Introduction

In the first ever UFC events the rules were simple: there are no rules. This was almost entirely true, and two fighters were essentially just thrown into a cage until one could no longer continue. Eventually, by UFC 5 it was realized that this format was not always effective, so a 30 minute time limit was added for these bouts that continued to stall. However, it was quickly realized that declaring these fights a draw that were stopped due to time was a very poor resolution for fans, so judges were added in UFC 8 to declare a winner in these fights. MMA has continued to evolve into more of a sport, and eventually by UFC 21 five minute rounds with a ten-point must scoring system were introduced, making scoring much more similar to what we see today. The role of a judge is extremely critical: just under half of all fights go to a decision so these judges can easily influence a fighter's entire career by the way they score just one round.

The addition of ten point must scoring system was essentially just implementing boxing's scoring system within MMA. However, while boxing judging is no simple matter, the introduction of this scoring system to the UFC opened a whole world of questions. Unlike boxing, where a fighter essentially has two basic options (a punch to the head or a punch to the body), MMA fighters have countless techniques that can be used. Just looking at striking, there is a new target (the leg) as well as the introduction of kicks, knees and elbows. And when grappling is involved in a fight it becomes even more complicated, as a judge must now decide how to score various actions such as body shots, head kicks, clinch knees, takedowns, submission attempts and ground & pound relative to each other. This seems like an almost impossible task.

As you can imagine, judging of fights became very ambiguous and in the early days judges were essentially at liberty to score fights however they want. There have been many changes to judging criteria since they were introduced, and commissions have attempted to specifically outline what judges should be looking for when scoring fights. While the art of judging will always remain subjective, there has certainly been improvements in terms of communicating what judges should be looking for.

One of these largest changes to scoring criteria occurred in 2012. When this change occurred, effective defense was removed from the scoring criteria as it was decided it should be all about offense. Striking and grappling are now supposed to be scored "with equal weight," whereas the previous criteria for scoring put more emphasis on striking. In terms of scoring striking, judges are now told to put more emphasis on how damaging strikes are as opposed to just how many are landed. Additionally, the scoring for grappling was more clearly outlined and "Takedowns, reversals, submissions, transitions, activity and threatening moves from the fighter on the bottom and attempted submissions that lead to the threatened fighter being tired" are now all points of emphasis in scoring.

While the clarification of this criteria in 2012 helped take some ambiguity away from judging, there is still plenty of subjectivity in judging and fans are constantly left questioning how a judge could have possibly scored a fight the way they did. In this analysis, I plan on examining the judging of rounds within the UFC. My main goal is to identify how important each of the judging criteria are when scoring these fights. Additionally, once I have identified what is important overall, I plan on examining some of the judges individually to see how their scoring preferences differ.

Literature Review:

This literature review covers past research involving Mixed Martial Arts (MMA) and other combat sports. I begin by examining research that analyzes judging within MMA, and determining which factors are most impactful. Still focusing on just MMA, this will then be expanded to identifying the main factors existent in successful fighters as well as key performance indicators for winning a fight. Mixed martial arts fights that did not result in a decision will then be considered, and I will highlight some research focused on these fights that resulted in finishes.

I will then expand to focusing on research that covers other combat & non-combat sports. I will start by examining papers that classify judging or winning performances in boxing and Muay Thai, similarly to much of the research examined for MMA. I will then showcase research focused on a potential handedness advantage within MMA, other combat sports, and tennis. Finally, I will examine studies of home advantage in MMA considering both fight location and judge nationality.

MMA Judging

Wimser (2021) wrote an article designed to study the factors that impact round judging within the UFC. Data was obtained from UFCStats, and individual round scoring was modeled using the difference between the two fighters for various fight statistics. The data was then normalized for logical comparison of coefficients. Significant strike disparity and takedown disparity were by far the two most impactful, with coefficients of 1.4531 and 1.2546 respectively. The next two where control time and knockdowns, having coefficients of 0.6753 and 0.647, so takedowns and significant strikes have the largest impact. The model had an r squared of 0.7168, and correctly predicted the winner of a fight (not a round) 85.5% of the time. When examining where the model fails, not surprisingly, this tends to be in cases where the metrics are close. There is certainly room for some improvement, but it would also be extremely difficult, if not impossible, to isolate the factors that led to winning these rounds. These rounds were likely won by the impact of shots or one fighter coming close to finishing the fight, which is almost impossible to quantify.

Somewhat like Wimser's model, JudgeAI is a machine learning algorithm created by Nate Latshaw (2021) designed to use publicly available statistics to judge UFC rounds. UFC fights that went to a decision from 2011-2020 were examined in the model. Fight metrics were also obtained from UFCStats.com, and 31 features are examined in the model (all as the difference between the red and blue corners). Time series cross validation was used to evaluate the model to account for the fact that judge and fighter tendencies can change over time. A random forest model was created with the scoring accuracy as a metric to evaluate the model. This is calculated as the percent of total rounds where the judges' score matched the models' predicted score (this means 10-9 or 10-8 was given correctly as well as predicting the fighter). The model's predicted score matched at least one judge's score in 90.8% of rounds, and predicted the same winner as one of the judges in 92.3% of rounds, meaning that no judges agreed with the model's predicted winner in only 7.7% of rounds. In terms of the specific predictors, the most impactful in the model were differences in significant and total strikes landed. Difference in control time was the next highest, and many of the other significant predictors had to do with striking placement and accuracy.

Feldman (2020) wrote a paper to analyze whether MMA judges accurately follow the criteria outlined by the ABC MMA rules committee. It examined data from 2000-2015, and included data from the UFC, Strikeforce and the WEC. The models found that differences in striking, control time, damage, and takedowns were all statistically significant factors to winning a round, and submission attempts were not statistically significant. Knockdowns and takedowns increased the chances of winning the round by the most. In terms of different types of strikes, it was found that all types of significant strikes are statistically significant predictors aside from ground strikes. The ABC MMA rules state that aggression should be used as a factor in scoring fights, but only when effective striking and grappling is very close. Another model was created which only included split decisions to see if judges follow these criteria. The model found that judges rewarded throwing more strikes in these close rounds but did not seem to reward grappling aggression through increased takedown attempts. However, it is also worth noting that even if it is indicative of increased aggression, a failed takedown attempt should be points for the other fighter as defending a takedown is considered effective grappling.

Holmes et al. (2024) used hierarchical models to model MMA judging. Differences in fight metrics from UFCStats were used as well as some fighter level data such as reach and ranking to test for bias with these. A crowd effect was also included with a binary variable for a fighter competing in their home country along with the interaction of this term with an indicator of whether there was a live audience. All the metrics examined, aside from takedowns missed, had a positive effect which is expected. Bigger moves, such as knockdowns and submission attempts, also had a larger effect on scoring. The actions that seemed to have the largest disagreement between judges were head strikes missed and control time. This makes sense as a missed head strike could be viewed as either positive, negative, or neutral. Additionally, it is widely debated how much control time should be factored into scoring, so this discrepancy makes sense. There was also a statistically significant positive effect for the interaction of a home fighter with the crowd variable.

Gift (2021) did a study focusing on the scoring of the 10-8 round within MMA. New criteria for 10-8 rounds were implemented in 2017, and this paper studies the impact of that change. The criteria for a round to be a 10-8 was significantly loosened, and the formal definition was changed from a round that is “overwhelmingly dominant” to “winning by a large margin”. A difference in differences framework is used to examine the 2016-19 period, to allow for the possibility of different groups of judges implementing the new criteria differently. A logit model was created for this period, containing an indicator variable for the new criteria, an indicator variable for a specific group of judges (whatever the group of interest is being studied), and a vector containing 29 fight metrics to serve as a control. Another model was created for 2001-2019, containing only data in Nevada to test for changes in 10-8 scoring over this time. This model contained a categorical variable for 3 different time periods. The model found that judges in New York, Nevada, and California already began implementing the new 10-8 criteria in 2016, before it was officially adopted. This is likely due to the large number of events in these states causing them to be at the forefront of judging, so these judges were made familiar and trained with this new criterion before it became official. Because many judges travel frequently and work in different jurisdictions, “traveling judges” were also examined, defined as any NY/CA/NV judge who appeared in at least two of these three states in the data. With these judges, there was essentially no difference between 2016 and the 2017-19 time period, indicating that they had already fully implemented the new 10-8 criteria in 2016.

Classifying Winning Fighters & Performances

There have also been research articles focusing on winning fighters and performances instead of individual rounds. An article by Latyshev (2021) et al. aimed to find the characteristics most prevalent in successful mixed martial arts fighters. The top 13-15 fighters were selected from each weight division in the UFC (depending on available data). The weight classes of these fighters were then grouped together into three weight classes: small (70 kg and below), medium (70 to 85 kg) and heavy (over 85 kg). Four different groups of indicators are examined: age and anthropometric, rating, impact performance and defense, takedown performance and defense. The average age for a top fighter was found to be 31.8 years old, and lower weight classes had younger fighters on average. For top 15 fighters, fighters in the lightweight group had an average of 7.2 wins with 2.7 losses. This was 9.6 wins and 3.5 losses for the middle weight group and 8.7 wins and three losses for the highest weight group. This is likely due to the fact that there are more UFC fighters in the central weight classes than the extreme ones, so a fighter has to build up more of a reputation to make it to the top 15. These fighters finished 58.2% of fights, and fighters in the higher weight classes had both more finishes and more knockdowns. It was also found that higher weight fighters tend to more striking in the clinch or on the ground, whereas in the lower weight classes most of the striking is on the feet at distance.

Crossley (2015) wrote a paper with the goal of identifying specific key performance indicators prevalent in winning mixed martial arts fighters. 16 professional fights with 32 different fighters were examined. Analysts examined each of the fights in slow motion, and each action was classified by pre-specified attributes. Striking offense and defense were found to be noticeably different between winners and losers, with winners executing more offensive techniques and losers executing more defensive techniques. While not deemed statistically significant at the 0.05 level, these effects were very close with p-values of 0.054 and 0.057 respectively. Losing fighters were found to use their lead limbs significantly more often, with about 60% of their strikes coming from lead limbs as opposed to about 40% for winners. A lot of fighters like to start off using a lot of lead strikes to set up their rear hand, so when a fighter isn't able to find any success with their lead hand, it is very difficult to set up power shots. Positionally, it was found that losing fighters executed most of their shots on the feet in neutral control. Winners, on the other hand, executed a lot more of their strikes while in dominant positions.

Collier (2011) focused on finding what factors are most important for winning a fight. A binary response model is used, so all coefficients will be interpreted as the effect of X on the probability of winning. Fight data was obtained from fightmetric, and fight statistics were used as well as some fighter level data (age, weight, height). The differences between the two fighters for each statistic will be used in the model. Summary statistics were also calculated for two categories: fights that ended in a decision vs did not end in a decision. T-tests were run on these, and some variables were found to have statistically significant differences in these two. Differences in knockdowns and multiple submission attempts were found to be significantly smaller in decisions, which seems to make sense as these are actions that frequently result in finishes. The probit model had an r squared of 0.634, and knockdowns, stand-ups, power shots landed, various types of takedowns and submission attempts were found to be statistically significant. The age variables had almost no impact on the model, which was somewhat surprising. Knockdowns had the largest impact, which lines up with most other research modeling fights.

Finishing Fights

Bianca et al. (2020) analyzed MMA fights that are finished in different rounds to see how characteristics of these fights differ. This study used time motion analysis to examine how fighters performed throughout the duration of fights. 1,564 rounds from 678 UFC fights were examined for this study. 28% of the fights were finished in the first round, 13% during the second round and 59% of them were finished in the third round. Time-motion and technical variables were examined. Time motion variables are defined as either standing up or doing groundwork with high/low intensity and technical variables are specific actions such as strikes, takedown attempts and submission attempts. A significant main effect was found with ending rounds & standing combat with low intensity. Fights ending in the first round had less time standing with low intensity than fights ending in the second and third and fights ending in the second round had less of this than fights that ended in the third. More high intensity striking seems to lead to an increase in knockouts, which is expected. Fights ending in the first round also had a higher percentage of overall strikes landed (88%) as opposed to fights that ended in the second (54%) or third (64%) round.

The puncher's chance, a concept stemming from boxing, is the idea that a strongly outmatched fighter still has a chance to win a fight if they can land one or two clean, powerful punches. Wild (2022) wrote an article to quantify this effect within MMA, and tried to figure out how often outmatched fighters can win by finding that one shot (within MMA this could be a punch, kick, knee, or elbow etc.). Data was examined from UFC 12 to UFC 250, and only fights that ended in. KO/TKO were included. Significant strike percentages are used as a measure of whether a fighter was being outclassed. A logistic regression was run on wins with significant strike percentage, yielding an R^2 of 0.58. As expected, there was a very strong correlation between significant strike percentage and wins. This effect really leveled off around 45% (win percentage ~ 10%), and win percentage dropped significantly below this value. Setting X (significant strike percentage) equal to 0 in the model to 0 results in an estimated win percentage of 0.001481 (or about 1:674). This would mean that according to the model a fighter who is getting completely outclassed (landing "infinitely" less strikes) this fighter would have an estimated 1 in 674 chance of winning. However, because only bouts that ended in TKO/KO were examined (only 1818 out of 5437 bouts), this data must be rescaled to the full number of fights. After rescaling, this fighter being completely outclassed has about a 1 in 2018 chance of winning. These odds increase to about 1 in 43 with a fighter who is landing only 45% of their significant strikes.

Stellpflug et al. (2022) performed a descriptive analysis of the types of fight ending chokes throughout UFC history. All UFC events up to October of 2020 were examined in this study, and the number of submissions in each event was documented. Each choke individually was examined, and the type of choke as well as the hand used and if it resulted in a loss of consciousness was recorded. Of the 5834 fights in the data, there were 1186 fight ending submissions, with 904 of them being chokes. This means fight ending chokes consist of 15.5% of all fight outcomes and 76.2% of all grappling submissions. About half of these were rear naked chokes, and the 5 most observed chokes (rear-naked, guillotine (no arm), arm-in guillotine, triangle, arm triangle) consisted of 89.4% of all chokes. The next two very similar chokes, d'arce and anaconda, had 27 and 20 observations respectively (about 5.2% of chokes combined), and any other chokes below these were extremely rare with less than 10 observations. In terms of handedness, it has been almost identical throughout history as there have been 453 right arm/leg chokes as opposed to 451 with the left leg/arm. 99 of these chokes (11%) resulted in a loss of consciousness, but due to video interpretation this may range between 97 and 107.

Classifying Judging/Outcome in Other Combat Sports

Myers (2013) wrote an article focusing on the differences in Muay Thai scoring within the UK and Thailand, and see if this has an effect on fighting style. Muay Thai scoring in the UK is pretty similar to international boxing scoring, whereas the scoring in Thailand follows the traditional scoring system that has been used for a while. Notational analysis was used, and fights from 32 fighters (16 Thai and 16 UK) were examined. The fights were watched in real time and slow motion, and each individual technique from the winning fighter was classified by type and target. Defensive techniques were analyzed similarly. Whether the fighter was off balance was recorded as well as the result of the technique, and the effectiveness was scored on scale of 1-3 from no effect to highly effective. ANOVA was performed on each group of fighters, and it was found with the Thai fighters that there was no significant difference in their pattern of techniques. With the UK fighters, there was a statistically significant difference in technique selection, suggesting they used a more varied pattern of techniques. There was also a statistically difference between the two groups, indicating UK and Thai fighters used different techniques. Thai fighters used significantly less punches, and also targeted the leg much less with their kicks.

Defensively, the two groups were very different. Raising the leg to block was by far the most utilized technique by Thai fighters (48.37%), while this was only used 3.09% of the time for UK fighters. Instead, the primary defensive technique used by them was a conventional boxing cover (33.3%). The results of this analysis seem to indicate that the different groups of fighters have adapted their techniques to the judging. It is generally understood that Thai judges like to see damage, and value kicks very highly due to a traditional Muay Thai mentality. These findings here seem to back this up, and the fact that Thai fighters have a lower punch rate but have more techniques that were deemed very effective shows that they are likely focused on delivering power punches as opposed to volume.

The purpose of this study by Latham (2018) was to make determinations about the judging criteria of boxing and see the impact the judging criteria has on fights that result in a decision. A sample of 174 bouts from the 2016 Rio Olympic Games and the 2017 Hamburg world boxing championships were analyzed. Similarly to the previous study mentioned, each move was examined in slow motion and categorized by fighter, punch type, target and outcome. A 3x4 ANOVA test is used to test for differences for each variable, using round and outcome as between factors. Within unanimous decisions, winners landed more punches than losers 89% of the time. For split decisions, this number decreased to 61%. Significant differences were found for punches landed between winner and loser for each round. Unanimous winners were found to land more punches than split decision winners are only rounds 2 and 3. Both unanimous and split winners landed a higher percentage of very successful punches than unanimous and split losers.

When looking at percent of punches that whiffed unanimous losers threw more than unanimous winners in rounds 1 and 3; while this was only true in round 1 for split winner vs loser. Punch accuracy to the head was significantly greater for unanimous winners vs losers in all three rounds, and it was higher for split winners vs losers in rounds 1 and 3. In terms of specific techniques, it was found that the straight lead (jab), straight rear, lead hook, and lead uppercut were the most likely to be used by winners. It was also found that punches thrown and landed in general decreased in rounds 2 and 3, and the performance rating given for techniques in the study also decreased in these rounds. While not known for sure, it would appear this is most likely just due to fatigue.

El Ashker (2011) sampled 33 boxing matches to identify performances between winning and losing boxers. Video recordings for all these fights were examined and punches were classified as either straights, hooks, or uppercuts coming from the lead or rear hand and directed at the body or head. Combinations were calculated (2 or more subsequent punches), and defensive movements with the arms, legs or trunk were recorded. It was found that straight punches, particularly to the head were by far the most utilized technique. Winners attempted more punches in all three rounds of the fights, but especially in rounds 1 and 2. This seems to back up the idea of output being the key to winning fights, and the importance of coming out and establishing a high level of offense early. The use of lead hand punches, however, was significantly higher for the winner in the third round. It appears that the use of the lead hand to manage distance throughout the fight was a key factor of winning. Punch output dropped significantly for losers in the third round as well as execution of defensive skills. This seems to indicate that fatigue is playing a key role in losses.

Another similar study by Dunn et al. (2017) examined the same concept with the addition of behavioral variables. Behavioral variables included movements such as guard drops, step time and bounce time. Analysis showed a significant main effect of winners vs losers for technical variables. Specifically, winners were found to have a much higher hit percentage than losers. There was no significant effect between winners and losers for the behavioral variables. The only moderate effects found here was that winners may bounce more, step less and drop their guard for longer durations then losers. Technical actions were generally consistent throughout all three rounds, but behavioral variables changed more throughout the fight. Bounce time decreased from rounds 1 to 3 and the average guard drop time increased. This is expected, as being heavier on the feet and dropping their guard are traditionally looked at as signs of fatigue.

Lachlan (2017) added to this area with a goal of seeing how a non-liner approach compares to a linear approach of classifying boxing outcomes. Win/loss is used as the response variable in this study, so these models will not be scoring rounds but instead seeing what influences the overall result of a fight. Fights that did not make it to the third round are not included in the data. A linear model was created with 13 performance indicators, utilizing the difference between each fighter, and a decision tree was used as the non-linear approach. Variables for total & significant strikes both landed and attempted as well as offensive passes were found to have the largest impact on winning and losing in the linear model. The decision tree analysis successfully classified bout outcome at about 71.8%, and ground strikes landed primarily influenced the classification of the model. Takedown accuracy and significant strikes also contributed. A second decision tree model was run, this time using rate dependent attributes as opposed to the raw numbers and had a classification accuracy of about 76.3%. Major contributors in this model were significant strikes per minute, significant ground strikes per minute takedown accuracy and significant strike accuracy. The second seemed to be better due to the increased classification accuracy as well as better specificity. This makes sense as well, as the metrics that account for time elapsed should be better indicators of performance.

Southpaw Advantage

It has been found that there is an advantage with being left-handed in many interactive sports. It is theorized that this is due to right-handed athletes seeing much less of left handers in practice, while the opposite is true for left-handed athletes. Hagemann (2015) studied this supposed advantage with the use of tennis footage. Participants in the study were shown tennis clips from two players, two of which were left-handed and two of which were right-handed. 96 clips of them striking the ball were examined. These clips were viewed originally and mirrored (so the handedness would appear opposite) to avoid a potential issue of movements being genuinely different between the two groups. The participants were tasked with predicting the depth and direction of each of the strokes they were shown. ANOVA was performed on the data. Participants' tennis level was classified as expert, intermediate or novice. Experts performed better than intermediates and intermediates performed better than novices, which was to be expected. There was also a significant main effect of playing hand ($P < 0.01$), and the strokes were significantly harder to recognize by left handers in all three groups. Additionally, left handers and right handers were both better at predicting the direction of right-handed strokes. With the mirrored strokes, the participants did a slightly better job of identifying original right handers as opposed to the mirrored left handers despite looking almost the same.

This raises the question of whether this southpaw advantage translates to MMA & other combat sports. Loffing (2015) examined the fight records of 2,403 left and right hand-oriented boxers from 1924 to 2012 to study this supposed advantage. Annual boxing ratings from the ring were examined consisting of the top 10 or 11 fighters for each year. Expected frequency of being a southpaw (21.23%) was calculated from a survey studying lateral preference boxing & other sports. Performance metrics to be examined for these boxers consisted of their overall win-loss percentage as well as their knockout percentage. Boxers were ranked using these metrics, and to study if southpaws varied over time a 2x10 (handedness x decade) ANOVA was performed. Overall, left handers within the data were less frequent than the estimated population value. Frequencies of southpaws varied greatly between 3.28% and 33.15%, and a significant excess of southpaws was only found in 7 of the 89 years. However, the percentage of southpaws has increased greatly from early to recent years and all of the year's southpaws were overrepresented have been in the last 30 years. When examining fight records, southpaws were found to have significantly better win-loss ratios than orthodox fighters. While this study does not show that southpaws have overrepresented boxing throughout history, it does seem to indicate there is a chance southpaws have actually been underrepresented. This would still mean left handed fighters have had an advantage throughout history, but there was an inefficiency. The correction of this inefficiency could be a viable explanation for the increased rate of southpaws that are now entering the sport.

Sorokowski et al. (2014) examined this advantage in boxing using fighter records. For this study, data was collected from boxrec for the top 20 fighters in each of the 17 weight classes (340 fighters). 25% of these fighters were southpaws and 75% were orthodox. Stance did not have an impact on winning percentages, and the winning percentage for southpaws (87%) was almost identical to that of orthodox fighters (88%). While the fighter being examined's stance did not have an impact, their opponent's stance did have an impact on winning percentage. These fighters were found to have a higher winning percentage against orthodox fighters (89%) than southpaws (85%) with a p-value of 0.002. For an additional study, the top 150 boxers in each weight class were examined. 20% of these fighters were southpaws, and it was found the proportion of southpaws in this sample was higher among the better rated boxers ($p < 0.01$). While this data did not show that the best boxers in the world are more likely to be southpaws, it did seem to show a southpaw advantage. There is likely not enough data here for this effect to be shown, but these fighters did fare better when going against orthodox fighters and the larger study showed more overall success from southpaws when looking at more boxers.

Because of the shown advantage that being a southpaw can have in boxing as well as other sports, one would expect this to carry over to MMA as well. Schorer (2013) examined 1468 UFC fights to examine this. All data was obtained from fightmetric. 80.3% of these fighters were orthodox while 17.4% were southpaws. Fighters who were considered switch stance were not included due to there not being enough data. Both winning percentage and number of fights were examined in this analysis. This is better than just examining winning percentage as fighters who have a high winning percentage with more fights have had a longer, more successful career. There was a significant difference between stance and number of fights ($p = 0.02$), with southpaws having just over two more fights on average. The t-test between winning percentage and stance indicated no differenced, but southpaws had the slightly higher winning percentage within the sample. The combination of having more southpaws in the sample (17%) than the overall population (about 11%) and southpaws having more fights on average does seem to indicate that there is a southpaw advantage within MMA. However, we cannot definitively relate a southpaw advantage to a left-handed advantage as there can be right-handed fighters who chose to fight southpaw. It is still unclear whether the inherent advantage of being a southpaw is leading to more natural left handers doing well or the perceived advantage of being a southpaw is causing more fighters to adapt the stance.

Home Advantage

Home field advantage, while still feasible in MMA, is a very different idea than most sports. Unlike a sport such as football where fans can try to weaponize their voices while the opponents are on offense, it is difficult to see this being the case in combat sports. It does, however, seem very possible that a crowd's reaction to a fight could have an impact on how judges see a fight.

Warnick (2007) wrote an article where they examined this concept in. boxing. 522 European boxing fights that resulted in a decision from 2013 to 2016 were examined. Additionally, only right-handed fighters were included to avoid a potential southpaw advantage having an impact. Variables examined in the study included weight, quality, age, and height of boxers. Quality scores were obtained from BoxRec and ranged from -721 to 721. These are calculated by starting with a fighter ranking model, and fighters are given points for a win or deducted points for a loss based on quality of opponent. For home advantage, a fighter was considered to have home advantage if only they shared a place of residence or birth with bout location. If this was the case for both or neither fighter, the bout is considered neutral. Classification trees were used, and it was found bout outcome primarily depended on boxers' quality and secondarily their age (older boxers were more successful) as well as height (taller were more successful). It was found that local boxers were higher quality, taller and younger than visiting opponents. The quality and height is advantageous for the home boxer, but them being younger on average is disadvantageous. The actual impact of being at home on winning was somewhat unclear due to this.

Balmer et al. (2005) also examined if there is home advantage in European boxing. Data was examined for all European championship bouts from March 1910 to June 2002 from BoxRec. Home competitors were defined as only those whose nationality matched the location of the fight; they did not consider place of residence. Any bouts that resulted in draws or did not have a single home fighter were removed. A binary logistic regression was used, and outcome type (decision vs stoppage) as well as relative quality (measured by their records) were used. The relative quality is important, as it was found that home boxers had a significantly higher winning percentage than away boxers. Four stages of analysis are used, beginning by just examining the influence of outcome type on a home win, followed by adding relative quality to the model, and then splitting the data into different time period. Finally, outcome type technical knockout is added for recent fights. It was found that a home win was significantly more likely if fights ended in a decision, 0.76, and only 0.67 if the fight ended in a knockout. Relative quality ended up being a significant predictor of the victor, and the influence of outcome type varied with relative quality. In the time period analysis, it was found that relative quality did not have much of an impact on the model in earlier years, and the likelihood of a home fighter winning a decision increased in more recent data. Overall, the probability of a home win was found to be higher in fights that went to a decision, and this was most evident when away boxers had better previous records than home boxers.

Myers (2006) et al. looked for national bias in the scoring of Muay Thai. Data was examined from the 2003 IFMA Muay Thai championships. 70 class A Muay Thai bouts are examined, each of which judged by between five and nine judges. Judges were classified as either red, blue, or neutral, based on whether they shared a nationality with either corner. A multilevel model was fitted with judges as the response variable. Scores were nested within bouts, and bout is included as a random effect due to judges scores likely being much more similar within bouts than between them. The model yielded an intercept of -0.3, showing that on average judges who shared nationality with the blue corner scored them about a third of a round higher than the fighter out of the red corner. It was also found that neutral judges on average scored red corner boxers 0.64 higher than judges who shared a nationality with the blue corner. While this difference in scoring is evident, this does not necessarily mean it is all that this has a large impact on the outcome. When removing non-neutral judges, only two out of the 43 bouts examined had a change in result. However, some of this was due to nationalistic bias both ways balancing itself out, meaning the potential for impact of national bias may be higher.

After examining the impact of just nationality, Myers & Balmer (2012) conducted an experiment studying the impact of crowd noise on Muay Thai judging. The study involved 30 qualified Muay Thai judges and had them score a Muay Thai bout in either a crowd noise or no crowd noise (with noise cancelling headphones) condition. Data was only included for bouts that went to a decision and had a hometown fighter against an out of town fighter, or there was a clear hometown advantage. 35 fights were examined, and there were two judges scoring each with crowd noise and 2 scoring with no crowd noise. The response variable is the difference between home score and away score, with these values ranging from -4 to 6, and a Markov chain model is used. Crowd noise did have a statistically significant impact on judging, and the model showed that exposing judges to crowd noise resulted in a difference of 0.53 extra points to the home fighter (about a half a round). In this dataset, 26.7% of the fights could have been impacted by this point difference based on how close they were, and this number was 29.6% when looking at a larger Muay Thai dataset.

Other Research

This paper by Herbert (2002) attempts to study boxing decisions by looking at the interrater agreement between the judges. The goal is to use this analysis to examine the Evander Holyfield vs Lennox Lewis JR decision. Both the official judges' scorecards as well as media scorecards recorded are examined, and these are all used to form agreement matrices. The judges' scorecards are essentially just being compared to each other to see if they stand out. Each judges score for each round is modeled as a Bernoulli, and a corresponding probability of scoring that round for a fighter is estimated by the percent of judges who scored the round for that fighter. The probability distributions of how the fight can be scored are examined to find extreme scorecards. When applying this analysis to the Holyfield fight, it is found that two of the judges' scorecards seemed to be extreme outliers. It was found that the probability of scoring at least as many rounds for Holyfield as the two judges who did not have him losing (seven) is 0.0256, and the estimated probability of Holyfield winning at least six rounds (enough to draw or win) is 0.132. While nothing is drawn here from the fight data itself, the analysis does seem to show that there was some bias (or at least statistically significant difference vs general opinion) in the scoring of the Lewis Holyfield fight.

This study by Hoelbling et al. (2023) was focused on the effectiveness JudgED, which is a game designed to evaluate and train martial arts judges. The performance data for the judges is obtained with a procedure that compares in-game inputs to expert defined decisions. After completing a session of training, judges are given a statistical performance summary which includes decision accuracy and reaction time of decisions. This JudgED system has not yet been evaluated, and the purpose of this study was to evaluate it. The study contained 2 video-based tests within the game, and 16 different judges were analyzed. The judges were given real kickboxing footage (2021 WAKO world championships) to watch and were given one day to become familiar with the system and data was collected on the next day.

The experiment found a mean decision accuracy of about 43% for the judges, and this varied greatly by discipline. Light contact or especially point fighting contests (tatami disciplines) had a significantly higher decision accuracy than full contact competitions (ring disciplines). This is expected, as there is a lot more to evaluate in full contact contests whereas point fighting is essentially identifying lands. Fleiss' kappa tests were run to test for inter judge agreement. For ring disciplines, this resulted in a kappa value of 0.371, indicating a significant amount of agreement between the judges. The same was true for decision accuracy; judges had a higher decision accuracy within tatami disciplines. When looking at the judgement difficulty of each scene in the study (as predetermined by an expert referee), the decision accuracy is much lower in the scenes rated more difficult to score. This backs up the validity of this training system, as we should expect to see accuracy decrease as people are being trained in difficult, faster paced scenarios.

Martens (2023) wrote an article with the goal of predicting the winner of mixed martial arts bouts; but he did this using very unconventional data. This study focused purely on nonverbal displays of pride from the fighters. For the first study, 158 undergraduate psychology students participated. They were each shown a clip of a fight where no clear winner was present, then presented with a picture of each boxer. One of the boxers was clearly displaying pride while one was in a neutral stance. The participant was asked who they thought won, and to what extent on a scale. They were also asked who they would rather train to fight with and who they would rather help them study for an exam. They were then asked how they thought the screenshots of the fighters they were shown showed pride and other emotion. T-tests showed that the proud fighters were rated significantly higher on displaying pride and emotion, so they were clearly perceived as the study designed. A one sample t-test was run on who the participants selected as the winner, resulting in a significant t-value of 3.81 ($p < 0.001$). Overall, 65% of the participants selected the fighter showing pride.

Study 1 supported the hypothesis that showing pride will increase chances of victory, but only in an experimental setting. The second study examined this effect in a real boxing environment. The nonverbal behavior of fighters was assessed in 252 boxing bouts judged by real judges. Research assistants went through these fights and rated each fighter's level of pride on a scale of 1-7. A t-test was run, and a resulting t-statistic of 18.97 ($p < 0.001$) indicated those who showed the most pride generally did win the fight. ANOVA was also run, and a significant interaction emerged when it came to decision type. The level of pride shows by the winner seemed to be the same in a split and unanimous decision, but the level of pride shown by the loser significantly increased in a split decision. These results show that a display of pride could very well influence a judge's decision of who won a fight. However, this is far from a conclusion as this could just be due to these fighters performing better, and displaying pride was a symptom of that instead of having a causal effect on victory.

Data & Methodology

The data for this analysis was gathered from 2 sources: ufcstats.com and mmadecisions.com. MMADecisions has the specific scorecard data for most UFC fights throughout history. This data dates back to UFC 80 with some missing fights but has pretty much every fight reliably after about UFC 120 or so. UFCStats has the fight statistics (as well as round by round statistics) for almost all UFC fights in history (at least reliably back to about UFC 30). The main statistics from this data that will be useful for this analysis are significant strikes landed, total strikes landed, takedowns landed, submission attempts, reversals and control time. Additionally, there are some further breakdowns of significant strikes in this data that can be used. There is data for significant strikes split by whether they were landed to the opponents head, body or legs. Similarly, there is another split for the significant strike data that splits these strikes whether they were landed at distance, in the clinch, or on the ground.

To begin this analysis of UFC judging, I will examine summary statistics for some of these key variables. I will examine the summary statistics for these variables among round winners as well as among the round losers to try to identify initial differences.

Round Winners

Variable	Total	Mean	SD	Min	Med	Max
Total Strikes	229,453	32	16	0	29	141
Knockdowns	453	0.062	0.27	0	0	4
Takedowns	4,771	0.66	0.94	0	0	9
Reversals	411	0.056	0.24	0	0	2
Sub. Attempts	859	0.12	0.42	0	0	5
Ctrl. Time (min)	9,956	1.4	1.4	0	0.82	5

Round Losers

Variable	Total	Mean	SD	Min	Med	Max
Total Strikes	147,144	20	12	0	18	130
Knockdowns	59	0.0081	0.093	0	0	2
Takedowns	1,789	0.25	0.55	0	0	5
Reversals	375	0.051	0.24	0	0	3
Sub. Attempts	427	0.059	0.27	0	0	4
Ctrl. Time (min)	3,675	0.5	0.78	0	0.12	4.8

As expected, the differences for all these variables between winners and losers is evident here. Winners seem to land around 50% more strikes than losers, which is not very surprising at all. I was a little surprised to see the minimum total strikes landed for round winners was 0. After looking at the data, I found that there was only one round where this happened: round 1 of Raul Rosas Jr. vs Christian Rodriguez. Rosas was able to take Rodriguez down 3 times and obtain just over 4 minutes of control time which won him the round, but he did not land or even attempt a single strike while doing so.

The relative difference in takedowns is a bit larger than striking, as round winners have landed about 3 times as many takedowns as the losers. Control time tells a pretty similar story as takedowns, and based on how control time will occur after a takedown/attempting to get one multicollinearity is certainly something to look out for with these variables. Reversals and submission attempts occur relatively infrequently. Round winners have went for about twice as many submission attempts as round losers, whereas the reversal numbers are pretty close between these groups.

The difference in knockdowns between winners and losers could not be more clear. Knockdowns are one of the most observable forms of damage in a fight, so landing a knockdown should be massive for scoring. You can see in the summary statistics that the winners of rounds are landing knockdowns at almost ten times the rate of losers. Looking a little deeper into this data, there were 453 total knockdowns landed by round winners. There were 375 rounds where the winner had one knockdown, 29 rounds where they had 2 knockdowns, 4 rounds where they had 3 knockdowns and 2 rounds with 4 knockdowns. There were only 59 total knockdowns landed by losers; 55 rounds where the loser had 1 knockdown and just 2 rounds in all the data where the loser had 2 knockdowns. As you can see, landing a knockdown is massive in terms of winning a round on the scorecards and once two knockdowns have been landed a fighter will almost never lose that round.

Round Winners Significant Strikes

Variable	Total	Mean	SD	Q1	Med	Q3	Max
Head	99,954	14	10	7	12	19	89
Body	32,769	4.5	4.1	2	3	6	45
Leg	24,982	3.4	3.8	1	2	5	38
	157,705	22	13	12	20	29	141

Round Losers Significant Strikes

Variable	Total	Mean	SD	Q1	Med	Q3	Max
Head	62,316	8.6	7.3	3	7	12	81
Body	24,173	3.3	3.2	1	2	5	25
Leg	19,648	2.7	3.2	0	2	4	34
	106,137	15	10	7	13	20	92

The tables above show the breakdown of significant strikes by target. As expected, winners are landing noticeably more significant strikes to each area of the body. This difference is the largest for head strikes (60%) followed by body (36%) then leg strikes (27%). Many people do believe that strikes to the head are ultimately the most damaging, so this is not too surprising. When modeling with this data we can expect head strikes to have the most value.

Round Winners Significant Strikes

Variable	Total	Mean	SD	Q1	Med	Q3	Max
Distance	118,874	16	12	6	14	23	139
Clinch	19,798	2.7	4	0	1	4	45
Ground	19,033	2.6	5	0	0	3	66
	157,705	22	13	12	20	29	141

Round Losers Significant Strikes

Variable	Total	Mean	SD	Q1	Med	Q3	Max
Distance	89,962	12	9.6	5	10	17	92
Clinch	12,702	1.7	2.7	0	1	2	27
Ground	3,473	0.48	1.5	0	0	0	29
	106,137	15	10	7	13	20	92

These tables show the alternative breakdown of significant strikes: separated by where in the fight they were landed. The difference in distance strikes is the smallest: round winners are landing about 32% more strikes at distance. They are landing 56% more strikes in the clinch, and the largest difference can be seen with significant strikes on the ground: round winners are landing over five times as many of these. What is considered a significant strike can be somewhat strict (but remember, ALL strikes at distance are significant), so it seems like the impactful strikes really are being separated out with these significant strikes on the ground.

Modeling

Binomial GLM Model

The first model I create will be a binomial regression model that uses many of the statistics to predict the winner of a round. In order to do this, the difference between any statistics used between the two fighters will be calculated. Initially, the data being used has all of the fight winners in one column (NOT the round winners) and the losers of the fight in another column. If models are created using this data, there will be some inherent bias in the model for being the winner of the fight. In order to correct this, I have randomly assigned each fighter in a round to one of the two fighter columns and all corresponding statistics as well as the dummy variable for whether that fighter won the round will also be moved to the data columns for that fighter. The differences for any statistics used between fighter 1 and fighter 2 are then calculated. With this setup, this means that the exactly what GLM model will do is predict the chances of one of the two fighters (randomly selected) winning that round based on the differences in fight statistics. The first model created will look at differences between the following statistics: significant strikes landed, non-significant strikes landed, knockdowns landed, takedowns landed, reversals, control time & submission attempts.

There have been many changes made to judging criteria over the years. A fight in 2024 will be scored extremely differently than a fight from the early 2000s, so trying to model these fights together is likely not a good approach. One of the potential uses for these models I create can be to try to measure these differences of judging over time, but for now I will just focus on using the right data to create the best scoring model. Aside from the changes to judging criteria in 2017 (which primarily revolved around the scoring of 10-8 rounds which does not factor into this model), the most recent large changes to judging criteria were in 2012. For now, I will use all UFC events from 2013 to present to model judging. Here are the results from the first model using the predictors listed above:

	Dependent variable:
	model_winner
sig_dif	0.054***
	(0.002)
nonsig_dif	0.031***
	(0.003)
kd_dif	2.067***
	(0.171)
td_dif	0.390***
	(0.038)
rev_dif	0.480***
	(0.107)
ctrl_dif	0.449***
	(0.025)
sub_dif	0.781***
	(0.077)
Constant	-0.033
	(0.030)
Observations	7,282
Log Likelihood	-3,332.412
Akaike Inf. Crit.	6,680.825
Note:	*p<0.1; **p<0.05; ***p<0.01

The coefficients for this model generally align with what is understood about scoring. Knockdowns are by far the most impactful predictor, and landing a knockdown will increase the log odds of winning a round by about 2.2. As we saw earlier, these are reasonably uncommon so will not frequently occur but when they do they will significantly increase the winning percentage. As expected, the coefficient for significant strikes landed is larger than the coefficient for non-significant strikes. However, I expected this difference to be larger. Most non-significant strikes are very small punches on the ground/in the clinch that don't do any noticeable damage. Because of this, I would expect a significant strike landed to be at least twice as impactful but this is not what the model shows.

Variable	VIF
sig_dif	1.05
nonsig_dif	1.02
kd_dif	1.02
td_dif	1.54
rev_dif	1.07
ctrl_dif	1.46
sub_dif	1.03

The grappling coefficients (takedowns, reversals, control time & submissions) generally line up with what I would expect. Multicollinearity is obviously a worry here, but the VIF numbers do not indicate that this is an issue, and all coefficients are significant and make sense. Takedowns and control time's coefficients' sound right given that one takedown will increase the log odds of winning the same as about 52 seconds of control time. The coefficient for reversals is a bit larger than takedowns, which many would argue aligns with scoring criteria. Both moves would fall under the scoring category of effective grappling. With a takedown, one fighter has gone from a neutral position to being in a dominant position. With a reversal, on the other hand, a fighter has gone from being in a bad position to a dominant position, so they have improved their position moreso than with a takedown. The final grappling coefficient, submission attempts, has a coefficient around twice as large as takedowns and reversals. Submission attempts are a bit weird to interpret (as by definition a fighter has failed to achieve a submission with a submission attempt), but according to the results of this model these attempts seem to be very helpful towards effective grappling.

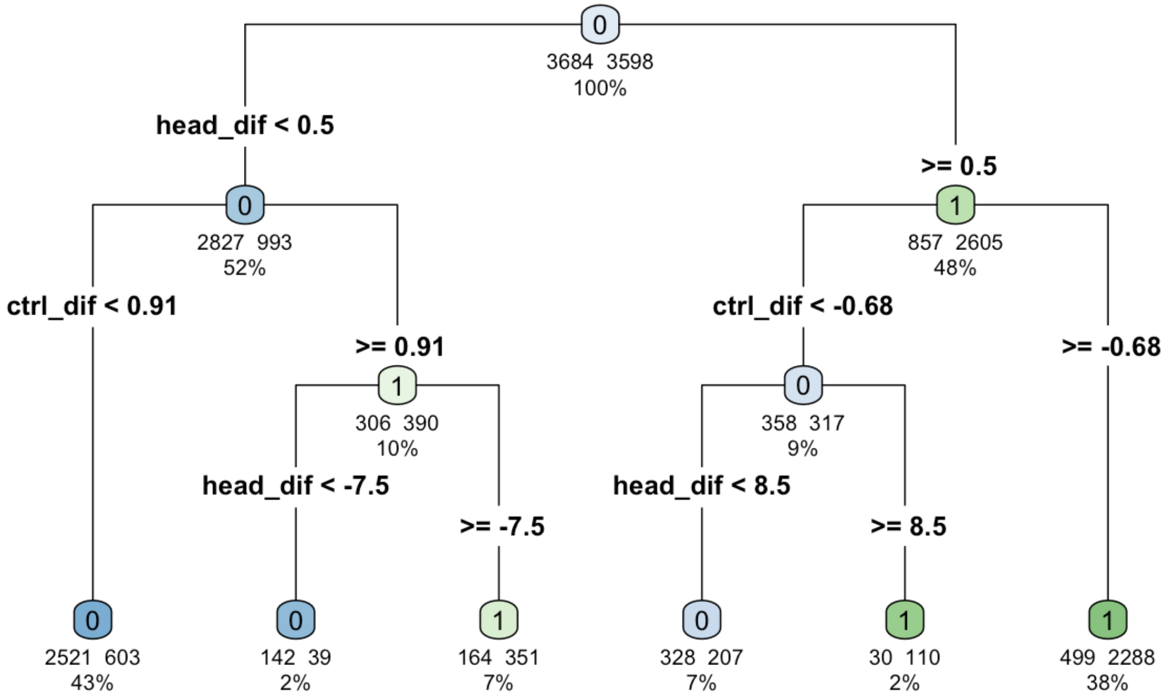
The next model examined will be very similar to the previous one, but significant strikes will now be split into significant strikes landed to the head, body & legs. The results for that model are as follows:

	Dependent variable:
	model_winner
head_dif	0.188***
	(0.006)
body_dif	0.141***
	(0.008)
leg_dif	0.115***
	(0.008)
nonsig_dif	0.028***
	(0.003)
kd_dif	1.967***
	(0.197)
td_dif	0.392***
	(0.042)
rev_dif	0.394***
	(0.118)
ctrl_dif	0.479***
	(0.029)
sub_dif	0.917***
	(0.087)
Constant	-0.050
	(0.034)
Observations	7,282
Log Likelihood	-2,711.757
Akaike Inf. Crit.	5,443.514
Note:	*p<0.1; **p<0.05; ***p<0.01

This model is certainly an improvement over the previous one. The striking coefficients seem to make more sense with how fights are scored. Non-significant strikes are now worth much less, and one significant strike to the head will improve the log-odds of winning by about as much as 7 non-significant strikes. The coefficient for strikes to the body drops by a bit, with another similar drop in the coefficient for leg strikes. Obviously any one of these strikes can be extremely effective, but this ordering of head > body > legs makes sense in terms of how judges might view fights. It is also important to consider that the use of leg kicks in MMA has changed a lot since 2012, so maybe this coefficient would change if more recent data were used

Decision Tree & Random Forests

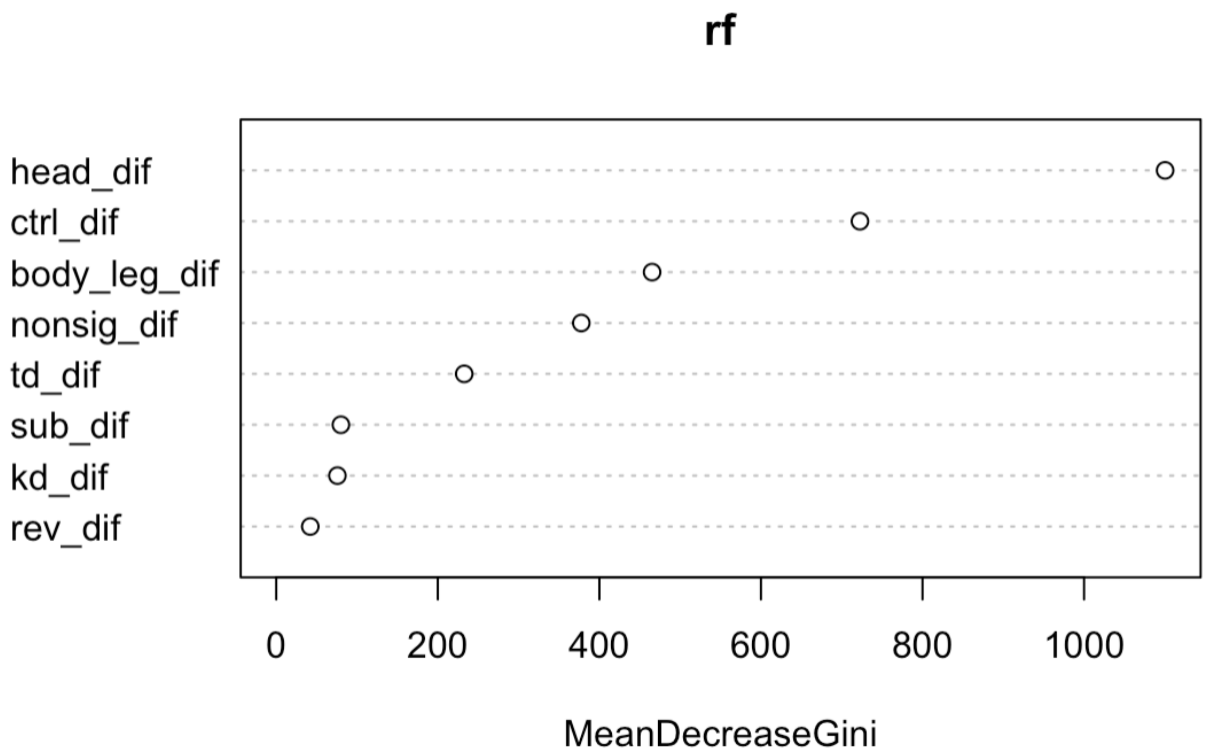
Now that I have created GLM models to model UFC judging, I will now do the same using a decision tree approach. The first tree examined will be similar to the last regression model. However, to help simplify the model and make splitting easier, I will consolidate the body & leg strikes difference into one variable which is the total difference in body and leg strikes.



The decision tree provides a very different approach to a scoring model. Rather than just placing a value on different fight actions, the model can now focus more on interactions and split into different types of fights (i.e a fight where one fighter won on the feet but lost the grappling). The left half of the tree is where the selected fighter lost or tied the head striking battle, and the right side is where this fighter out struck their opponent to the head. If the selected fighter lost or tied the head striking battle and did not have at least 55 more seconds of control time than their opponent, the tree predicts they lost. If they had less than this much more control time, the tree predicts they lost only if they were out struck to the head by more than 7 significant strikes.

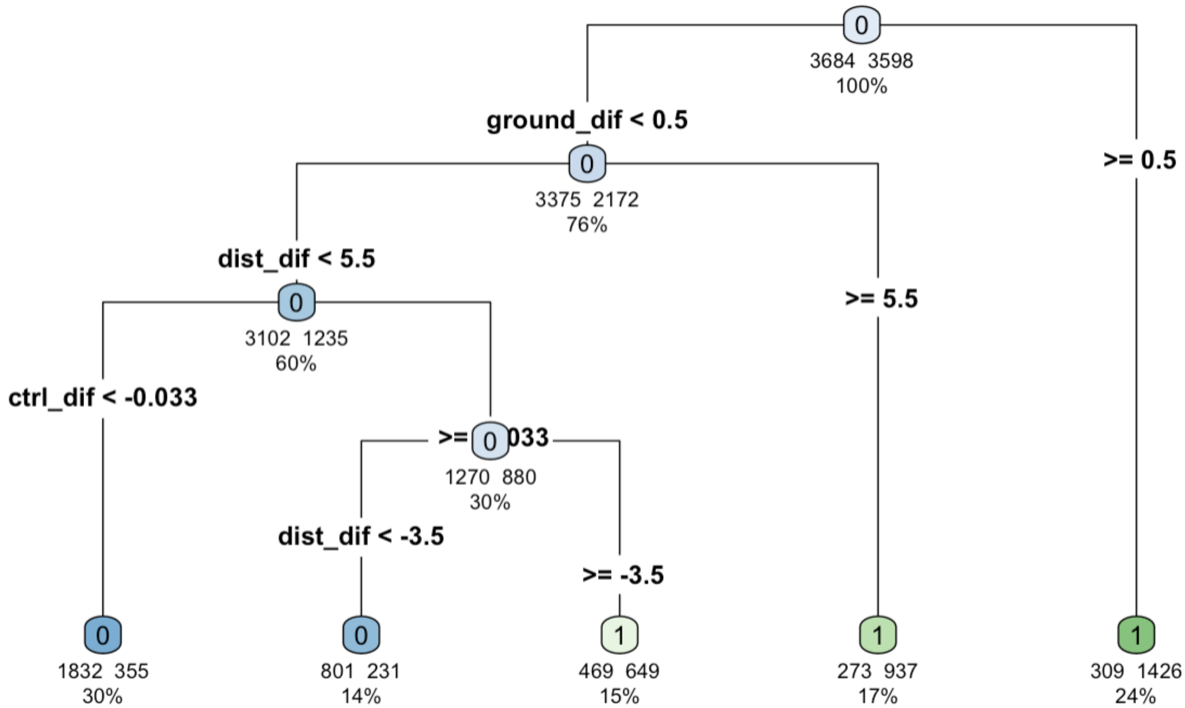
On the other side of the tree (where the selected fighter out struck their opponent to the head) a win is predicted if this fighter did not lose the control time battle by more than 40 seconds. If they were out controlled by this much, the tree only predicts they won if they outstruck their opponent to the head by more than 8 strikes.

The decision tree backs up the importance of strikes to the head. It also seems to indicate control time is the best measure of who won the grappling, and you can see it has completely simplified the scoring of a fight to control time and head strikes. The predictions are most accurate at both ends of the tree, where one fighter has won both the head striking and the control time. While overfitted, this tree approach seems like a reasonable way to score a fight that accounts for both striking and grappling. I will now create a random forest model using the same predictors. The following variable importance plot is the result of this random forest:



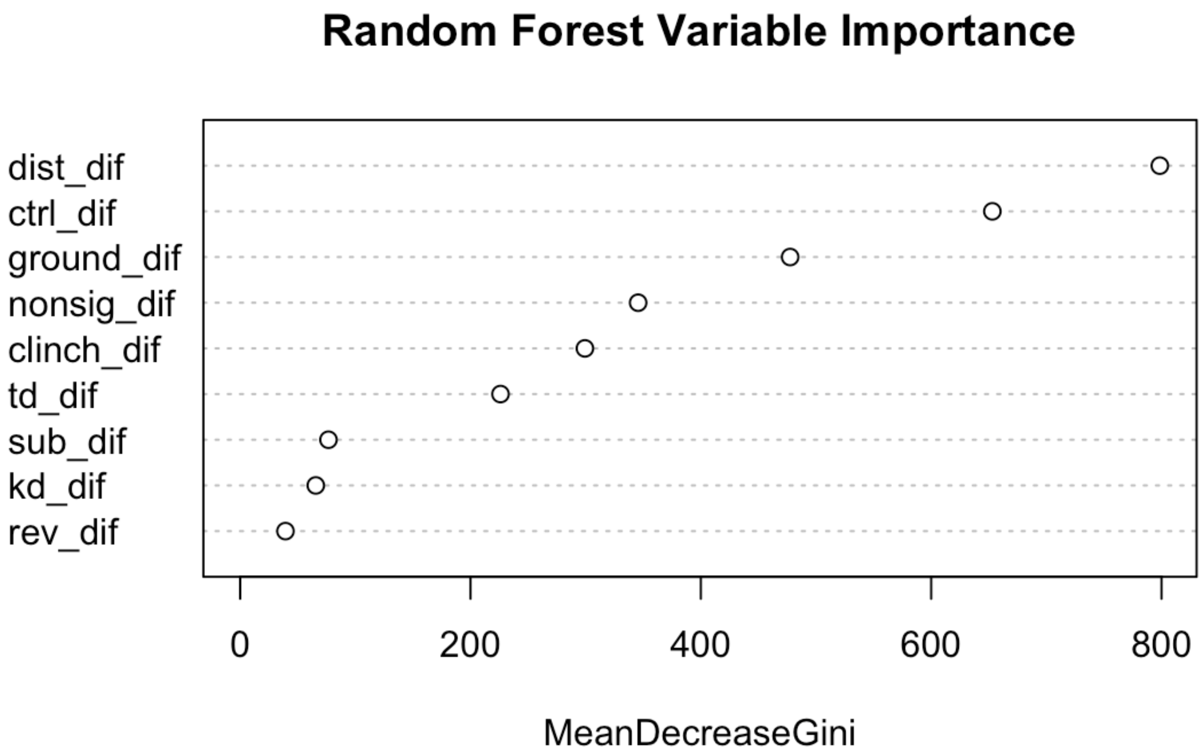
As seen from the decision tree, head strikes and control time are the most important predictors in this model. Significant strikes to the body & legs are the next most important. Takedowns are not as important as in the GLM models (likely due to most splits instead occurring on control time). Knockdowns, submission attempts & reversals have very little impact on the model. These do not occur very frequently, so it is unlikely any of the trees will split based on these. There is certainly a large advantage in this model that it can pick up on the interactions between the variables, but the linear models seem to do a better job of valuing these important events that occur less frequently.

The next decision tree & random forest examined will be for the other split of significant strikes: ground, clinch & ground strikes. The results for a GLM model using these were not great, but I think decision trees will pick up on the interaction of these variables much better.



The right side of this tree is very simple: if the selected fighter out struck their opponent with significant strikes on the ground the tree predicts a win. This is accurate 82.2% of the time. On the left side of the tree (where the ground striking battle was even or lost), the tree predicts a win if the selected fighter out struck their opponent by more than 5 strikes at distance. If this was not the case, the tree splits based on who won control time. A loss is predicted if the selected fighter was out controlled by more than 1 second. If they won this control time battle, the tree only predicts a loss if this fighter was out struck by more than 3 strikes at distance.

The interactions for this tree look even better than the last one. In terms of profiling fights into different types, it seems the tree can do this much better with the striking split up by area. In the previous tree with head & body + leg strikes, the tree was not able to split off where these strikes occurred (so these head strikes may have been on the feet or on the ground). While this tree can no longer split based off where they were targeted, the fact that it can split based on who won the striking purely at distance on the feet is highly valuable. I will now create a random forest model using these same variables.



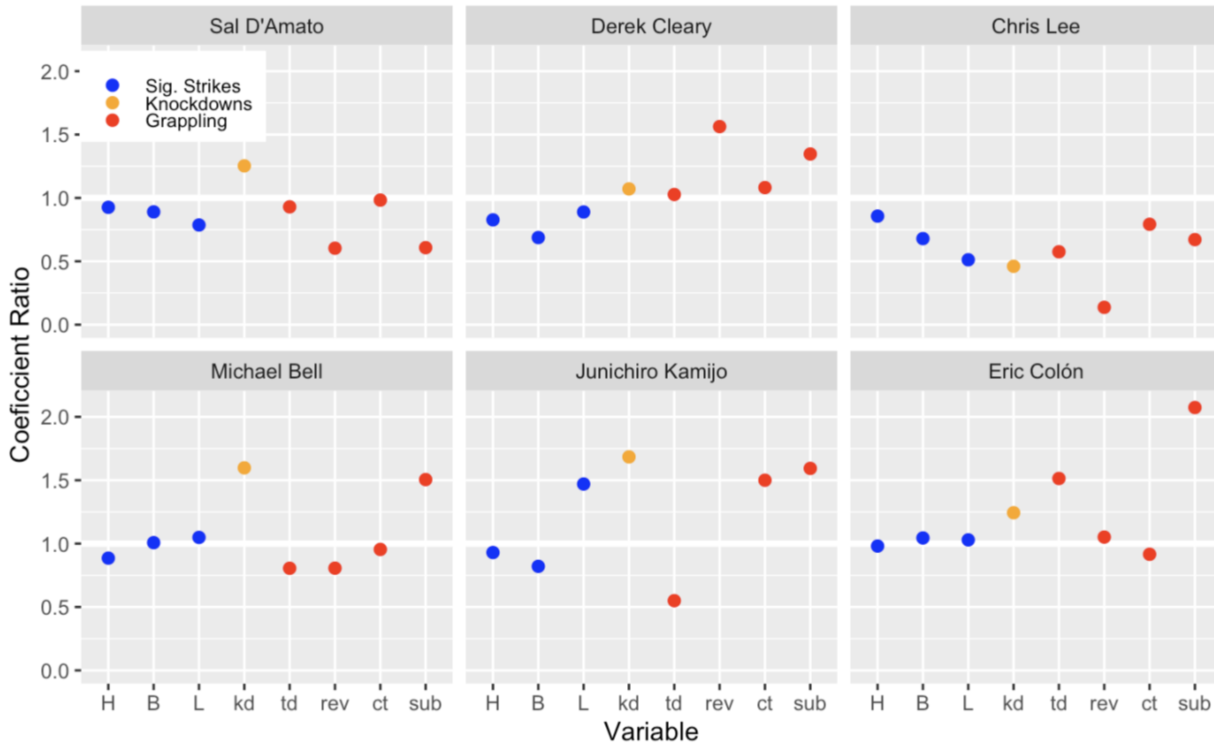
This random forest model appears to be an improvement from the previous one. The error rate has decreased just slightly (from 17.85% to 17.36%) but the predictors themselves look better in this model. The previous random forest put an extremely large emphasis on head strikes (about twice as impactful as body & leg strikes and by far the most important predictor)

Judge Profiles

As I mentioned earlier, one of the uses for these models can be applying them to individual judges to identify differences in how they score fights. I will now do this, and the process for each judge will be as follows:

1. An individual judge to examine is picked – we will call this Judge A.
2. Dataset will be split into two partitions: one dataset containing all fights judged by Judge A (judge data) & a dataset of all fights Judge A did not judge (non-judge data)
3. The judge data will be duplicated and two versions are created: In the first dataset the winner (response variable) will be whoever the Judge A had winning that round. This will be the dataset used to create the judge model.
4. In the duplicate judge dataset, the winner variable will be whoever the other two judges had winning the round (leaving out the judge A's score)
    a. There will be rounds where these two judges disagree on the winner. This is inherently a tie here, so these fights will not be included
5. The duplicate judge dataset will be combined with the original non-judge dataset. This combined data will be used to create the non-judge model for each judge.

This process can be used with both models (the GLM model & random forest). Due to the easy interpretation of the GLM coefficients, I will primarily be using the coefficient ratios between the judge model & non-judge model for each judge to see how they score fights. The following graph shows the coefficient ratios for each of the judges being examined:



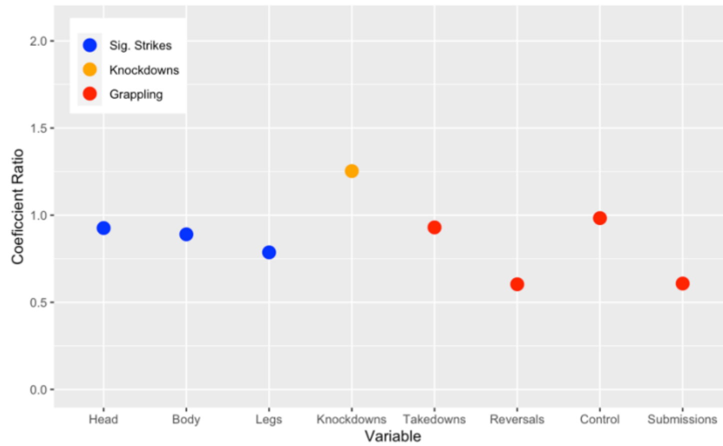
**Junichiro Kamijo's rev coefficient ratio is -0.37 and not displayed on the graph

Judges and their striker vs. grappler preference (Sum of striking coefficient ratios/Sum of all coefficient ratios):

Judge	STR %	Rounds
Chris Lee	60.4	1323
Junichiro Kamijo	57.8	1033
Sal D'Amato	52.5	2305
Michael Bell	50.6	1179
Eric Colon	43.1	790
Derek Cleary	39.8	1720

Judge 1: Sal D'Amato (2,305 rounds judged)

	Dependent variable:	
	model_winner	
	Non-Judge	Sal D'Amato
	(1)	(2)
head_dif	0.198***	0.184***
	(0.006)	(0.010)
body_dif	0.147***	0.131***
	(0.009)	(0.014)
leg_dif	0.121***	0.096***
	(0.008)	(0.013)
nonsig_dif	0.028***	0.024***
	(0.003)	(0.006)
kd_dif	2.087***	2.615***
	(0.208)	(0.404)
td_dif	0.410***	0.382***
	(0.045)	(0.079)
rev_dif	0.469***	0.283
	(0.123)	(0.214)
ctrl_dif	0.503***	0.494***
	(0.031)	(0.052)
sub_dif	1.010***	0.614***
	(0.091)	(0.159)
Constant	-0.028	-0.059
	(0.036)	(0.061)
Observations	6,948	2,285
Log Likelihood	-2,453.502	-843.112
Akaike Inf. Crit.	4,927.004	1,706.223
Note:		*p<0.1; **p<0.05; ***p<0.01

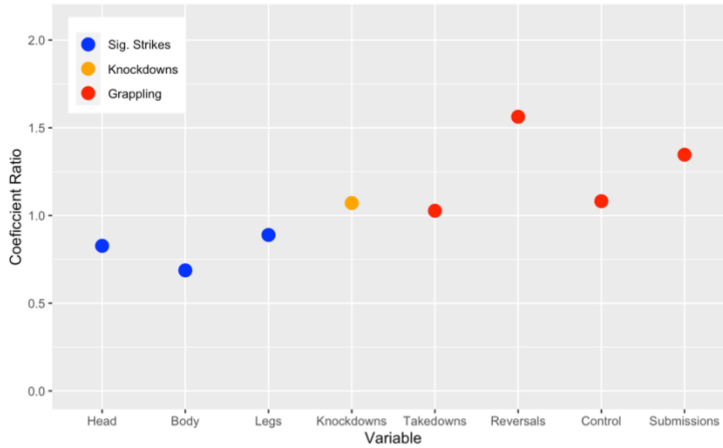


Sal D'Amato appears reasonably neutral, and he is scored 52.5% striker friendly. His coefficient ratios go in descending order from head to body to leg strikes so he appears to slightly prefer strikes higher up on the body relative to other judges. The striking coefficient that really stands out is knockdowns; D'Amato appears to prefer these more than the other judges.

While he does not appear to lean very strongly one way in terms of striking vs. grappling, D'Amato does appear to have some specific grappling preferences. His coefficients for both reversals (not significant) and submissions are much smaller than in the non-judge model, whereas the coefficients for takedowns and control time are around the same as the non-judge model. If the model results are assumed to be correct, D'Amato prefers traditional wrestlers over grapplers who favor jiu jitsu. Instead of focusing on what offense a fighter can generate on the ground, he seems to be more concerned with which fighter is able to secure a takedown and keep their opponent there.

Judge 2: Derek Cleary (1720 rounds judged)

	Dependent variable:	
	model_winner	
	Non-Judge	Chris Lee
	(1)	(2)
head_dif	0.195***	0.167***
	(0.006)	(0.013)
body_dif	0.149***	0.101***
	(0.009)	(0.018)
leg_dif	0.124***	0.063***
	(0.008)	(0.015)
nonsig_dif	0.028***	0.036***
	(0.003)	(0.007)
kd_dif	2.092***	0.964**
	(0.207)	(0.410)
td_dif	0.426***	0.245***
	(0.044)	(0.089)
rev_dif	0.469***	0.064
	(0.123)	(0.261)
ctrl_dif	0.495***	0.392***
	(0.030)	(0.058)
sub_dif	0.955***	0.641***
	(0.091)	(0.170)
Constant	-0.050	0.050
	(0.035)	(0.075)
Observations	7,083	1,319
Log Likelihood	-2,534.372	-548.093
Akaike Inf. Crit.	5,088.744	1,116.186
Note:	*p<0.1; **p<0.05; ***p<0.01	

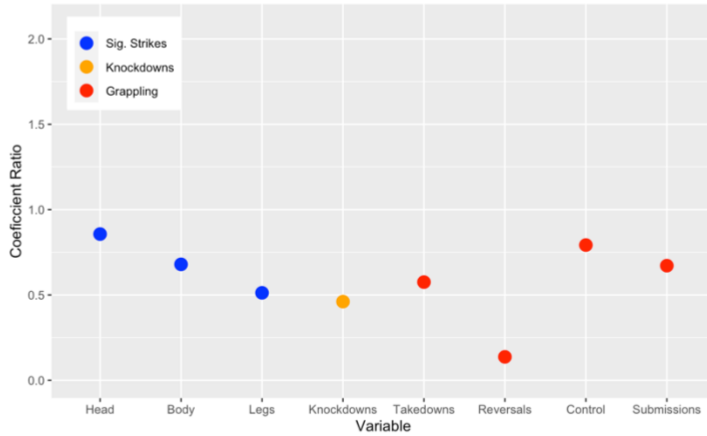


Derek Cleary appears to be the most beneficial judge for grapplers, and he scores 60.2% grappler friendly. There are not massive differences in his significant striking coefficient ratios, but body strikes are valued slightly less. The only other aspect that stands out with his striking is that he values knockdowns slightly higher than other judges.

Unlike Sal D'Amato, Derek Cleary appears to prefer grapplers with a more jiu-jitsu-oriented style. He has a very large coefficient for reversals as well as submission attempts, whereas he appears to score takedowns and control time around the same as other judges. While he still values taking down your opponent and controlling them, Cleary seems more interested in the offense that fighters can generate once the fight has hit the mat. These numbers suggest that even if a fighter ends up in a bad position, they can still do well on Cleary's scorecards if they threaten with submissions and/or reverse their opponent. This is not very surprising given that Cleary is a jiu-jitsu black belt, and his experience with jiu-jitsu very well may be influencing his scoring preferences.

Judge 3: Chris Lee (1323 rounds judged)

	Dependent variable:	
	model_winner	
	Non-Judge	Derek Cleary
	(1)	(2)
head_dif	0.196***	0.162***
	(0.006)	(0.011)
body_dif	0.149***	0.102***
	(0.009)	(0.015)
leg_dif	0.119***	0.106***
	(0.008)	(0.015)
nonsig_dif	0.027***	0.024***
	(0.003)	(0.007)
kd_dif	2.213***	2.371***
	(0.214)	(0.397)
td_dif	0.395***	0.405***
	(0.044)	(0.090)
rev_dif	0.409***	0.639**
	(0.122)	(0.259)
ctrl_dif	0.493***	0.533***
	(0.030)	(0.062)
sub_dif	0.939***	1.264***
	(0.091)	(0.185)
Constant	-0.035	-0.101
	(0.035)	(0.068)
Observations	7,018	1,713
Log Likelihood	-2,511.295	-672.927
Akaike Inf. Crit.	5,042.590	1,365.855
Note:	*p<0.1; **p<0.05; ***p<0.01	

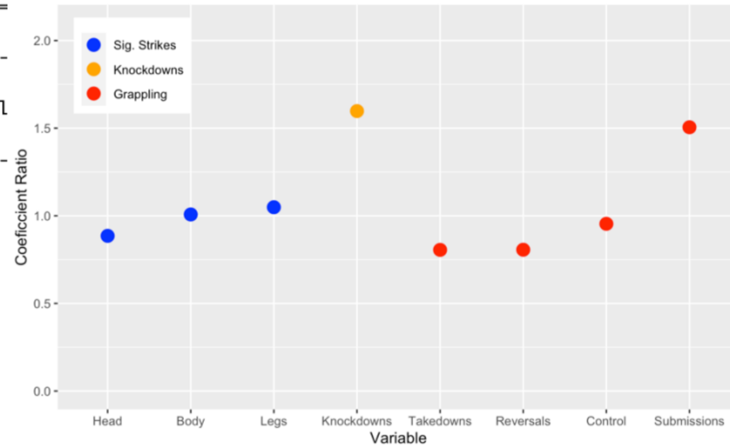


The Chris Lee model has much smaller coefficients overall, so the model was not able to identify as strong preferences in his scoring. He scores as the most striker friendly judge (60.4%) largely due to a very small coefficient for reversals. Similarly to Sal D'Amato, Chris Lee seems to favor strikes that land higher on the body, but he does not seem to value knockdowns nearly as much as the other judges.

The biggest standout with the grappling coefficients is the very small (and not significant) coefficient for reversals. There does not appear to be an extremely favored grappling style here, but we can see that Lee does seem to score control time relatively highly. While the model results are certainly the weakest for Chris Lee, they still do seem to indicate that he likes strikes that land higher on the body as well as control time.

Judge 4: Michael Bell (1179 rounds judged)

	Dependent variable:	
	model_winner	
	Non-Judge	Michael Bell
	(1)	(2)
head_dif	0.190***	0.168***
	(0.006)	(0.013)
body_dif	0.144***	0.145***
	(0.009)	(0.018)
leg_dif	0.117***	0.123***
	(0.008)	(0.020)
nonsig_dif	0.028***	0.034***
	(0.003)	(0.008)
kd_dif	2.073***	3.312***
	(0.206)	(0.687)
td_dif	0.411***	0.331***
	(0.044)	(0.104)
rev_dif	0.406***	0.328
	(0.121)	(0.297)
ctrl_dif	0.491***	0.468***
	(0.030)	(0.070)
sub_dif	0.892***	1.342***
	(0.090)	(0.253)
Constant	-0.039	-0.055
	(0.035)	(0.087)
Observations	7,108	1,173
Log Likelihood	-2,585.038	-416.824
Akaike Inf. Crit.	5,190.077	853.648
Note:		*p<0.1; **p<0.05; ***p<0.01

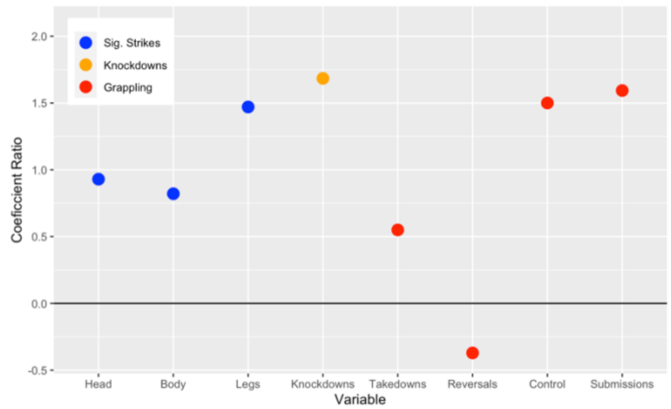


Michael bell comes in as the most neutral judge (50.6% striker friendly) and seems to have pretty unique scoring preferences. The coefficient ratios for head, body & leg significant strikes are all very close to one, but they do suggest he slightly prefers strikes lower on the body. What really stands out for striking, however, is how highly knockdowns are valued in Michael Bell's model. He values them second highest relative to the other judges examined with a coefficient ratio just above 1.5.

In terms of grappling coefficients, Michael Bell seems to value takedowns & reversals almost the same, which is just a bit less than the other judges. The coefficient ratio for control time is just below 1, but what really stands out for Bell is the ratio of around 1.5 for submission attempts. Michael Bell does not appear to have a clear style of fighter that he prefers, but these numbers do seem to indicate that he values large moments with the potential to end a fight very highly.

Judge 5: Junichiro Kamijo (1033 rounds judged)

	Dependent variable:	
	model_winner	
	Non-Judge	Junichiro Kamijo
	(1)	(2)
head_dif	0.195***	0.181***
	(0.006)	(0.016)
body_dif	0.147***	0.121***
	(0.009)	(0.021)
leg_dif	0.116***	0.170***
	(0.008)	(0.022)
nonsig_dif	0.028***	0.035***
	(0.003)	(0.009)
kd_dif	2.011***	3.386***
	(0.203)	(0.721)
td_dif	0.415***	0.228*
	(0.044)	(0.117)
rev_dif	0.433***	-0.160
	(0.121)	(0.385)
ctrl_dif	0.481***	0.722***
	(0.029)	(0.086)
sub_dif	0.925***	1.474***
	(0.089)	(0.312)
Constant	-0.038	-0.045
	(0.035)	(0.094)
Observations	7,131	1,021
Log Likelihood	-2,584.630	-357.040
Akaike Inf. Crit.	5,189.260	734.080
Note:	*p<0.1; **p<0.05; ***p<0.01	



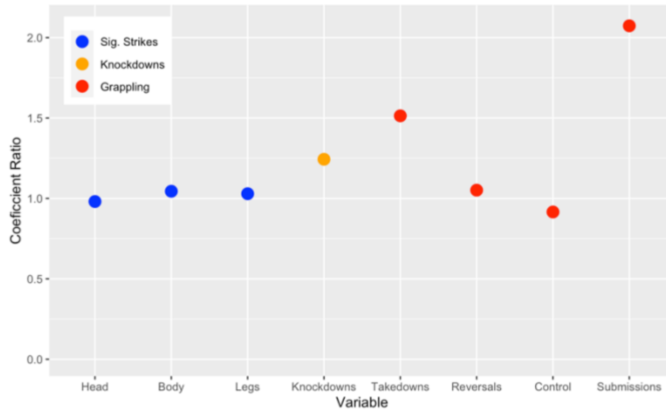
Junichiro Kamijo appears to score fights uniquely compared to the other judges. He is ranked as the second most striker friendly judge at 57.8%. This model is the only one where there is a negative coefficient: reversals. The p-value here is extremely high (~0.67) and this is really a product of multicollinearity more than anything. To put this coefficient into context, a coefficient of -.16 for reversals means that one reversal is valued around the same (but negative) as 13 seconds of control time in this model. This much control time (and probably more) will usually be obtained anyway after a reversal, so if you factor in everything that comes from them reversals should still be helping fighters in this model.

Looking at striking preferences, 2 clear aspects stand out for Kamijo: leg kicks and knockdowns. He appears to value leg kicks much more highly than other judges, and Kamijo's model is the only one where the coefficient for significant leg strikes is around the same as the coefficient for significant head strikes. The coefficient ratio for knockdowns is the highest among all the judges, and is slightly larger than the one previously seen for Michael Bell.

The grappling coefficients for Kamijo are unique and seem to paint a picture of how he scores grappling. As discussed earlier, this negative coefficient for reversals is misleading but it does seem to show that he does not value reversals nearly as highly as other judges. Additionally, Kamijo has the smallest coefficient ratio for takedowns out of all 6 judges examined. However, he has the largest coefficient ratio for control time as well as the second largest coefficient ratio for submission attempts. This wide discrepancy in grappling coefficient ratios does not appear to show a clear preference for/against grappling, but it seems to indicate what he likes to see in grapplers. Rather than being concerned with how the fight got to where it is (through a takedown or reversal), Kamijo seems much more interested in what a fighter can do once it gets there. Fighters who are able to control their opponent and threaten with submissions should fair well with these judging preferences, whereas fighters who get their opponents to dominant positions but are unable to control them should not score as well.

Judge 6: Eric Colón (790 rounds judged)

	Dependent variable:	
	model_winner	
	Non-Judge	Eric Colón
	(1)	(2)
head_dif	0.191***	0.187***
	(0.006)	(0.017)
body_dif	0.143***	0.150***
	(0.009)	(0.025)
leg_dif	0.119***	0.123***
	(0.008)	(0.023)
nonsig_dif	0.027***	0.031***
	(0.003)	(0.009)
kd_dif	2.017***	2.508***
	(0.201)	(0.756)
td_dif	0.396***	0.599***
	(0.043)	(0.143)
rev_dif	0.405***	0.426
	(0.119)	(0.372)
ctrl_dif	0.482***	0.441***
	(0.029)	(0.093)
sub_dif	0.916***	1.899***
	(0.088)	(0.310)
Constant	-0.042	-0.085
	(0.034)	(0.106)
Observations	7,166	787
Log Likelihood	-2,625.846	-282.673
Akaike Inf. Crit.	5,271.692	585.347
Note:	*p<0.1; **p<0.05; ***p<0.01	



Eric Colon ranks as the second most grappler friendly (56.9%) judge behind Cleary. His striking preferences align very closely with the rest of the judges, as his coefficient ratios for all the significant striking areas are extremely close to 1. The ratio for knockdowns is slightly larger, and he appears to prefer these around as much as D'Amato while not quite as much ad Bell & Kamijo.

The grappling coefficients seem to indicate that grapplers should fair well on Colon's scorecards. His coefficients for control time & reversal are close to the non-judge model, but the coefficients for both takedowns & submission attempts are both extremely larger. Given that Eric Colon is also a jiu-jitsu black belt, it is not surprising to see submission attempts valued so highly in his model. These coefficient ratios suggest that grapplers in general should have an advantage with Colon judging, and grapplers who are able to secure takedowns and threaten with submissions will do especially well on his scorecards.

Conclusion

Overall, I have been able to identify aspects of MMA judging are most important. While it is an almost impossible task to ask a judge to accordingly balance various strikes & grappling actions when scoring a fight, these models have helped show how these actions compare to each other.

These models back up the idea that at least at the extremes judges are scoring based off damage through knockdowns. Of course, all knockdowns are created differently and many may think these would be worth even more, but unfortunately the largest limitation of these models is being unable to capture the damage of strikes aside from these knockdowns. This research does indicate that striking higher to the body will be preferred by judges. Of course, this is subjective and depends on the strike, but it seems to back up the common belief many people have that strikes to the head are most important.

Grappling wise, it is not surprising that takedowns and reversals are valued around equally. The high emphasis on submission attempts may surprise some people (these people may instead call these "failed submissions") but because submission attempts have the potential to end the fight and are certainly a form of effective grappling, the rules would certainly say these are valuable. In fact, the rules mention submission attempts 4 times and they are already mentioned two sentences into defining effective striking and grappling (the number one criteria).

The individual judge analysis identified some judging preferences that could be of value to fighters, coaches or fans. While it is unlikely a fighter would drastically modify their game plan if at all (remember: this is a fight) it still may be of use to know how judges are looking at fights. I would not expect a fighter to change their striking targets much due to judge preferences, but certain aspects could still be of use. For example, if there is a judge who appears to heavily favor submission attempts, this could be of benefit. If a fighter is in a grappling exchange in a close round, and they (or their coach) are aware of this information, they could potentially focus on trying to offer a submission threat. Even if they know they likely won't be able to finish it, they could go for it knowing just threatening a submission could be the difference in a close round. At the end of the day a fighter has to fight their fight, so it is unlikely they will change their style regardless of how a fight is being judged, but having this information could certainly be of use to fighters and coaches.

References:

Wismer, David. "Predicting Ultimate Fight Championship (UFC) Judging Decisions Using Multivariate Linear Regression" Medium, 5 Oct. 2021, https://davidrwismer.medium.com/predicting-ultimate-fight-championship-ufc-judging-decisions-using-multivariate-linear-regression-450d2fa1a5bb.

El Ashker, Said. "Technical and tactical aspects that differentiate winning and losing performances in boxing" International Journal of Performance Analysis in Sport, vol. 11, no. 2, 2011, pp. 356-364, https://doi.org/10.1080/24748668.2011.11868555.

Stellpflug, Samuel, et al. "Analysis of the fight-ending chokes in the history of the Ultimate Fighting Championship" Phys Sportsmed, vol. 50, no. 1, Feb. 2022, pp. 60-63, https://doi.org/10.1080/00913847.2020.1866958.

Feldman, Todd. "The Way of the Fight: An Analysis of MMA Judging" Journal of Applied Sport Management, vol. 12, no. 2. 1 Dec. 2020, https://doi.org/10.7290/jasm120205
Lee, Herbert K., et al. "Did Lennox Lewis beat Evander Holyfield?: methods for analysing small sample interrater agreement problems" Journal of the Royal Statistical Society: Series D (The Statistician), vol. 51, no. 2, 20 Jun. 2002, pp. 129-146, https://doi.org/10.1111/1467-9884.00306.
Myers, Tony, and Balmer, Nigel. "The Impact of Crowd Noise on Officiating in Muay Thai: Achieving External Validity in an Experimental Setting" Frontiers in Psychology, vol. 3, no. 346, 2012, https://doi.org/10.3389/fpsyg.2012.00346.
Gift, Paul. "The impact of new judging criteria on 10-8 scores in MMA" Journal of Sports Analytics, vol. 7, no. 4, 2021, pp. 235-246, https://doi.org/10.3233/jsa-200478.
Martens, Jason P. "Raise Your Hands: The Influence of Post-Fight Nonverbal Pride on Fight Decisions" Journal of Nonverbal Behavior, vol. 47, no. 4, 2023, pp. 435-448, https://doi.org/10.1007/s10919-023-00435-8.
Holmes, Benjamin, et al. "Detecting individual preferences and erroneous verdicts in mixed martial arts judging using Bayesian hierarchical models" European Journal of Operational Research, vol. 312, no. 2, 2024, 16 Jan. 2024, pp. 733-745. https://doi.org/10.1016/j.ejor.2023.07.004.
Sorokowski, Piotr, et al. "The influence of the boxing stance on performance in professional boxers" Anthropological Review, vol. 77, no. 3, 20 Dec. 2014, pp. 347-353, https://doi.org/10.2478/anre-2014-0025.
Balmer, NJ, et al. "Do judges enhance home advantage in European championship boxing?" Journal of Sports Sciences, vol. 23, no. 4, 23 Apr. 2005, pp. 409-416, https://doi.org/10.1080/02640410400021583.
Crossley, Tom. "Performance profiling in professional mixed martial arts (MMA): comparing winning and losing performances." University of Chester, 2015, https://chesterrep.openrepository.com/handle/10034/618441.
Miarka, Bianca, et al. "Ending an MMA combat bout: specific striking techniques which determine the type of outcome" Ido Movement for Culture. Journal of Martial Arts Anthropology, vol. 20, no. 3, Jul. 2020, pp. 9-17, https://doi.org/10.14589/ido.20.3.2.
James, Lachlan, et al. "Identifying the performance characteristics of a winning outcome in elite mixed martial arts competition" Journal of Science and Medicine in Sport, vol. 20, no. 3, 20 mar. 2017, pp. 296-301, https://doi.org/10.1016/j.jsams.2016.08.001.
Latham, James. "An appraisal of judging criteria in relation to performance in elite male amateur boxing" University of Chester, 28 Sep. 2018, https://chesterrep.openrepository.com/handle/10034/621876.
Latshaw, Nate. "Introducing JudgeAI" LiteralFightNerd, 19 Feb. 2021, https://literalfightnerd.com/posts/2021-02-19-introducing-judgeai/.
Warnick, J. E. "Specification of variables predictive of victories in the sport of boxing" Perceptual and Motor Skills, vol. 105, no. 1, 2007, pp. 153-158, https://doi.org/10.2466/PMS.105.5.153-158

Myers, Tony, et al. "Evidence Of Nationalistic Bias In Muaythai" Journal of Sports Science & Medicine, 5 Jul. 2006, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3863918/.
Latyshev, Sergii, et al. "Determination of model characteristics of martial mixed arts fighters" E3S Web of Conferences, 2021, pp. 09035, https://doi.org/10.1051/e3sconf/202127309035.
Loffing, Florian, and Hagemann, Norbert. "Pushing through evolution? Incidence and fight records of left-oriented fighters in professional boxing history" Laterality, vol. 20, no. 3, 20 May 2015, pp. 270-286, https://doi.org/10.1080/1357650x.2014.961471.
Wild, Graham. "The puncher's chance in professional mixed martial arts competition" Journal of Human Sport and Exercise, vol. 17, no. 2, 2022, pp. 370-377, https://doi.org/10.14198/jhse.2022.172.12.
Dunn, Emily C., et al. "Human behaviours associated with dominance in elite amateur boxing bouts: A comparison of winners and losers under the Ten Point Must System" PLOS ONE, vol. 12, no. 12, 2017, https://doi.org/10.1371/journal.pone.0188675.
Hagemann, N. "The advantage of being left-handed in interactive sports" Attention, Perception & Psychophysics, vol. 71, no. 7, 2009, pp. 1641-1648, https://doi.org/10.3758/app.71.7.1641.
Baker, Joseph, and Schorer, Jörg. "The Southpaw Advantage? - Lateral Preference in Mixed Martial Arts" PLoS ONE, vol. 8, no. 11, 2013, https://doi.org/10.1371/journal.pone.0079793.
Collier, Trevor, et al. "Aggression in Mixed Martial Arts: An Analysis of the Likelihood of Winning a Decision" Violence and Aggression in Sporting Contests, vol. 4, 1 Jan. 2011, pp. 97-109, https://doi.org/10.1007/978-1-4419-6630-8_7.
Miarka, Bianca, et al. "Differences by Ending Rounds and Other Rounds in Time-Motion Analysis of Mixed Martial Arts: Implications for Assessment and Training" J Strength Cond Res, vol. 32, no. 2, Feb. 2018, pp. 534-544, https://doi.org/10.1519/JSC.0000000000001804.
Hoelbling, Dominik, et al. "JudgED: Comparison between Kickboxing Referee Performance at a Novel Serious Game for Judging Improvement and at World Championships" Applied Sciences, vol. 13, no. 17, 2023, pp. 9549, https://doi.org/10.3390/app13179549.
Myers, Tony, et al. "A comparison of the effect of two different judging systems on the technique selection of Muay Thai competitors" Journal of Human Sport and Exercise, vol. 8, no. 3, 5 Dec. 2013, pp. 761-777, https://doi.org/10.4100/jhse.2013.83.01.