A dark red overlay image of a UFC fight scene. In the center, a referee in a black shirt is raising the right arm of a shirtless male fighter. The fighter has his left arm raised in victory. Another shirtless fighter is visible in the background. The background is a blurred arena with spectators and a banner that reads "UFCSTORE.COM".

DAVID WISMER

UFC SCORE PREDICTIONS

LINEAR REGRESSION

INTRODUCTION

- ▶ Project Goal
 - ▶ Create a multivariate linear regression model to determine the winner of a UFC fight using fight metrics (striking statistics, grappling statistics, control time, etc.).
- ▶ Ultimate Fighting Championship (UFC) Judging
 - ▶ There are 3 to 5 rounds per fight.
 - ▶ For each round, each judge gives 10 points to the winner. The loser receives 9 points, or in some cases, 8 points. Both fighters get 10 points in the event of a tie.
 - ▶ Judges score fights based on a hierarchy of criteria: 1) effective striking and grappling, 2) effective aggression, and 3) control of the fight area.

MODEL DATA

► Features

- ▶ Each feature of the model represents the difference (disparity) between fighter_1 and fighter_2 for a given metric (for example, significant strikes landed becomes ***sig_strike_land_disparity***).
- ▶ The dependent variable, or the per round score for a fighter, is based on all judge and media scoring data available on MMADecisions.com.

► Train Test Split

- ▶ I applied cutoffs that split the data 2,258 fights (4,516 records) into 60% training, 20% validation, and 20% testing data.
- ▶ Using a random split results in fighter_1 and fighter_2 from the same fight sometimes appearing in separate datasets (fighter_1 in training, fighter_2 in testing)
- ▶ There were no significant rule changes or other factors to suggest a time based split is unreasonable. I tested both random and time-based splits, and performance was very similar.

DATA SOURCES AND TOOLS

Data Sources:



Web Scraping:



Data Storage:



Database Interactions:



Data Cleaning:



Regression Model:



Visualization:



WEB SCRAPING - UFCSTATS.COM

1 **Events & Fights**

Enter Event Name...

Completed Upcoming

NAME/DATE	LOCATION
NEXT UFC Fight Night: Rozenstruik vs. Sakai June 05, 2021	Las Vegas, Nevada, USA
UFC Fight Night: Font vs. Garbrandt May 22, 2021	Las Vegas, Nevada, USA
UFC 262: Oliveira vs. Chandler May 15, 2021	Houston, Texas, USA
UFC Fight Night: Rodriguez vs. Waterson May 08, 2021	Las Vegas, Nevada, USA

2 **UFC Fight Night: Font vs. Garbrandt**

DATE: May 22, 2021 LOCATION: Las Vegas, Nevada, USA

Click on a row below to see in-depth event stats. Fight, Perf, Sub, and KO of the Night Bonuses: [#FIGHT](#) [#PERF](#) [#SUB](#) [#KO](#)

W/L	FIGHTER	KD	STR	TD	SUB	WEIGHT CLASS	METHOD	ROUND	TIME
WIN	Rob Font Cody Garbrandt	0	176	2	0	Bantamweight	U-DEC	5	5:00
WIN	Carla Esparza Yan Xiaonan	0	27	3	0	Women's Strawweight	KO/TKO Punches	2	2:58
WIN	Jared Vauderaa Justin Tafa	0	121	0	0	Heavyweight	U-DEC	3	5:00
WIN	Norma Dumont Felicia Spencer	0	68	1	0	Women's Featherweight	S-DEC	3	5:00

3 **UFC Fight Night: Font vs. Garbrandt**

WOMEN'S FEATHERWEIGHT BOUT

L [Felicia Spencer](#)
"FEENOM"

W [Norma Dumont](#)
"THE IMMORTAL"

METHOD: Decision - Split ROUND: 3 TIME: 5:00 TIME FORMAT: 3 Rnd (5-5-5) REFEREE: Chris Tognoni

DETAILS: Sal D'Amato 27 - 30. Brian Miner 29 - 28. Junichiro Kamijo 28 - 29.

SIGNIFICANT STRIKES

FIGHTER	SIG. STR	SIG. STR. %	HEAD	BODY	LEG	DISTANCE	CLINCH	GROUND
Felicia Spencer	47 of 114	41%	27 of 76	14 of 27	6 of 11	38 of 102	9 of 12	0 of 0
Norma Dumont	68 of 135	50%	40 of 100	14 of 20	14 of 15	59 of 123	9 of 12	0 of 0

PER ROUND ▾

ROUND 1
Felicia Spencer 11 of 37 Norma Dumont 21 of 46
ROUND 2
Felicia Spencer 20 of 51 Norma Dumont 35 of 60
ROUND 3
Felicia Spencer 16 of 26 Norma Dumont 12 of 29

WEB SCRAPING - MMADECISIONS.COM

1

EVENTS													
2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008
2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994
Date												#	
May 22, 2021	UFC on ESPN+ 46: Font vs. Garbrandt											8	
May 21, 2021	Bellator 259: Cyborg vs. Smith 2											6	
May 21, 2021	Invicta FC: Rodriguez vs. Torquato											4	
May 15, 2021	UFC 262: Oliveira vs. Chandler											5	
May 08, 2021	UFC on ESPN 24: Rodriguez vs. Waterson											6	
May 07, 2021	Bellator 258: Archuleta vs. Pettis											5	
May 01, 2021	UFC on ESPN 23: Reyes vs. Prochazka											8	
Apr 24, 2021	KSW 60: De Fries vs. Narkun 2											3	
Apr 24, 2021	UFC 261: Usman vs. Masvidal 2											4	
Apr 17, 2021	UFC on ESPN 22: Whittaker vs. Gastelum											8	

2

UFC on ESPN+ 46: Font vs. Garbrandt			
UFC Apex			
Las Vegas, Nevada, USA			
May 22, 2021			
Fight		Scores	
Font def. Garbrandt Unanimous	CLEARY 48 - 47	D'AMATO 50 - 45	KAMIJO 50 - 45
Vanderaa def. Tafa Unanimous	BELL 30 - 27	COLÓN 29 - 28	WEEKS 30 - 27
Dumont def. Spencer Split	D'AMATO 30 - 27	KAMIJO 29 - 28	MINER 28 - 29
Ramos def. Algeo Unanimous	BELL 30 - 27	CLEARY 29 - 28	COLÓN 30 - 27
Hermansson def. Shahbazyan Unanimous	BYRD 29 - 27	D'AMATO 29 - 27	KAMIJO 29 - 27
McGee def. Silva Unanimous	BELL 29 - 27	BYRD 30 - 26	CLEARY 30 - 26
Culibao def. Nuerdanbieke Unanimous	CLEARY 29 - 28	HAGEN 29 - 28	MINER 29 - 28
Ismagulov def. Alves Unanimous	BELL 29 - 28	HAGEN 29 - 28	KAMIJO 29 - 28

3

Norma Dumont
defeats
Felicia Spencer

SPLIT DECISION

UFC on ESPN+ 46: Font vs. Garbrandt
May 22, 2021
Las Vegas, Nevada, USA
REFEREE: Unknown

TALE OF THE TAPE	
Spencer	Dumont
Quebec, Canada	Minas Gerais, Brazil
AGE	30
HEIGHT	5'6"
WEIGHT	145 lbs.
REACH	68"
FIGHTING OUT OF	Belo Horizonte, Minas Gerais, Brazil
Feenom	The Immortal

Sal D'Amato

ROUND	Dumont	Spencer
1	10	9
2	10	9
3	10	9
TOTAL	30	27

Junichiro Kamijo

ROUND	Dumont	Spencer
1	-	-
2	-	-
3	-	-
TOTAL	29	28

Bryan Miner

ROUND	Dumont	Spencer
1	-	-
2	-	-
3	-	-
TOTAL	28	29

MEDIA SCORES

Dayne Fox <i>BloodyElbow.com</i>	29-28	Dumont
Matthew Wells <i>TheBodyLockMMA.com</i>	29-28	Dumont
Ryan Frederick <i>WrestlingObserver.com</i>	29-28	Dumont
Drake Riggs <i>TheBodyLockMMA.com</i>	29-28	Dumont
Rob Tatum <i>CombatPress.com</i>	29-28	Dumont
Marcel Dorff <i>mmdna.nl</i>	29-28	Dumont
Shawn Bitter <i>Cageside Press</i>	29-28	Dumont
MMAJunkie.com	29-28	Dumont
MMANmania.com	29-28	Dumont
Jay Petry <i>Sherdog.com</i>	28-29	Spencer
Ben Duffy <i>Sherdog.com</i>	28-29	Spencer
Tyler Treese <i>Sherdog.com</i>	28-29	Spencer

YOUR SCORECARD

ROUND	Dumont	Spencer
1	-	-
2	-	-
3	-	-
TOTAL	-	-

Show Results

Submit Scorecard

CROSS VALIDATION - R-SQUARED

	Linear Regression	LASSO Regression	Ridge Regression	SGD Regression	Random Forest
Unaltered Data	0.6832	0.6832	0.6832	N/A	0.6955
MinMaxScaler (Normalized)	0.6832	0.6831	0.6832	0.6661	0.6954
StandardScaler (Standardized)	0.6832	0.6832	0.6832	0.6832	0.6958

- ▶ The best R-Squared in the models tested was approximately **0.6958**.
- ▶ Random Forest had the highest R-Squared, but also had the largest generalization error from training to validation. Hyperparameter tuning only partially fixed overfitting.
- ▶ Normalizing with MinMaxScaler and standardizing with StandardizedScaler had minimal impact on performance.
- ▶ SGDRegressor performed differently each iteration. Unaltered data produced nonsensical results.

CROSS VALIDATION - RMSE

	Linear Regression	LASSO Regression	Ridge Regression	SGD Regression	Random Forest
Unaltered Data	0.2025	0.2025	0.2025	N/A	0.1984
MinMaxScaler (Normalized)	0.2025	0.2025	0.2025	0.2075	0.1987
StandardScaler (Standardized)	0.2025	0.2025	0.2025	0.2026	0.1986

- ▶ The minimum RSME in the models tested was approximately **0.1984**.
- ▶ The reasonable actual per round score range is 8.0 - 10.0. The RMSE results range from 9.4% - 10.4% of this range.
- ▶ Normalizing with MinMaxScaler and standardizing with StandardizedScaler had minimal impact on performance.
- ▶ SGDRegressor performed differently each iteration. Unaltered data produced nonsensical results.

SIMPLE VALIDATION - WINNER PREDICTION

	Linear Regression	LASSO Regression	Ridge Regression	SGD Regression	Random Forest
Unaltered Data	85.2%	85.2%	85.2%	47.2%	84.4%
MinMaxScaler (Normalized)	85.2%	85.8%	85.2%	86.6%	84.0%
StandardScaler (Standardized)	85.2%	85.2%	85.2%	85.1%	84.4%

- ▶ The best winner prediction accuracy in the models tested was approximately **86.6%**.
- ▶ SGDRegressor was worse than 50/50 using unaltered data. When data was normalized, it produced the best results of all models tested in simple validation.
- ▶ Normalizing with MinMaxScaler and standardizing with StandardizedScaler had minimal impact on prediction accuracy.
- ▶ Random Forest Regression had the lowest predictive ability, which makes intuitive sense given the established hierarchy of judging criteria.

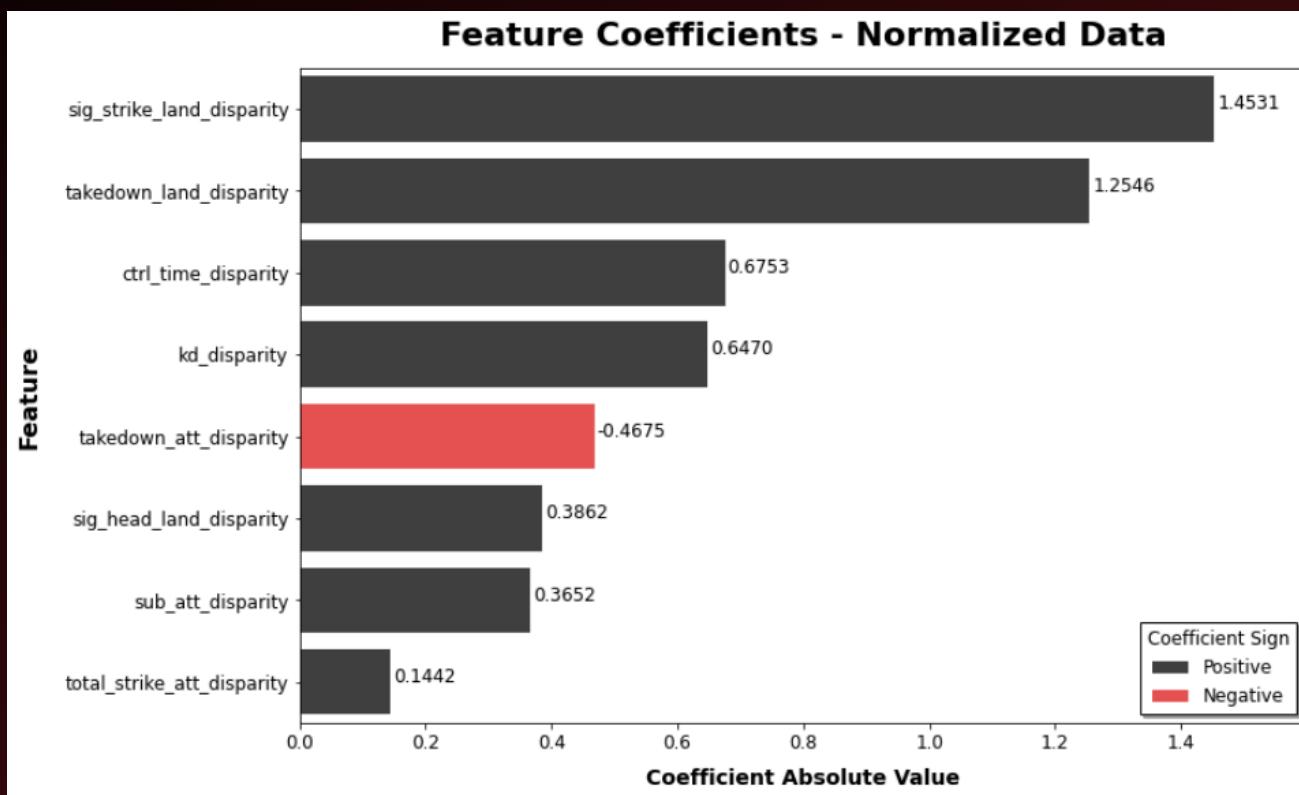
TEST RESULTS - TOP 3 MODELS

	Linear Regression Normalized	SGD Regression Normalized	Random Forest Regression Standardized
R-Squared	0.7168	0.7209	0.7149
RMSE	0.2043	0.2048	0.2050
Prediction Accuracy	85.5%	86.6%	83.7%

- ▶ Predictions over 10.0 were reduced to 10.0, as this is the maximum possible points per round. This produced a **1-2% positive impact** to Linear and SGD R-Squared and RMSE.
- ▶ My preferred regression model is Linear Regression with Normalized Data
- ▶ Linear Regression is easily interpretable and normalized features allow for a more natural comparison of feature coefficients
- ▶ All models performed very similarly in cross validation, simple validation, and in testing
- ▶ Random Forest Regression had strong cross validation results, but even with hyperparameter tuning, the model was overfit to the training data.

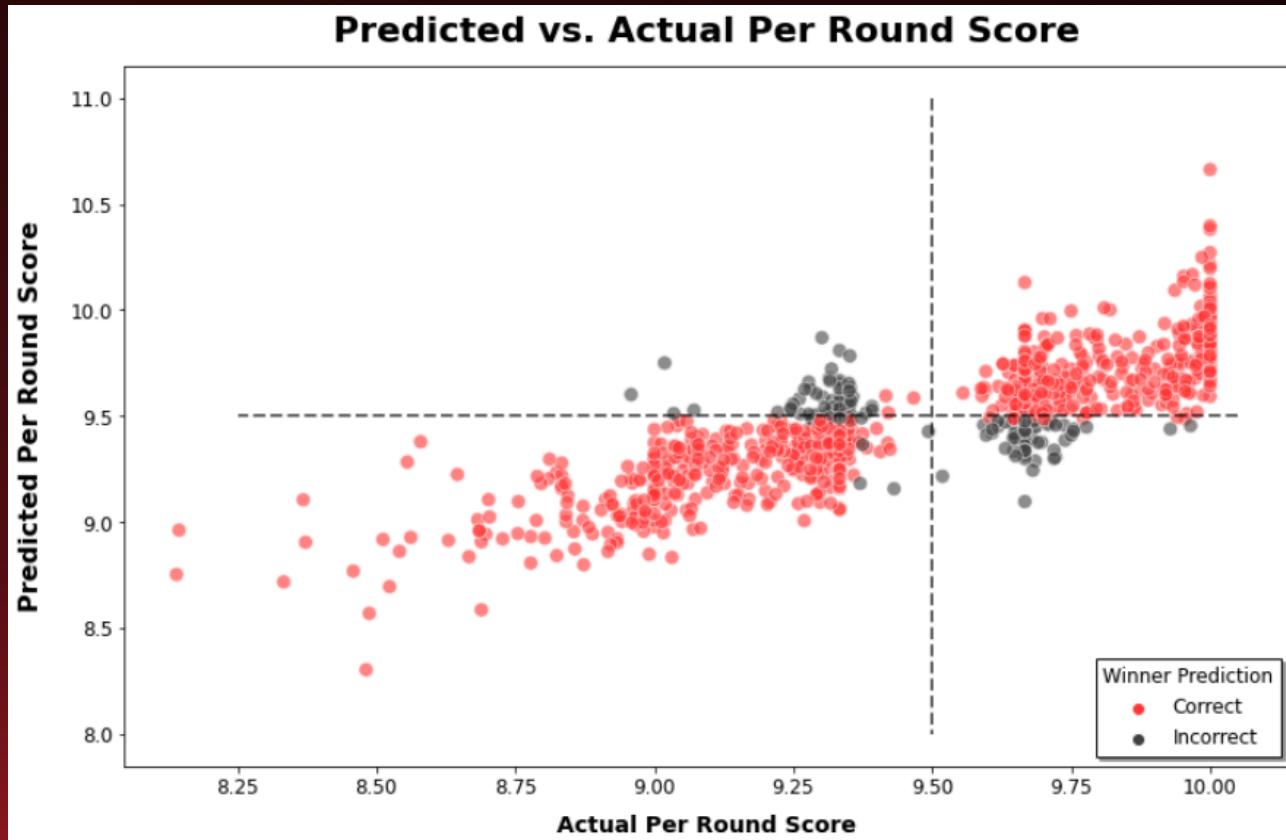
FEATURE IMPORTANCE - LINEAR REGRESSION, NORMALIZED DATA

Per Round Score = $1.4531X_1 + 1.2546X_2 + 0.6753X_3 + 0.6470X_4$
 $- 0.4675X_5 + 0.3865X_6 + 0.3652X_7 + 0.1442X_8 + 7.2543$



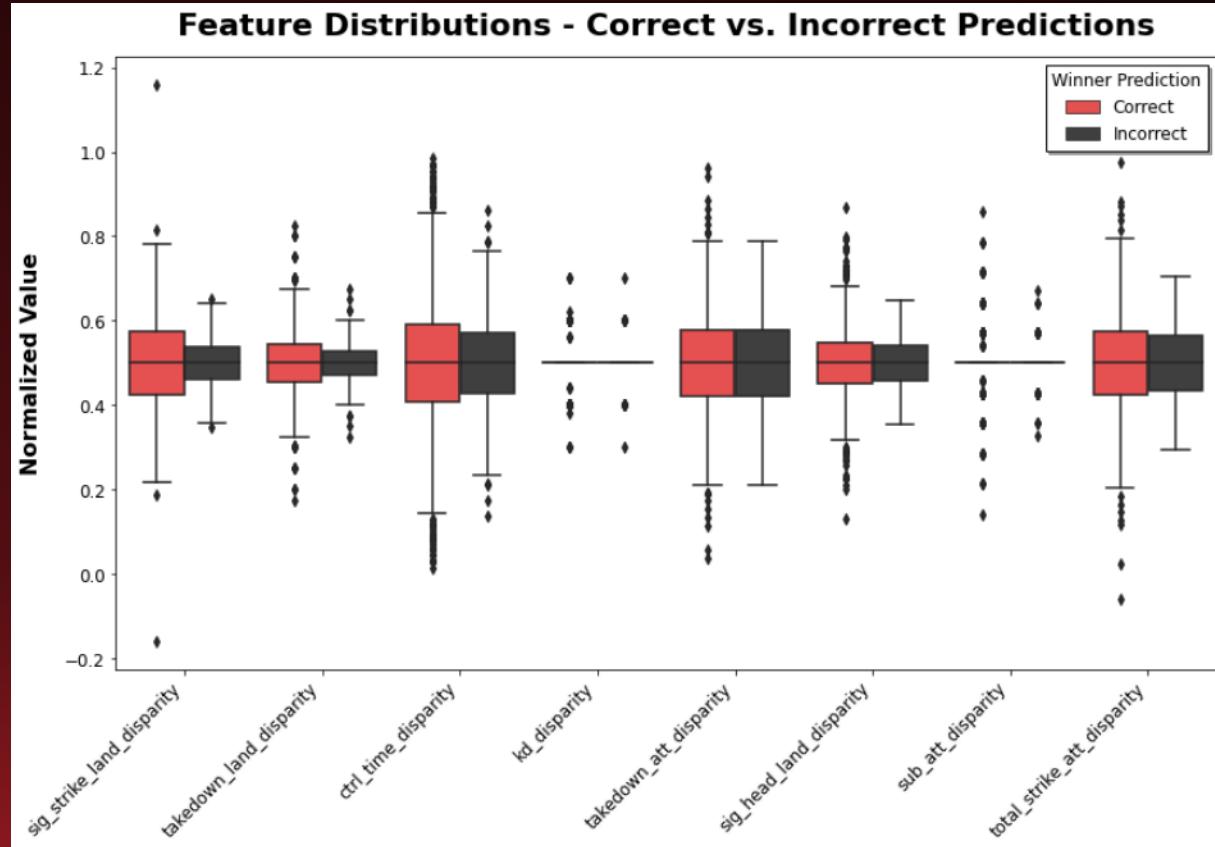
- ▶ The metrics with the greatest impact on predictions are in line with intended UFC judging criteria:
 - ▶ 1) effective striking and grappling
 - ▶ 2) aggression
 - ▶ 3) fight area control
- ▶ The error term for this model is **7.2543** points per round.
- ▶ Takedown attempts have a negative impact on projected score. For every takedown attempt, the fighter is penalized. For each successful takedown, the reward exceeds the penalty.

WINNER PREDICTION - LINEAR REGRESSION, NORMALIZED DATA



- ▶ Winner Prediction Accuracy: **85.5%**
- ▶ Typically, the model is correct when both predicted and actual per round score fall on the same side of 9.5 points per round.
- ▶ The model predicts the winner more accurately the further the actual score gets from 9.5 points per round.

WHERE STATISTICS FALL SHORT

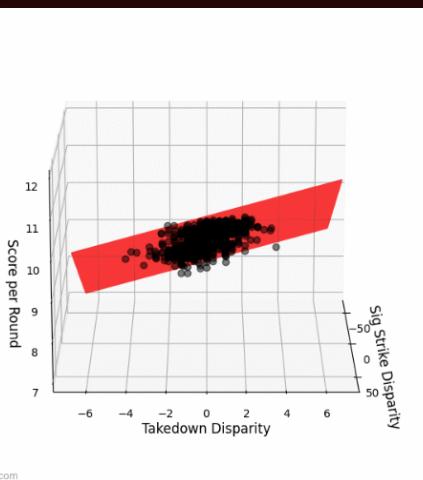
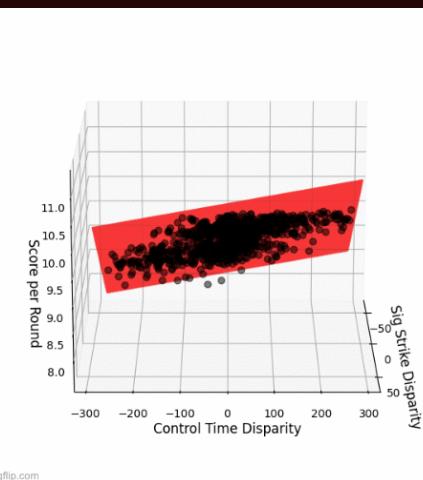
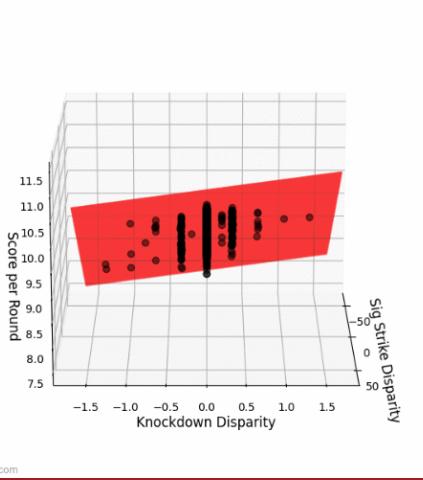
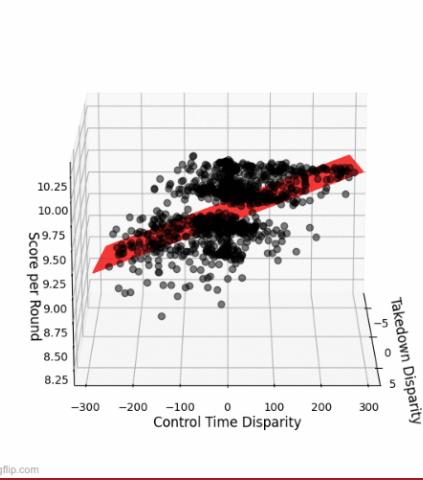


- ▶ Fight metric distributions were narrower in fights where the model made the wrong prediction. Closer fights are harder to judge.
- ▶ Statistics struggle to capture aggression and damage inflicted by fighters, both key factors in judging close fights.
- ▶ Knockdowns are much more rare than, for example, significant strikes. Any knockdown disparity is rewarded heavily, as even the 3rd quartile for knockdown disparity per round is zero. Knockdown disparity has the 4th largest coefficient in the regression model.

WHAT'S NEXT?

- ▶ Model Enhancements
 - ▶ Logistic Regression on a per round basis
 - ▶ Inclusion of point deductions (all fights with deductions are currently excluded from the data)
 - ▶ Additional hyperparameter tuning, particularly for random forest and SGD
- ▶ Analysis of Rule Changes
 - ▶ 10-8 round scoring criteria, only adopted by certain fight commissions
- ▶ Judging the Judges
 - ▶ Which judges historically deviate furthest from the regression model?

THANK YOU!

Feature 1	Sig Strike Disparity		Feature 1	Sig Strike Disparity
Feature 2	Takedown Disparity		Feature 2	Control Time Disparity
R-Squared	0.6190		R-Squared	0.6624
Feature 1	Sig Strike Disparity		Feature 1	Takedown Disparity
Feature 2	Knockdown Disparity		Feature 2	Control Time Disparity
R-Squared	0.5071		R-Squared	0.3283