-103.805
Surfer stance: Regular	0
HWP Event 2 2014	0
Heats surfed event 2	0
HWP Event 2 2015	0
HWP Event 2 2016	0
National Advantage	0
Nationality: 1	0
2014 Placement	0
2016 Placement	0
2015 Placement	0
Number of Excellent heats 2016 - Event 2	0
Average Heat Score	8.024
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	3.624
2016 Score	0
2016 Heats surfed	0
Max Heat Score	0
Previous 2 Event Points	0
Previous Heat Score	0
Previous Heat Score	0.035

Table 19: Event 4 Coefficients

Variable	Coefficient (s1)
(Intercept)	12.0842
Surfer Stance: Regular	0.9897
Heats surfed event 8	0
National Advantage	-3.9750
2014 Placement	0
2015 Placement	0
2016 Placement	0
Number of Excellent heats 2016 - Event 8	0
Average Heat Score	1.7353
2014 Average Heat Score	0
2014 Score	0.3537
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0
2015 Heats Surfed	-3.1648
2016 Heat Score	0.6518
2016 Score	0
2016 Heats Surfed	0
Max Heat Score	0
Previous 7 heat average	0
Previous Heat score	0
Previous 7 Event Points	0

Table 20: Event 5 Coefficients

Variable	**Coefficient**
(Intercept)	34.025
Surfer Stance: Regular	0
Heats surfed event 5	0
2014 Placement	0
2015 Placement	0
2016 Placement	0
Number of Excellent heats 2016 - Event 5	0
Average Heat Score	0
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	0
2016 Score	0
2016 Heats surfed	0
Max Heat Score	0
Previous 4 heat average	0
Previous Heat score	0
Previous 4 Event Points	0

Table 21: Event 6 Coefficients

Variable	**Coefficient**
(Intercept)	20.531
Surfer Stance: Regular	0
Heats surfed event 6	0
National Advantage	0
2014 Placement	0
2016 Placement	0
2015 Placement	0
Number of Excellent heats 2016 - Event 6	0
Average Heat Score	0
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	0
2016 Score	0.335
2016 Heats surfed	0
Max Heat Score	0
Previous 5 heat average	0
Previous Heat score	0
Previous 5 Event Points	0.001

Table 22: Event 7 Coefficients

Variable	**Coefficient**
(Intercept)	-13.341
Surfer Stance: Regular	0
Heats surfed event 7	0.052
National Advantage	0
2014 Placement	0
2016 Placement	-0.019
2015 Placement	0
Number of Excellent heats 2016 - Event 7	0.206
Average Heat Score	0
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	2.903
2016 Score	0
2016 Heats surfed	0
Event Part 2016?	0
Max Heat Score	0
Previous 6 heat average	0.308
Previous Heat score	0.167
Previous 6 Event Points	0

Table 23: Event 8 Coefficients

Variable	**Coefficient**
(Intercept)	5.968
Surfer Stance: Regular	0
Heats surfed event 8	0
National Advantage	-3.686
2014 Placement	0
2015 Placement	0
2016 Placement	0
Number of Excellent heats 2016 - Event 8	0
Average Heat Score	0
2014 Average Heat Score	0
2014 Score	0.446
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	-0.166
2015 Heats Surfed	0
2016 Heat Score	2.452
2016 Score	0
2016 Heats Surfed	0
Max Heat Score	0
Previous 7 heat average	0
Previous Heat score	0
Previous 7 Event Points	0

Table 24: Event 9 Coefficients

Variable	**Coefficient**
(Intercept)	-60.914
Surfer Stance: Regular	-1.272
Heats surfed event 9	0
National Advantage	14.821
2014 Placement	-0.078
2016 Placement	0
2015 Placement	0
Number of Excellent heats 2016 - Event 9	0
Average Heat Score	2.532
2014 Average Heat Score	5.780
2014 Score	0.285
2014 Heats Surfed	0
2015 Heat Score	0.656
2015 Score	0
2015 Heats Surfed	-1.466
2016 Heat Score	0
2016 Score	0.119
2016 Heats Surfed	0
Max Heat Score	0
Previous 8 heat average	0
Previous Heat score	-0.326
Previous 8 Event Points	0

Table 25: Event 10 Coefficients

Variable	**Coefficient**
(Intercept)	7.818
Surfer Stance: Regular	0
Heats surfed event 10	0
National Advantage	0
2014 Placement	0
2016 Placement	-0.281
2015 Placement	0
Number of Excellent heats 2016 - Event 10	0
Average Heat Score	0
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	-2.024
2015 Heat Score	0.898
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	0
2016 Score	0
2016 Heats Surfed	0
Max Heat Score	0
Previous 9 heat average	0.382
Previous Heat score	0.225
Previous 9 Event Points	0

Table 26: Event 11 Coefficients

Variable	**Coefficient**
(Intercept)	-10.249
Surfer Stance: Regular	0
Heats surfed event 4	0
National Advantage	17.023
2014 Placement	0
2016 Placement	-0.367
2015 Placement	0
Number of Excellent heats 2016 - Event 11	0
Average Heat Score	0
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	1.659
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	0
2016 Score	0
2016 Heats Surfed	0
Max Heat Score	2.012
Previous 10 heat average	0
Previous Heat score	0.049
Previous 10 Event Points	0

Table 27: Point Break Model Coefficients

Variable	**Coefficient**
(Intercept)	-0.030
National Advantage: 1	0
2014 Placement	0
2016 Placement	0
2015 Placement	0
Excellent Heats	0
Average Heat Score	0
2014 Average Heat Score	0.050
2014 Score	0.010
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	0.180
2016 Score	0.000
2016 Heats Surfed	0
Max Heat Score	0.210
Average Previous Heat Score	0
Championship Score	0.170

Table 28: Reef Break Model Coefficients

Variable	**Coefficient**
(Intercept)	-0.020
Surfer Stance: Regular	0
Total Heats	0
National Advantage: 1	0
2014 Placement	0
2016 Placement	0
2015 Placement	0
Excellent Heats	0.160
Average Heat Score	0
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	0
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	0
2016 Score	0
2016 Heats Surfed	0
Max Heat Score	0.160
Average Previous Heat Score	0.030
Previous Heat Score	0.080
Championship Score	0.010

Table 29: Beach-Break Coefficients

Variable	**Coefficient**
(Intercept)	-0.000
Surfer Stance: Regular	0
National Advantage: 1	0
2014 Placement	-0.083
2016 Placement	0
2015 Placement	-0.041
Average Heat Score	0.028
2014 Average Heat Score	0
2014 Score	0
2014 Heats Surfed	0
2015 Heat Score	0.094
2015 Score	0
2015 Heats Surfed	0
2016 Heat Score	0
2016 Score	0
2016 Heats Surfed	0
Max Heat Score	0
Average Previous Heat Score	0.023
Previous Heat Score	0
Championship Score	0

7.4 Appendix C: Nationality Table

Country	**Abbreviation**
Australia	AUS
Japan	JAP
Brazil	BRA
France	FRA
Tahiti	TAH
Hawaii	HAW
Italy	ITA
Portugal	POR
United States	USA
South Africa	ZAF

Table 30: List of Countries and their Abbreviations

7.5 Appendix D: University of Cape Town Research Ethics Committee Approval Letter

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

2024/08/19

SCI/00938/2024

RE: Research Ethics Committee Project Approval Letter

Dear Neil Watson,

Your application for ethics review of your project titled

Performance analysis of professional surfers in the World Surfing League

has been reviewed and evaluated by the

Science Research Ethics Committee.

You may proceed with your research project titled:

Performance analysis of professional surfers in the World Surfing League

Expiration date of approval: 2026/08/19

Please note that should:

(i) any serious or adverse effects to participants occur and/or,

(ii) aspect(s) of your current project change and/or

(iii) any unforeseen events that might affect continued ethical acceptability of the project occur then you should immediately report this to the approving REC. You may be required to submit an amendment to this application, in order to determine whether the changed aspects increase the ethical risks of your project.

Based on the information supplied your application has been successful and is approved.

Please note the following additional conditions associated with this approval:

(i) None.

Regards,

Science Research Ethics Committee.