Connor Lovejoy, Dillon Tranquillo, Raanan Gurewitsch

July 31, 2018

INFSCI 1092 - Final Project

Machine Learning, Human Lifting: Using Data Analytics to Inform Rules in Powerlifting

Abstract

As the 105th annual Powerlifting Competition at the University of Pittsburgh School of Computing and Information quickly approaches, the rules committee, under the leadership of former heavyweight champion Dmitriy Babichenko, has enlisted the help of University of Pittsburgh's finest amateur data scientists to assist in determining two important rules for the competition. The first regards the use of weightlifting equipment and the second concerns the distribution of participants into fair age groups. To do this, we obtained a dataset from OpenPowerlifting (PS), an organization which tracks the results of competitors at powerlifting competitions around the world. The dataset, which contains information on over 3,000 meets and over 300,000 lifts, enabled us to analyze specific patterns relating to these two specific rules. We developed to predictive models to helps determine how equipment use affects performance. The first model was a Random Forest Classifier which attempted to predict competitor placing (as a binary win or lose). The second model was a Linear Regression model which attempted to predict Total amount lifted for a meet based on Wilk's score. We found that our Random Forest classifier had a moderate predictive ability and the equipment usage had little effect on a competitors ability to win. We also found in our regression analysis that the predictive ability of the Wilk's score does improve over more extreme equipment usage, from which we can

conclude that equipment may improve the Wilk's score's ability to normalize lifter strength across equipment usage.

Introduction

In this study we analyze a hotly debated question in the weightlifting community: does equipment usage impact performance and if so, how? Equipment usage is a cause of significant controversy among powerlifters**.** Some individuals prefer using certain supports or wearing specific gear to improve grip, balance and form, while others see equipment as a hindrance on their performance. For those who lift weights only to stay in shape, this is a matter of personal preference and little other significance. However, in the context of a powerlifting competition, whether certain equipment offer an unfair advantage is especially relevant. Therefore, using the OP dataset, we explore the empirical evidence for equipment's impact on performance in order to inform suggestions for equipment usage guidelines.

Methodology

Random Forest Model

The first experiment we ran was predicting an athletes ability to win based on equipment usage. This algorithm makes the most sense for what we are trying to model because what we want to actually create it a hierarchy of features that are important and we want to keep track of these feature values. Random Forest is good for this because the features are built into a binary decision tree which orders the features in a hierarchy of importance. We created two models with this algorithm, one that does not use equipment usage as a feature and one that does use equipment usage as a feature. The feature for the first model are Body weight, best squat, best bench, best deadlift, Wilks score and the features for the second model are body weight, best squat, best bench, best deadlift, Wilks score, equipment_wraps, equipment_raw, equipment_multi, equipment_single. The values that we are trying to predict are place, which is defined as a 1 (for first place) and 0 for all of the other places.

To organize our experiment, we split the dataset by equipment type into three categories; raw (no equipment), single-ply (bodysuit), wraps, and multi-ply. We then generate a correlation matrix to evaluate the predictive power of our dataset's continuous variables.

Table 1: Correlation Matrix

	Age	**Bodyweight (Kg)**	**Best Squat (Kg)**	**Best Bench (Kg)**	**Best Deadlift (Kg)**	**Total (Kg)**	**Wilks**
Age	1.000000	0.105273	-0.033177	0.034536	-0.037746	-0.016277	-0.071883
Bodyweight (Kg)	0.105273	1.000000	0.653410	0.663659	0.639776	0.674403	0.207940
Best Squat (Kg)	-0.033177	0.653410	1.000000	0.910896	0.907176	0.976353	0.789318
Best Bench (Kg)	0.034536	0.663659	0.910896	1.000000	0.877395	0.956746	0.715040
Best Deadlift	-0.037746	0.639776	0.907176	0.877395	1.000000	0.962885	0.727985
Total (Kg)	-0.016277	0.674403	0.976353	0.956746	0.962885	1.000000	0.773686
Wilks	-0.071883	0.207940	0.789318	0.715040	0.727985	0.773686	1.000000

In each dataset we isolate the Wilks Coefficient, a measure of overall weightlifting ability (in Kg) which controls for both body weight and gender[1] and is the only continuous value that is predictive of an individual's performance, in order to evaluate the impact of equipment usage on weightlifting performance and whether certain equipment offer an unfair advantage.

Then, using the sklearn library in Python 3, we trained a simple linear regression model to predict performance in the three events at the SCI powerlifting competition; squat, bench press and deadlift. The following table organizes the nine models, which predict for each type of equipment the performance of an individual in each event with the generic equation,

$$y_i = \alpha + \beta x_i + \varepsilon_i .$$

where α is the intercept, β is the coefficient for Wilks, and ε is the standard error.
Once the models have been trained, we use them to plot (by equipment type) the predicted outcomes of all three powerlifting events for a random sample of powerlifters based on their Wilks index. The differences between equipment groups in predicted performance for each event then informs our advice on implementing a rule concerning equipment usage.

Results and Discussion

When we first began working with this dataset we were going to analyze the relationship between age and powerlifting performance, however when we began working with the dataset it became clear that age was not a clear indicator of lifting performance. In Figure 4, we can see that a 40 year old lifts just as much as a 20 year old. Since we would not be able to predict age based on performance, we shifted our attention to equipment usage and its effect on performance.

The results for the linear regression model are depicted in the two figures below. The first figure shows the coefficients and the intercepts of all of the models.. The second figure, which is more interesting, shoes the R-squared values for each equipment type and each lift. What we see here is that, except for deadlift, the Wilk's scores predictive ability does increase over more extreme equipment use except for deadlift. This is an interesting abnormality. A possible explanation for this is that deadlifting is already more normalized over body weight than the other lifts. What we see with the other lifts however is that, especially when using a Multi-ply suit when compared to Raw, a competitors Wilk's score is a much better predictor of performance in that equipment class. This may suggest that including equipment in the calculation actually augments the Wilk's score to better adjust and normalize it against body weight.

Table 2: Linear Regression Models

	Raw		**Single-Ply**		**Wraps**	
	α	β	α	β	α	β
Squat	-63.89	.66	-37.68	.60	-88.99	.77
Bench	-45.26	.44	-37.97	.42	-50.47	.48
Deadlift	-34.31	.66	29.20	.43	-31.36	.67

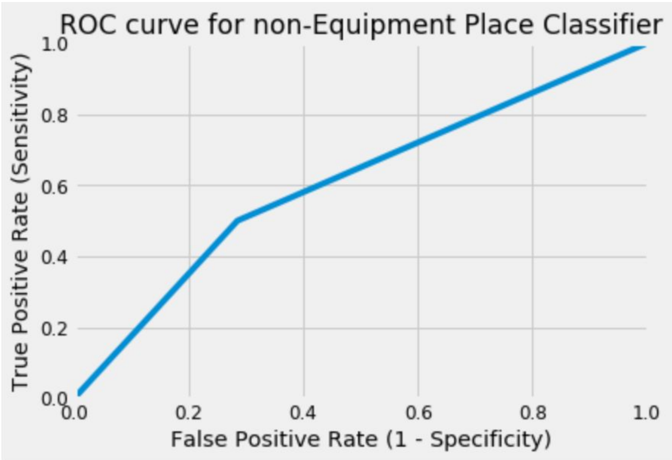
	Raw	Single	Wraps	Multi
Bench	.57	.61	.64	.64
Deadlift	.72	.69	.74	.67
Squat	.70	.75	.77	.80
Total	.76	.76	.83	.89

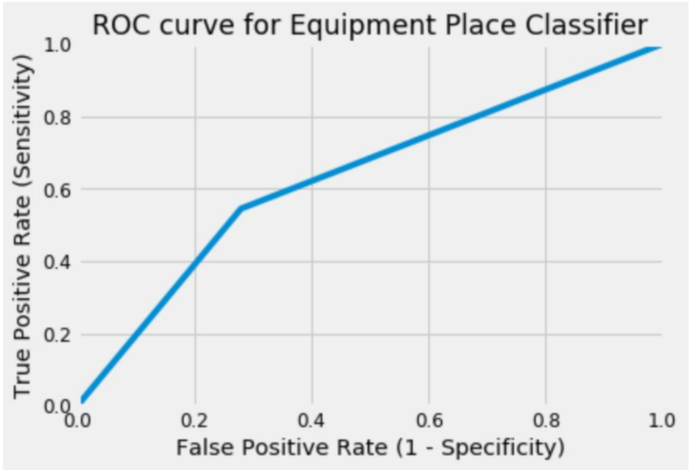
When we first began working with this dataset we were going to analyze the relationship between age and powerlifting performance, however when we began working with the dataset it became clear that age was not a clear indicator of lifting performance. In Figure 4, we can see that a 40 year old lifts just as much as a 20 year old. Since we would not be able to predict age based on performance, we shifted our attention to equipment usage and its effect on performance.

The results for the Random Forest Models are depicted in the figure below.

	Accuracy (R^2)	RMSE	Cross-validated Score
Wilks	.61	.62	.57
Wilks with equipment	.64	.59	.58

What we see here is that there is some marginal gain in performance with equipment and that overall we have a moderate ability to predict whether an athlete will win or not based on this model. To help validate the results we did a 6-fold cross validation on the data and found that it did lower the accuracy values in both models. The ROC curves do show this marginal gain in predictive ability as well, and they are depicted in the figures below.





Limitations and Future Considerations

There are a few limitations to this study and a few things that further study into this question could illuminate. The first thing is that our regression model is incomplete, as the team was not able to complete the modeling of performance using the formulas that the regression model developed. With this information we could drill down into even greater detail into how equipment affects performance. Another limitation may be that Wilk's is not actually the best score to help normalize across different categories of lifter. In analysis done by Peidi Wu, she finds that the Wilk's score may actually have a bias towards heavier lifters. While our inclusion of equipment may help limit this bias as shown by our regression model, it would still be a useful endeavor to analyze other methods as well

Another limitation is that our Random Forest model does not predict place along more than a win or lose binary. While it is possible that equipment usage does not affect performance enough to reliably predict a win, it may help competitors rise through the other places as performance may break down differently depending on place. In addition to this, we could add a more granular analysis of Division to the Random Forest model. One of the reasons we may only see moderate predictive ability with this model is that the actual requirements to win across divisions are different enough that it becomes hard to actually measure the differences.

Appendix

[1] The Wilks Coefficient is a coefficient that measures the strength of a powerlifter against other powerlifters despite the different weight of the lifters. The formula for the coefficient is as follows: $Wilks = 500/(a+bx+cx^2 + dx^3 + ex^4 +fx^5)$ where x = the body weight of the lifter

in kilograms and for men a=-216.0475144, b=16.2606339, c=-0.002388645,d=-0.00113732,e=7.01863E-06,f=-1.291E-08 and for women a=594.31747775582, b=-27.23842536447, c=0.82112226871, d=-0.00930733913, e=4.731582E-05, f=-9.054E-08. This coefficient is then multiplied by the total weight lifted so that we are left with a standard amount lifted that is normalized across all body weights and gender. We are focusing our analysis on the powerlifter's Wilks Score because it reflects the powerlifter's performance better than if we were to simply look at the total amount that they lifted (usapowerlifting.com, 2018).

http://www.usapowerlifting.com/lifters-corner/wilks-formula-for-men-lbs/

Data Cleaning

There are many different reasons why a powerlifting might be disqualified from a powerlifting competition. In the USA Powerlifting Technical Rules, a rulebook that is adapted from the International Powerlifing Federation Technical Rulebook, there are 27 different ways to be disqualified across the events of squat, deadlift, and bench press (http://www.usapowerlifting.com/wp-content/uploads/2014/01/USAPL-Rulebook-2017.pdf Pages 34,36,37). Because of the various reasons that powerlifters can get disqualified, and due to the fact that we are not given the specific reason for each disqualification, we are choosing to omit data where the lifter was disqualified from our analysis. In addition to this, when a powerlifter has been disqualified, their Wilks score was not included in our dataset, and since this is the most important variable the we are analyzing, the data where powerlifters have been disqualified is essentially useless for our purposes.

We will be omitting the rows where the powerlifter did not record a weight lifted for all three parts of the competition (squat, deadlift, bench press). The reason we are doing this is because the variable that we care the most about (Wilks Score) is dependent on Total Kilograms lifted, and if a powerlifter did not perform in all three lifts, their Wilks Score is greatly affected.

References

Wilks Score Formula.

http://www.usapowerlifting.com/lifters-corner/wilks-formula-for-men-lbs/, 2018

Peidi Wu, A Better Wilk's formula, https://peidiwu.com/a-better-wilks-formula/, 2018

Openpowerlifting main repository. https://github.com/sstangl/openpowerlifting

Figure 1: Histogram of Men's equipment usage of Raw, Single-Ply, Multi-Ply Wraps and Straps

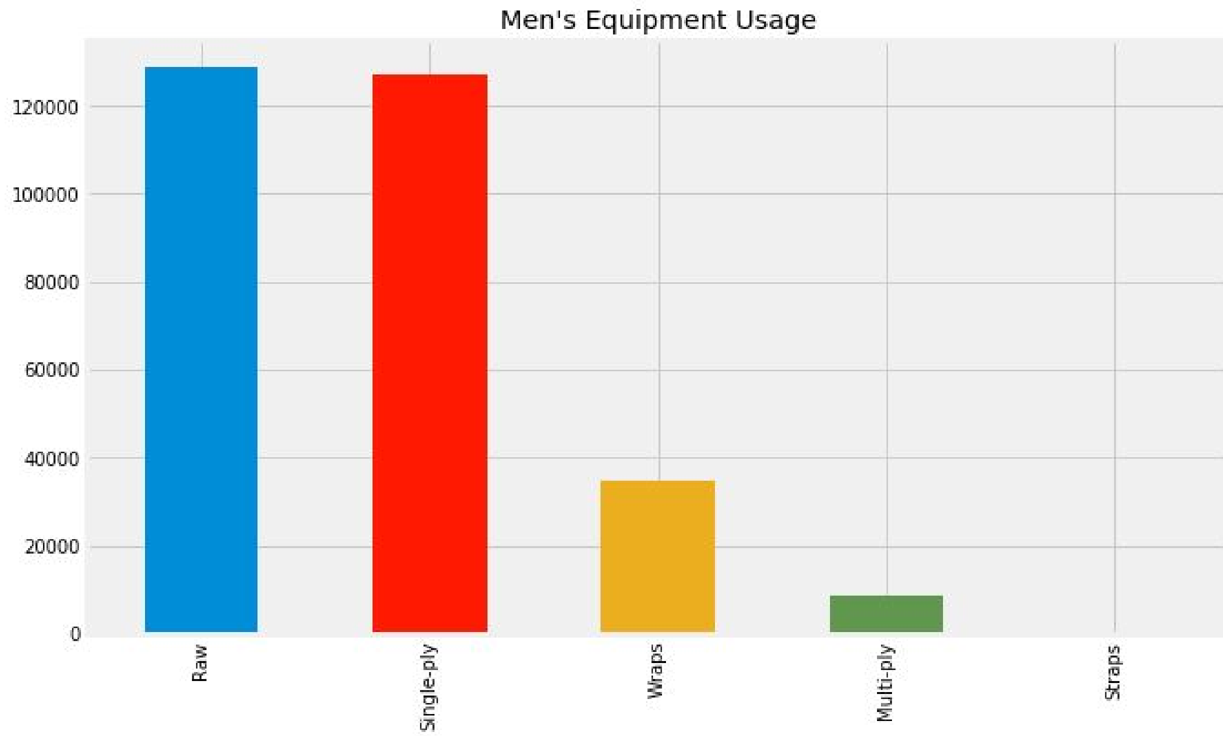
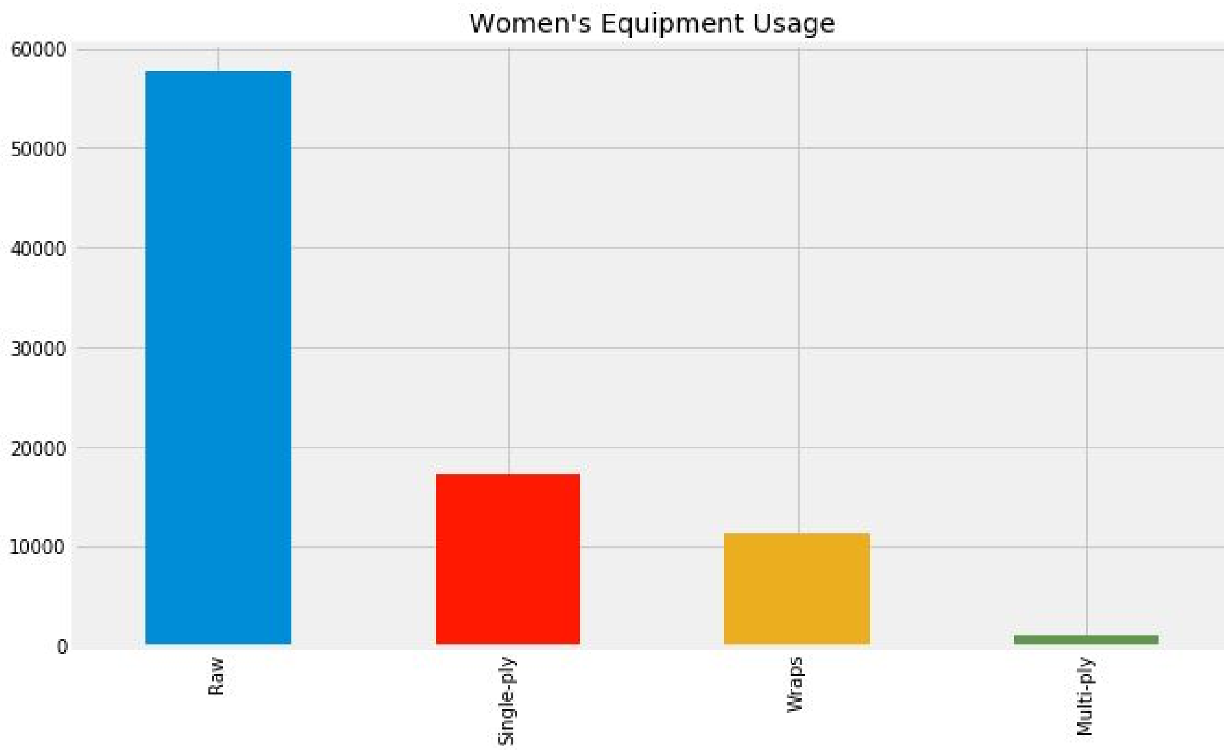
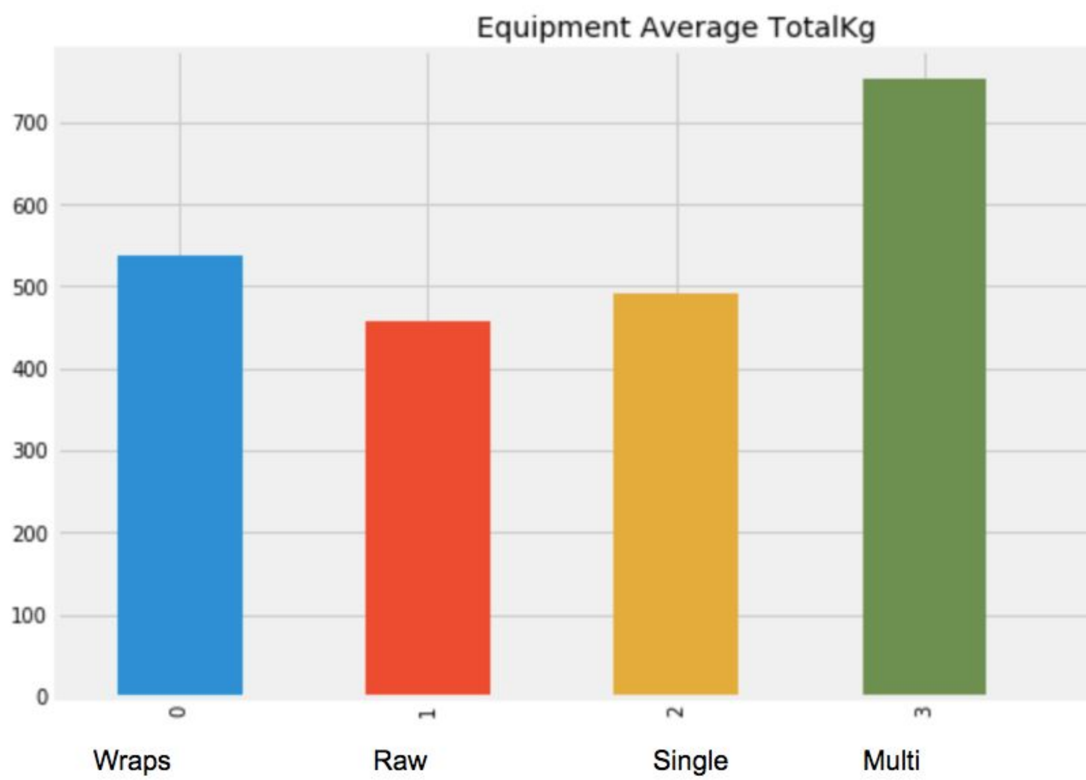


Figure 2: Histogram of Women's equipment usage of Raw, Single-Ply, Multi-Ply and Wraps



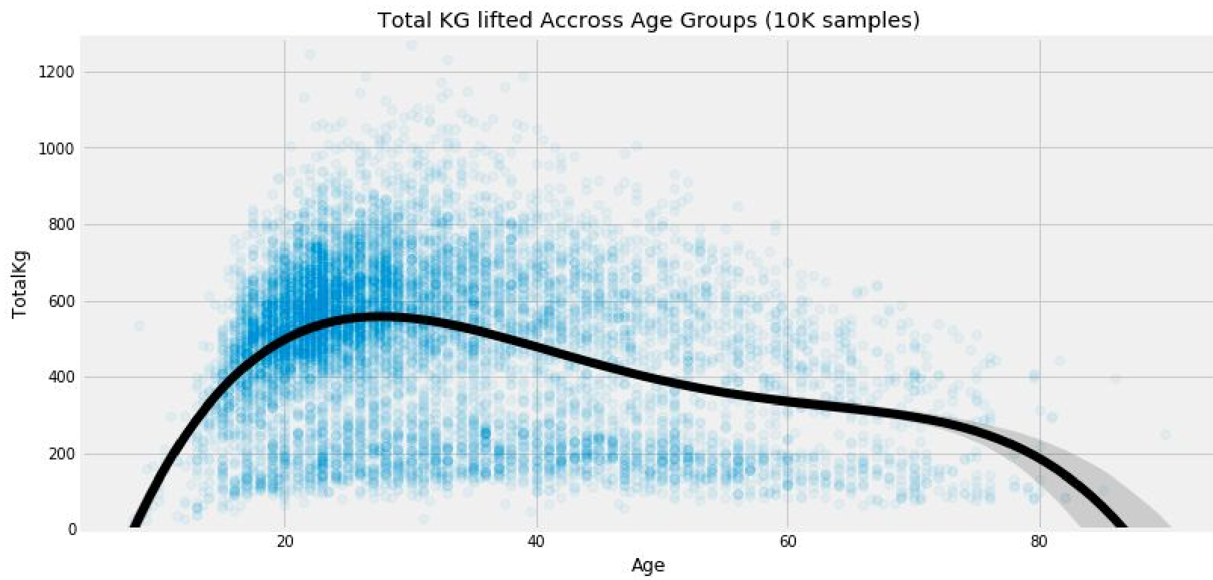
Here, we see how women's equipment usage is far more rare than men's.

Figure 3



Here we see that those who use multi-ply gear tend to lift a lot more, however, this is more indicative of necessity rather than an unfair advantage because multi-ply suits are only used for the heaviest lifts.

Figure 4



Here we see that age is not a great indicator of performance, as those who are 20 and 40 years old tend to lift about the same.