The Wrestling Pickems Generator

University of Northern British Columbia

Department of Computer Science

CPSC 499 Final Project Report

Elizabeth Norman, Nicholas Slugocki

December 15, 2022

Declaration

We, Elizabeth Norman and Nicholas Slugocki, declare the proposed project work is based on our original work, except on ideas or data within acknowledged citations. We declare the proposed work is carried out solely by ourselves and has not been submitted previously or concurrently for any other course or degree from UNBC or other institutes.

Abstract

This paper will detail the process of building a linear regression model that will predict the outcomes of professional wrestling matches. A lengthy description of professional wrestling is provided, noting that the scripted nature of wrestling was going to make it difficult to create a model out of. Many features, such as championship belts held by a wrestler or their social media following, were explored in the creation of the model. A scoring system such as the ELO system is considered as well.

As many features (such as a wrestler being a heel or face) were of a non-numeric categorical variety, the paper describes how to handle such features through two methods: one hot encoding and dummy encoding. These turn categorical features into useable numeric values for use in our regression model. We then describe the usefulness of selected features, providing graphs to show which features were useful or not in our model. The model went through many incarnations before a winning combination of features producing a 72% success rate was found and is discussed.

The paper concludes with an overview of potential applications of the model, difficulties faced when building the model, and an analysis of our model's results.

Table of Contents

List of Figures

Chapter 1

Introduction

Professional wrestling is a wildly popular combat sport, with companies such as World Wrestling Entertainment (WWE) or All Elite Wrestling (AEW) producing weekly television shows and special pay-per-view events such as WWE's Wrestlemania, showcasing matches and promotional segments performed by professional wrestlers.

Since we are both fans of professional wrestling, we became interested in creating a model that will accurately predict the outcome of wrestling matches. Sports prediction models exist for nearly all sports already, in both combat sports such as mixed-martial arts and team or racing sports, as they are a useful asset to gamblers, fantasy sport players, sports analysts and more. However, a search for a professional wrestling prediction model came up empty. This is likely because professional wrestling is not structured like an actual sport, as the outcomes are predetermined and the matches are staged, with the wrestlers performing under melodramatic personas to tell a story. This illusion that the wrestling companies and wrestler's maintain that everything the audience see is authentic is known in the industry as "kayfabe". To incorporate this, we will create our model using features that are both in and out of kayfabe, such as a wrestler being a heel or face (villain or hero) and their previous match data, to see what kind of outcomes we will get.

Chapter 2.1 provides a background for our model: describing our expectations, data collection methods, and features collected and a description of said features. We finish this section with discussing which tools we will use to solve our problems.
In 2.2, theoretical results are discussed.
Numerical results are in 2.3

Chapter 3 concludes our paper, starting with 3.1's discussion on future possibilities of our methods and ideas.
Moving into 3.2, we cover the difficulties we faced throughout our project.
Finally, 3.3 is an analysis on the significance of our findings.

Chapter 2

Main Results

2.1 Background

As noted in our introduction, professional wrestling is a staged combat sport popular around the world and organized by a wide array of production companies such as WWE and AEW, who put on shows for viewers to enjoy wrestling matches. Due to a personal preference towards All Elite Wrestling's products, our professional wrestling predictions model will analyze match data from all of AEW single's matches in combination with a wide array of data from 60 individual professional wrestlers who are employees of AEW or have competed in AEW matches.

Ultimately, professional wrestling is an art form. Television shows such as AEW's Dynamite or WWE's Smackdown are tell a dramatic storyline pushed forward through staged matches full of high adrenaline stunts and rehearsed, theatrical combat. The illusion that what we are seeing is real is referred to as 'kayfabe' in the industry. To create a model to accurately predict the outcome of these matches, we will need to consider features that incorporate data from both in and out of kayfabe.

2.1.1 A note on expectations

The outcomes of wrestling matches are often determined by a sole show runner, such as a company's president/CEO (AEW's Tony Khan) or head of creative (WWE's Paul Levesque, aka Triple H) and are thus often unpredictable. The outcomes of matches serve a grand story that have deeper meaning than mere brawn and fighting skills. We entered this project truly as pure fans of wrestling, fully aware that our model was unlikely to be successful, but our passion for the sport overwhelmed us and we felt we had to research this. Thus, we ventured forth in exploring what happens when you try to make a model out of something that is so subjective and unmodel-able, just for the sake of curiosity.

2.1.2 Data Collection

Data was separated into two sections: match data and wrestler data. Data was difficult to gather due to the nature of professional wrestling, as easily accessible tables clearly defining a wrestler's statistics such as weight, height, and match history are not available in the same manner that a mixed-martial artist's statistics would be accessible for analysts to use. In the following two sections, we will provide a background on how data was collected as well as provide a list of features gathered for both the wrestler and match data. This will also serve as a useful glossary of wrestling terms for those unfamiliar with professional wrestling.

2.1.3 Match Data Gathering & Features Selected

For the match data, we had to build a website scraper that scraped up the information from cagematch.net, an online database of wrestling data. We received permission from them to scrape as we desired, so as long as it was not in excess. The scraped strings were then parsed into useful bits of data and the data was placed into a useable .csv file.

Wrestler 1 & 2 The match data naturally started with the wrestler's involved.

Winner Who won the match.

Location Where the specific match took place.

Time The length of the match.

Show Many of the larger wrestling companies such as WWE or AEW will have a variety of shows available to watch across many platforms, including cable television programming, streaming exclusives, YouTube shows, and pay-per-view events. We took note of which show each match took place on.

Championship Match A boolean to flag if a match was for a championship belt or not. Wrestler's compete for belts and there are a wide array of belts in any one particular company such as World Championship belts for both men and women's wrestling.

Title The particular title being fought for if the match was a champsionship match.

Champion at start & champion at end Two columns to keep track of a transfer of a belt. This appears clunky and was the result of the difficulties parsing data from the web scraper.

Title change This inform us if the title changed at the end of the match.

Date Day the match took place

2.1.4 Wrestler Data & Features Selected

Unfortunately, it was not feasible to put together a scraper like what was done for the match data for the wrestler's stats and the wrestler data .csv file was filled manually. We compiled the statistics of 60 wrestlers who have performed at AEW events.

Name Wrestler's often perform under stage names and take on exaggerated personas.

Heel or Face Wrestler's can be seen as a good guy, a face, or a bad guy, a heel. Their roles shape the story line that is told on the stage.

Recent turn Wrestler's usually don't stay as a heel or face throughout their entire career. When they switch roles, through a scripted scene (called a "promo") or match, this is referred to as "a turn". This was filled out in a subjective manner, where we said true if the turn had happened in the past 2-6 months depending on how the storyline was unfolding.

Has Belt & Tenure Does the wrestler currently hold a belt and for how long? Belt's are often changed cyclically to ensure that there's always excitement around upcoming special events.

Total Belts Held We totalled not just AEW belts held but belts held across most major companies in a wrestler's career. We only included major companies as we felt it would be difficult to consider the value of belts from smaller, independent, regional companies. An example of some of the major companies we included were: WWE, New Japan Pro Wrestling, Ring of Honor, and more.

Physical Traits We included height and weight in our data, as it is easily obtainable information. Wrestler's weight and height are announced at the beginning of matches but are both often exaggerated to make wrestler's appear much bigger or smaller than they actually are.

Win percentages We included the overall win percentages of a wrestler's career and their career at AEW. Here would be an ideal time to mention the concept of a "jobber": a slang term for a professional wrestler who's job (hence the name) it is to lose. We included a variety of jobbers in our data.

Number of Dave Meltzer's 5 Star Ratings Professional wrestling journalist and sports historian Dave Meltzer has assigned ratings, of up to 5 stars (occasionally, some notable matches have received higher stars), to all matches in professional wrestling since 1982. To receive 5 stars is a huge and rare honour so we included how many matches a wrestler has achieved this within our data.

PWI Ranking Pro Wrestling Illustrated is a magazine that has been in print since 1979. Since 1991 they have published a list of the top 500 male wrestlers in the industry and added the top 150 women list in 2008. This ranking list, and magazine, are published "in kayfabe", which we felt could be a useful asset in determining the win factor of a wrestler.

Social media followers We were curious if a wrestler's online presence, where they can promote their matches to a wide audience, factored into their success. So we totalled their Instagram and Twitter followers and included that in our data.

shopaew.com Merch Count Following from the above's implication that this is all a popularity contest, we totalled the number of merchandise a wrestler has available for sale.

Hometown Where the wrestler is from.

Tenure at AEW Length of time as a wrestler contracted to AEW.

Ultimately only some of these features became part of our model as time progressed completing this project. Which features ended up as part of the project will be discussed further in the paper.

2.1.5 Models we used

When trying to consider how to assign rankings to individual wrestler's, we experimented with the ELO rating system. The ELO system was developed by Arpad Elo, a Hungarian-American physicist who developed this system to rank chess players. The performance of a person in the ELO system is inferred from wins, losses, and draws against other players. The score of a person is dependent on the ratings of their opponents and the results scored against them. A winning person takes points from a losing person and the number of points taken is determined by the difference in their ratings.

- Higher ranking winners will take a few points from the lower ranking losers

- Lower ranking winners will take a lot of points from the higher ranking winner
- In the case of a draw, the lower rated player gains some points

A player's expected score is their probability of winning plus half of their probability of drawing. Elo assumed that the chess performance of a player in a game is a random variable and that it follows a normally distributed bell-shaped curve over time, meaning that a player could perform worse or better in successive games, the mean value of their performances would remain the same. This mean value for a given player then changes slowly over time.

The predictor of a match then is determined by the different in the ratings between two wrestlers. Let wrestlers A and B have ratings R^A and R^B, respectively. The expected outcome for wrestler A would then be [1]:

$$E_A = \frac{1}{1 + 10^{(R^B - R^A)/400}} \tag{2.1}$$

And similarly, the expected outcome for wrestler B is:

$$E_A = \frac{1}{1 + 10^{(R^A - R^B)/400}} \tag{2.2}$$

As an example, after feeding our program the entirety of AEW's match data the wrestler's Jungle Boy Jack Perry received a score of 1577 and Luchasaurus received 1639. The expected scores are then:

Luchasaurus:

$$\frac{1}{1 + 10^{(1577-1639)/400}} = 58.8\% \tag{2.3}$$

Jungle Boy Jack Perry:

$$\frac{1}{1 + 10^{(1639-1577)/400}} = 41.2 \tag{2.4}$$

After utilizing the match data to create ELO scores for our selected wrestler's, we then created a simple linear regression approach utilizing a combination of our wrestler's features to calculate a model. Simply put, a linear regression model describes a relationship between a dependant variable Y and one or more independent variables X. In our models, our Y was the ELO score of a wrestler, later just the win/loss ratio at AEW, and our X was a combination of some of the features listed above in the wrestler data section.

Many of our selected features were categorical features, which are discrete, yet non-numerical features. Incorporating these into a mathematical model was the primary problem of our project. We explain in our theoretical results section how we approached this problem.

2.1.6 Tools used

We utilized a handful of Jupyter notebooks and python programs to complete our work alongside several Python packages. The packages used were:

- Numpy
- Pandas
- Seaborn
- Matplotlib

- scikit-learn
- BeautifulSoup

2.2 Theoretical results

The math theoretical problem we wished to explore with this project was discrete and categorical features using a linear regression machine learning model.

2.2.1 Discrete Valued Features

Discrete features are not dissimilar to continuous features as the same principles apply. The linear regression takes the form

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ...$$

And all one must do is treat discrete integers as if they were continuous to get the appropriate results. At that point our θ estimators can be calculated using methods such as the least squares method, as normal.

2.2.2 One Hot Encoding

One of the problems that need to be solved when dealing with a linear regression model is how to incorporate *categorical data*. Categorical data is data that falls into discrete categories but is non-numerical.

One way to achieve this is using the method of one-hot encoding. The term *one-hot encoding* comes from a technique in electrical engineering where an encoding is created by enabling a single bit in a series of binary registers. A numerical example and a categorical example of one-hot encoding would look like the following:

Integer	**One-Hot Encoding**
0	0001
1	0010
2	0100
3	1000

Colour	**One-Hot Encoding**
Red	001
Green	010
Blue	100

We can apply this idea to categories. In the case of our wrestling outcome predictions, we would select a set of categories such as whether the wrestler is a face ("good guy") or heel ("bad guy"), for example:

Wrestler Name	**Heel/Face**
Jon Moxley	Face
MJF	Heel

In such a case, we could split the Heel/Face category into their own separate columns. We could then assign a 1 in the column is the data set is included in that category, or a 0 if they are not.

Wrestler Name	**Face**	**Heel**
Jon Moxley	1	0
MJF	0	1

Given this data set, suppose we have a linear regression equation. Let us define β to be our unknown least square estimator and z to be one of our categorical variables. In that case we have a data vector that looks like:

$$\begin{bmatrix} x_0 \\ x_1 \\ z \end{bmatrix} \Rightarrow \begin{bmatrix} x_0 \\ x_1 \\ z_1 \\ z_2 \\ z_3 \end{bmatrix} \quad (e.g. \begin{bmatrix} \text{belts held} \\ \text{career length} \\ \text{face/heel} \end{bmatrix} \Rightarrow \begin{bmatrix} \text{belts held} \\ \text{career length} \\ \text{face} \\ \text{heel} \end{bmatrix})$$

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3$$

It follows that for any wrestler that is a face character, we have:

$$\begin{aligned} \hat{y} &= \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \beta_1(1) + \beta_2(0) + \beta_3(0) \\ &= \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \beta_1(1) + \cancel{\beta_2(0) + \beta_3(0)} \\ &= \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \beta_1 \end{aligned}$$

This would apply likewise for any wrestler that is a heel. In such a case the resulting equation would be $\hat{y} = x_0 + \theta_1 x_1 + \theta_2 x_2 + \beta_f(0) + \beta_h(1) = x_0 + \theta_1 x_1 + \theta_2 x_2 + \beta_h$. This gives us a system where a single value in the linear regression equation can be "active" for each data item, but "deactivated" or set to zero if it is not. This can then influence the linear regression model along one discrete axis if it is part of that category, but not others.

2.2.3 Dummy Encoding & Collinearity

One problem with one-hot encoding is that it does not account for collinearity between items in a category. Some categorical features have items where if it is not in one category, it then follows that it must be the other category.

Let's once again take the example of a wrestler being a heel or face. One way to illustrate this collinearity that most people would be familiar with is the idea of probabilities. Suppose we have a population of n wrestlers, and the number of faces are f while the number of faces are h. We can then deduce that the amount of heels implicitly from f because $h = n - f$. This means that having a separate category encode for both values is encoding for redundant information, and can skew the results of our model.

To get around this we can simply drop one of the categories and only include one in our model.

Wrestler Name	Face	Heel
Jon Moxley	1	0
MJF	0	1

$\Rightarrow$

Wrestler Name	Face
Jon Moxley	1
MJF	0

This may not be suitable for every kind of variable. For instance, one possible categorical feature that we considered looking is a wrestlers country of origin. AEW is a wrestling brand from the USA, so we wondered if country had any role in the likelihood of a wrestler winning matches. However, a wrestler not being from, say, the USA does not necessarily imply with any kind of certainty that there is any other "default" country they must be assigned to, so it would be difficult to argue that there is a sufficient collinear relationship.

2.3 Numerical results

2.3.1 Feature Selection: Worse features

While doing feature selection we tried to look at country to see if there was any bias for a particular country. As mentioned earlier, AEW is a brand from the USA so we wondered if there was more bias towards American wrestlers. It appears, however, that our country data just did not provide enough information to make any valid predictions, and there didn't seem to be any particular trend or biases that could be ascertained intuitively so we didn't explore that much.

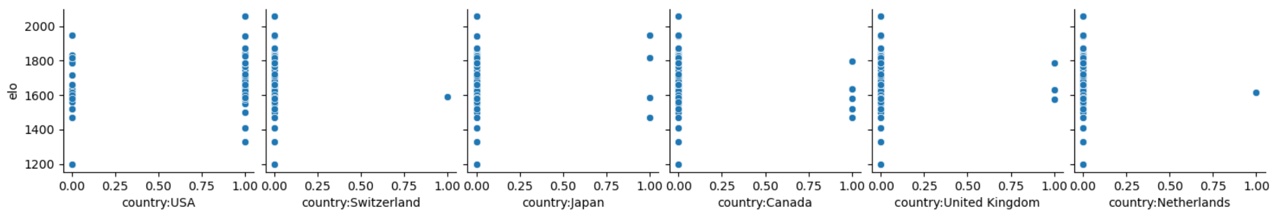

Figure 2.1: First set of countries.

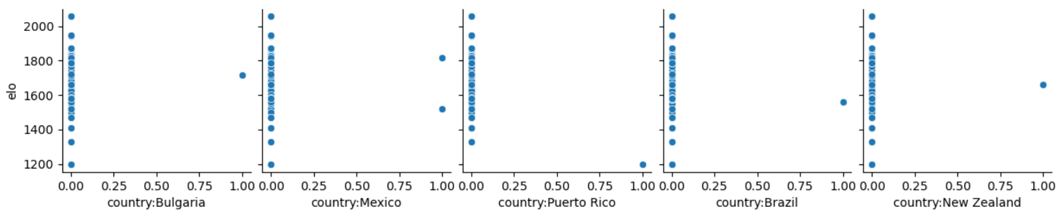

Figure 2.2: Second set of countries.

Other examples of features that were particularly bad were things like height and weight. We expected this to the be the case. Although professional wrestling requires an incredible amount of athleticism, it is largely designed for entertainment purposes and often about telling a story. The build of a particular wrestler often doesn't play as big of a part as it may in genuine combat sports.

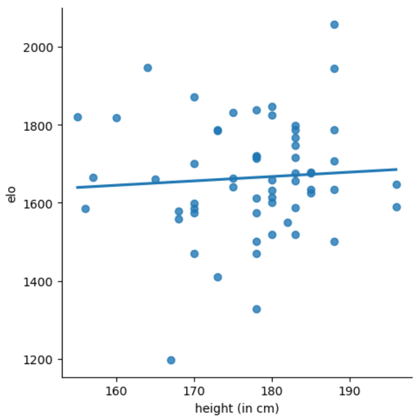

Figure 2.3: Height vs Elo rating.

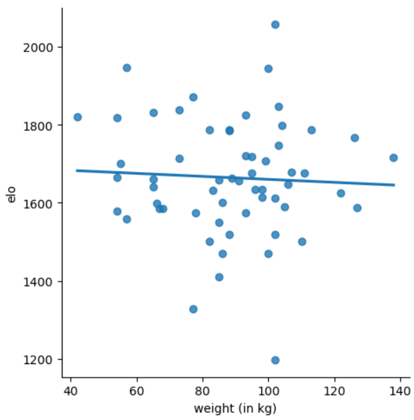

Figure 2.4: Weight vs Elo rating.

2.3.2 Feature Selection: Better Features

Some of the betters features we found were:

- Wrestler currently has belt?: Yes
- How many belts the wrestler has held previously
- How many merch items the wrestler has in the AEW store

The first two would seem to make intuitive sense. Generally, if a wrestler has a recent history of winning, they will be in possession of a belt, and presumably a wrestler company would want to take the opportunity to promote that wrestler with fans if they're popular. If a wrestler gets defeated too quickly, it might cause the title to lose "prestige". One may also conjecture that it's a good marketing opportunity for the company to spend some time promoting wrestlers that fans are interested in seeing as the winner.

Our thoughts that marketing would seem to have some kind of correlation to the likelihood a wrestler's win rate seems to have been correct as well.The number of merch items in the AEW store seemed to have an above average correspondence with the wrestlers' win outcomes.

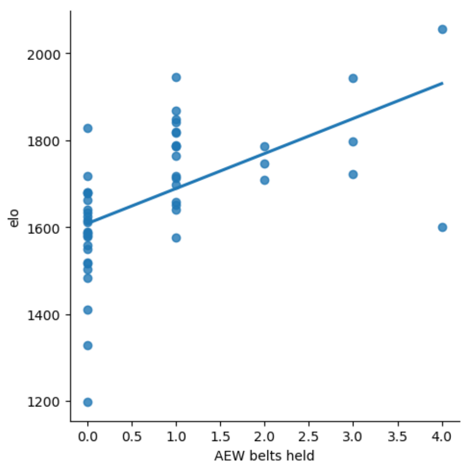

Figure 2.5: Number of Belts Held vs Elo rating.

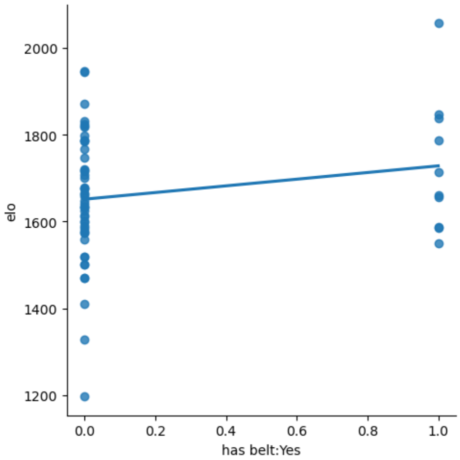

Figure 2.6: Wrestler Has Belt vs Elo rating.

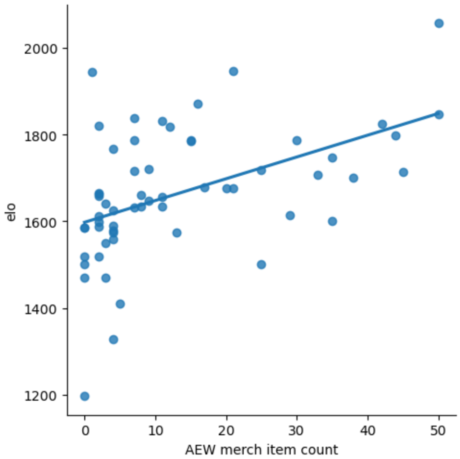

Figure 2.7: Wrestler Has Belt vs Elo rating.

2.3.3 Final Results

Our final attempt ended with our model guessing 13/18 matches correct, and an RMSE of 10.37. For us, this is a huge achievement.

Our final selection of features were the following:

- AEW belts held over time
- Social media followers (Twitter + Instagram)
- Has belt?: Yes
- AEW merch item count
- Belts held over time (from all major companies)

Additionally for our predicted outcome, $\hat{y}$, we ended up using a simple percentage winrate instead of Elo. After various tests, we ultimately found this to be a better predictor than the Elo system. The end result is that our model was able to guess 13/18 matches correct. For us, that is a huge accomplishment. Two of the most satisfying results from this is that our model correctly predicted the outcome for two wrestlers that not only had our model not seen before, but had never participated in AEW as well. In the case of Saraya in particular, we credit this to the inclusion of the social media account. Saraya is a wildly popular wrestler who had wrestled in past promotions, with 8,700,000 followers on social media.

	w1	w2	winner	show	predicted_1	predicted_2	predicted_winner	correct
0	Eddie Kingston	Jun Akiyama	Eddie Kingston	Full Gear	62.785987	57.522760	Eddie Kingston	True
1	Eddie Kingston	Tomohiro Ishii	Eddie Kingston	All Out	62.785987	54.262219	Eddie Kingston	True
2	Jon Moxley	MJF	MJF	Full Gear	77.109777	68.097929	Jon Moxley	False
3	Wardlow	MJF	Wardlow	Double or Nothing	89.661258	68.097929	Wardlow	True
4	Dr. Britt Baker DMD	Saraya	Saraya	Full Gear	71.946958	86.215226	Saraya	True
5	Brian Cage	Ricky Starks	Ricky Starks	Full Gear	62.902005	63.625480	Ricky Starks	True
6	Powerhouse Hobbs	Ricky Starks	Powerhouse Hobbs	All Out	58.529935	63.625480	Ricky Starks	False
7	Jack Perry	Luchasaurus	Jack Perry	Full Gear	65.392492	64.358326	Jack Perry	True
8	Jade Cargill	Nyla Rose	Jade Cargill	Full Gear	169.548665	65.420514	Jade Cargill	True
9	Jade Cargill	Anna Jay	Jade Cargill	Double or Nothing	169.548665	59.738372	Jade Cargill	True
10	Jade Cargill	Athena	Jade Cargill	All Out	169.548665	61.074913	Jade Cargill	True
11	Toni Storm	Jamie Hayter	Jamie Hayter	Full Gear	65.369841	58.555669	Toni Storm	False
12	Toni Storm	Dr. Britt Baker DMD	Toni Storm	All Out	65.369841	71.946958	Dr. Britt Baker DMD	False
13	CM Punk	Adam Page	CM Punk	Double or Nothing	80.468331	72.466314	CM Punk	True
14	CM Punk	Jon Moxley	CM Punk	All Out	80.468331	77.109777	CM Punk	True
15	Chris Jericho	Bryan Danielson	Chris Jericho	All Out	88.011774	83.925542	Chris Jericho	True
16	Hook	Angelo Parker	Hook	All Out	87.827986	57.940601	Hook	True
17	Christian Cage	Jack Perry	Christian Cage	All Out	55.516349	65.392492	Jack Perry	False

```
fg_matchup_df['correct'].value_counts()
```

```
True     13
False     5
```

Figure 2.8: Wrestler Has Belt vs Elo rating.

Chapter 3

Conclusion

3.1 Difficulties

Discovering how to score wrestlers and how to handle categorical data were the two problems we had to take time to research to solve, but we came through with answers rather quickly.

Gathering wrestling data was a personal difficulty just due to the intensely tedious nature of it, as the data was gathered from a variety different websites. The time to create a dozen or so web crawlers for all of the pages used would have been too great for it to be a feasible option. If we decide to pursue this topic further, however, we would take the time to design web crawlers to scrape the necessary data.

Despite the effort put into the ELO system and categorical features, our results worked better without the ELO and most of the categorical features. We are choosing to view this as a lesson in keeping things simple.

As can be gathered from the short length of this section, working on our project was overall a smooth experience.

3.2 Future considerations

It would be interesting for us to pursue creating models for other wrestling companies, to ask what statistics would create the most accurate results for a company like the WWE or New Japan Pro Wrestling. Our results showed that the AEW features "sweet spot" highly valued features such as amount of belts held and merchandise counts, and would that carry over to other promotions? A promotion that values verisimilitude over theatrics, such as Ring of Honor, could likely value a wrestler's physical traits over their online presence.

3.3 Significance

With the 13/18 correct results produced by our linear regression model formed from a variety of wrestler's and match statistics, we can say that we are incredibly thrilled to see that we produced a some-what successful model. Going into this project fully aware that we were likely to come out on the other side with a model that produced at best random results to see this outcome has been immensely rewarding.

It is uncertain how useful a model like this could be in any commercial application, as sports betting for professional wrestling is basically non-existent as an industry due to the scripted nature of it all. Our inspiration for this project stems from the "pickems belt", a toy replica of the WWE World Championship belt that our social circle often competes for with significant pay-per-view's such as Wrestlemania. The concept is simple: whoever guesses the most correct matches wins custody of the belt, ties are determined with tie breakers such as "how many times with this wrestler use their signature move?" Will we use this formula for future AEW events? Likely not. Trying to guess how you think a story will turn out and discussing it with your friends is what makes wrestling fun- we as fans are aware of the theater of it all, and trying to interpret the story for yourself is part of the appeal.

Ultimately we recognize that our project was built for novelty's sake, an experiment in seeing if it was even possible to make a successful model from something so subjective. It may not have many application uses, but it was an interesting journey into a passion of ours. The knowledge we gained from this regarding regression models and scoring systems could be very useful if we aspired to pursue creating a model for other sports.

Bibliography

[1] Raghav Mittal. What is an elo rating?, 2020. Last accessed 13 December 2022.