

COLORADO STATE SOFTBALL PITCHING (FALL 2023)

Danielle Contreras

Advisor: Dr. Aaron Nielsen

RESEARCH QUESTION

- How does **release height**(numeric), **spin efficiency**(numeric), and **spin direction**(factor) affect the end result of a pitch (**ball or strike**), by **pitch type**(factor)?
- And can we use this to predict future pitches?

DATA I

- Collected from Rapsodo:

- Radar Technology that produces instant data
- Data from the Pitching 2.0 Machine
- Collected over the time period of **September 28, 2022 – September 14, 2023**
- 4 pitchers (**Danielle, Delaney, Giselle, and Sydney**)
- 5 Pitch Types (**Changeup, Curveball, Dropball, Fastball, and Riser**)

3,311 Pitches all together for our data (**Changeup, Curveball, Dropball, Fastball, and Riser**):

- Danielle: **834**
- Delaney: **943**
- Giselle: **847**
- Reagan: **154**
- Sydney: **687**
- Changeup: **484**
- Dropball: **1309**
- Riser: **537**
- Curveball: **521**
- Fastball: **460**

DATA 2

CSU Softball Averages

Pitch Type	Strike	Avg. Release Height	Avg. Spin Efficiency	Avg. Side	Avg. Height	Mode Spin Direction
Dropball	N	1.87	68.97	0.85	22.31	7
Dropball	Y	1.85	68.80	-0.63	27.37	7
Curveball	N	1.82	45.87	-6.12	31.55	11
Curveball	Y	1.83	48.81	-0.69	30.31	11
Riser	N	1.77	48.23	4.22	38.71	7
Riser	Y	1.77	44.48	1.37	31.60	7
Changeup	N	1.80	70.54	-0.22	24.65	7
Changeup	Y	1.80	71.37	0.40	28.80	7
Fastball	N	1.76	58.31	3.45	34.81	7
Fastball	Y	1.77	59.89	0.18	30.26	7

CSU Softball Strike Percentages

Pitch Type	Strikes	Balls	Total	Strike Percentage
Dropball	411	898	1309	31.40
Curveball	161	360	521	30.90
Riser	156	381	537	29.05
Changeup	158	326	484	32.64
Fastball	136	324	460	29.57

METHODS

To discover how **release height, spin direction, and spin efficiency** affect the end result of a **pitch (Strike or Ball)**, by **pitch type**, I used a few methods

1. Visualizations and Summary Statistics

2. Modeling:

- Multivariate Regression
- Random Forest
- KNN

(Did not combine Regan's data due to low count and left handed pitcher)

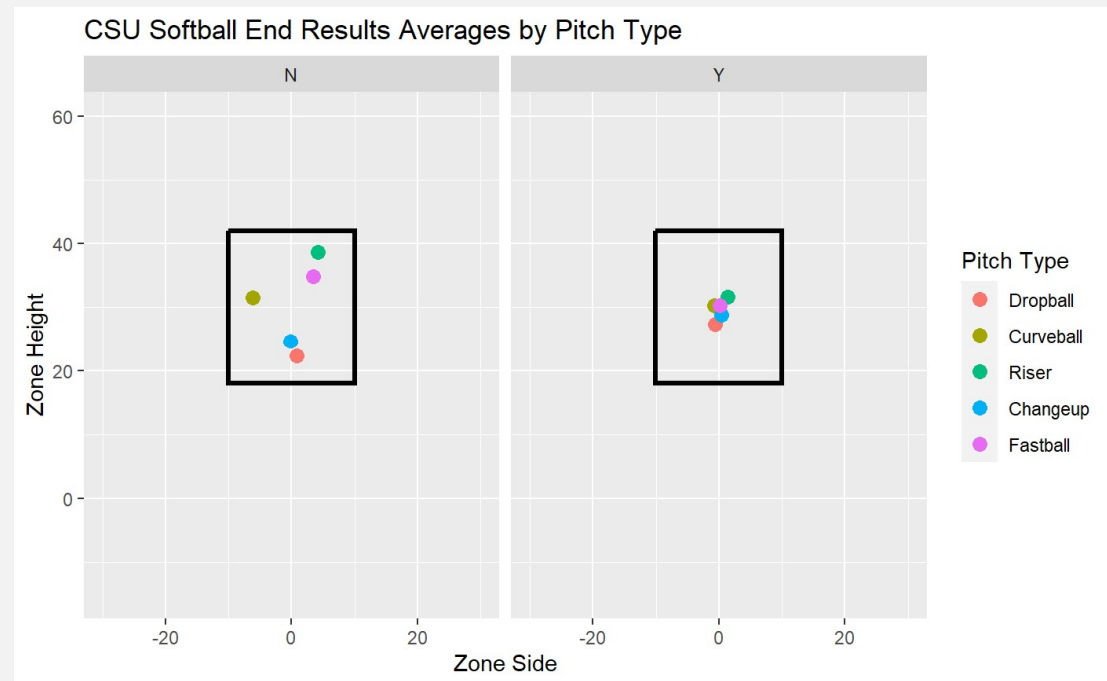
RESPONSE VARIABLES

For Multivariate model:

- Strike Zone Height (Numeric)
- Strike Zone Side (Numeric)

For Random Forest and KNN:

- Strike (Yes or No)



RELEASE HEIGHT

- How many feet off the ground the ball is released

Release Height CSU Softball

Pitch Type	Strike	Mean	Max	Min	Median	Sd
Dropball	N	1.87	2.38	0.50	1.90	0.20
Dropball	Y	1.85	2.36	-0.32	1.87	0.23
Curveball	N	1.82	2.34	0.70	1.85	0.19
Curveball	Y	1.83	2.47	0.64	1.86	0.23
Riser	N	1.77	2.42	0.72	1.80	0.19
Riser	Y	1.77	2.09	-0.40	1.77	0.22
Changeup	N	1.80	2.31	0.54	1.80	0.19
Changeup	Y	1.80	2.28	1.30	1.82	0.18
Fastball	N	1.76	2.46	0.43	1.77	0.20
Fastball	Y	1.77	2.15	1.38	1.75	0.16

SPIN EFFICIENCY

- Useful Spin : total spin (a pitch with all gyro will have spin efficiency of 0)

From Rapsodo Ideal Values:

- Dropball:** 85%-95%
- Curveball:** 55%-65%
- Riser:** 55%-65%
- Changeup:** 60%-70%
- Fastball:** 80%-100%

Spin Efficiency CSU Softball

Pitch Type	Strike	Mean	Max	Min	Median	Sd
Dropball	N	68.97	100.0	6.5	71.25	18.26
Dropball	Y	68.80	99.9	11.1	71.80	18.51
Curveball	N	45.87	100.0	0.2	41.40	24.83
Curveball	Y	48.81	100.0	1.4	44.70	24.43
Riser	N	48.23	100.0	2.6	45.00	23.82
Riser	Y	44.48	100.0	1.7	40.95	26.26
Changeup	N	70.54	100.0	20.3	74.10	18.39
Changeup	Y	71.37	99.9	11.4	73.80	18.36
Fastball	N	58.31	100.0	3.7	55.95	25.59
Fastball	Y	59.89	100.0	2.0	56.50	25.79

SPIN DIRECTION

- The Rotation of spin on a 2D plane from the pitcher's point of view (*Rapsodo*)

- Effects movement of the ball (tilt), while gyro (spin axis along the direction of motion) effects rotation and not movement

- Original Spin direction was measured in format of time (H:M)
- Mutated "H" so that any "M" ≥ 45 rounded up the "H" (11:45 \rightarrow 12)
- Mutated 0 to 12
- All Spin Directions in our modeling and visualizations are in "H" format only

From Rapsodo Video:

- 12 represents backspin (Riser)
- 6 represents topspin (Fastball, Dropball)
- Curveball for righthand: 9
Curveball for lefthand: 3
- Changeup depends on which type of spin but in summary 5-12

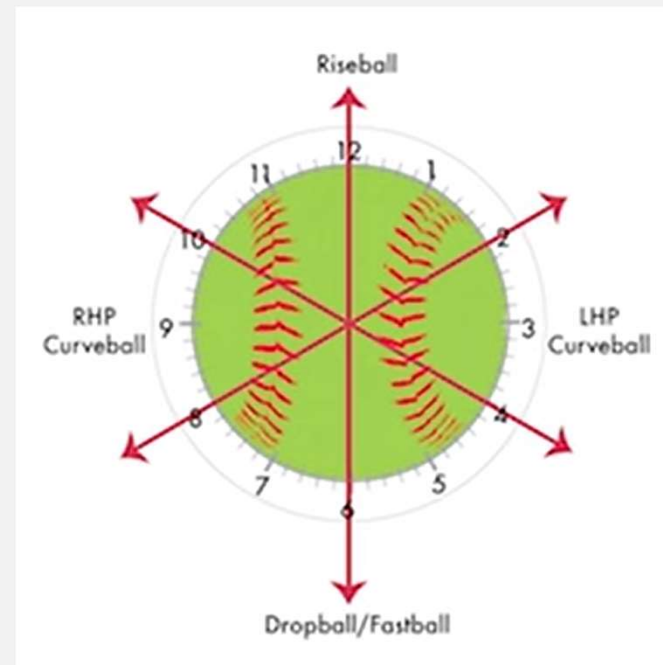
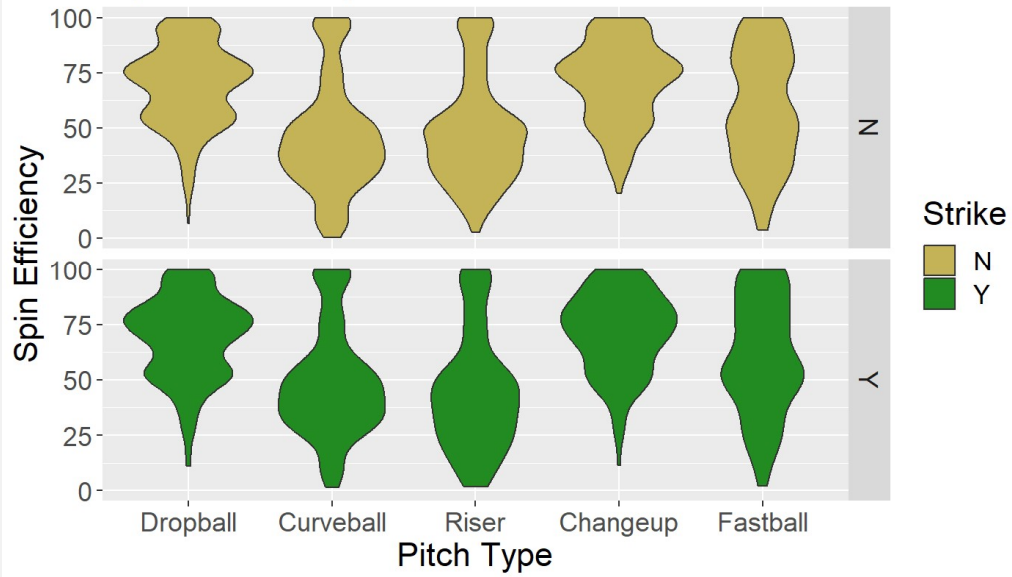


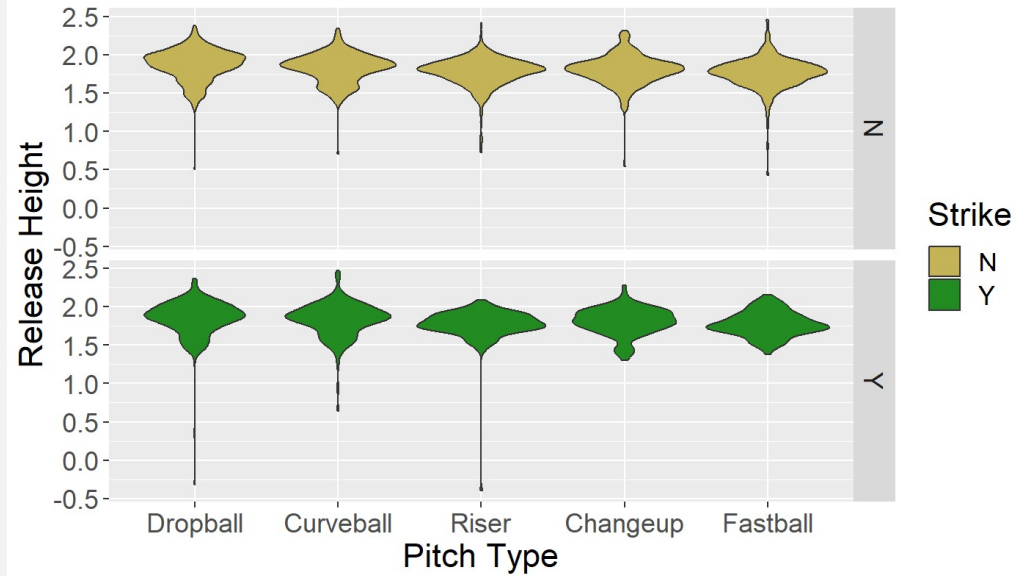
Image from:
Rapsodo Spin Direction
training video (0:30)

VISUALS

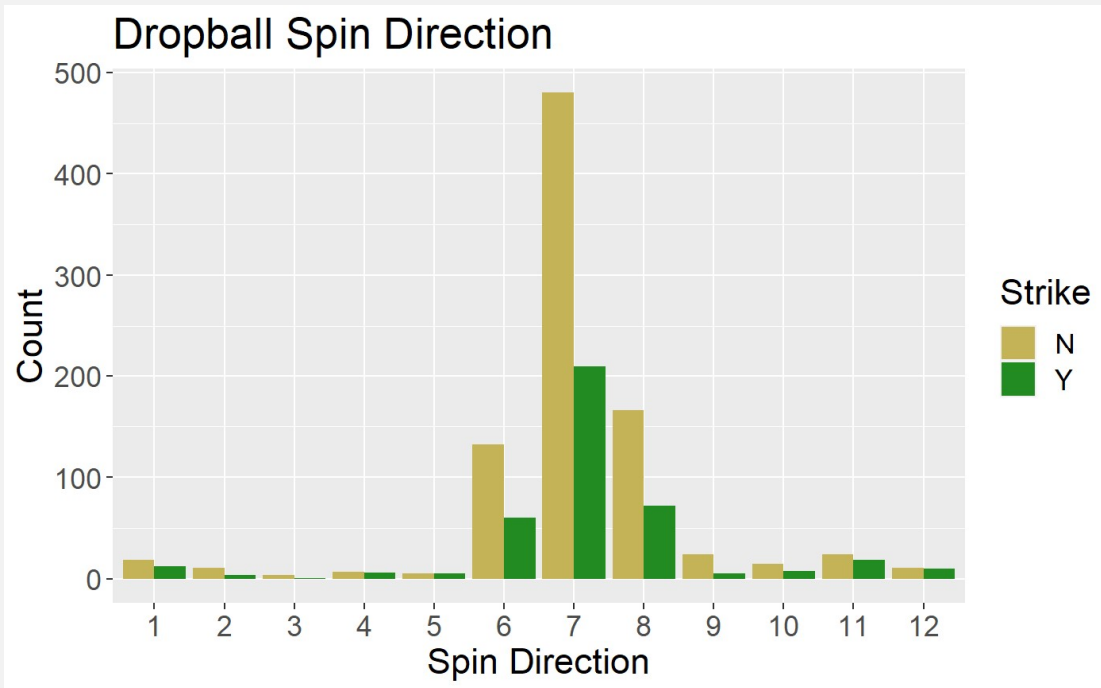
Spin Efficiency All Pitches



Release Height All Pitches



VISUALS 2



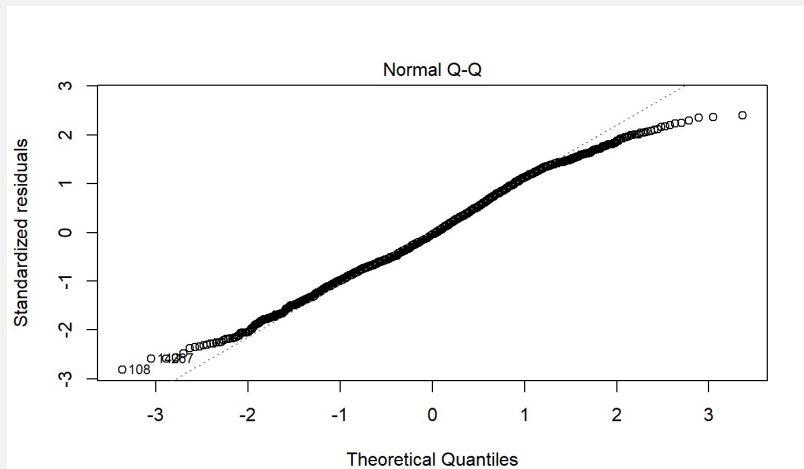
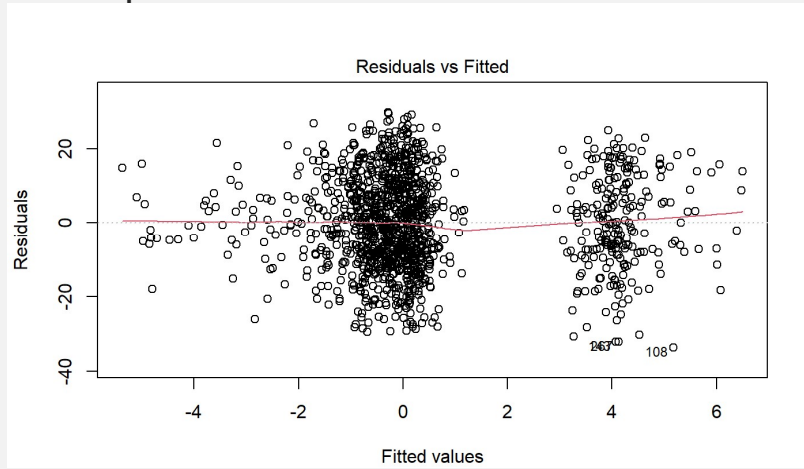
Based off of rapsodo video: 6 is ideal for dropball

Spin Direction: Dropball CSU Softball

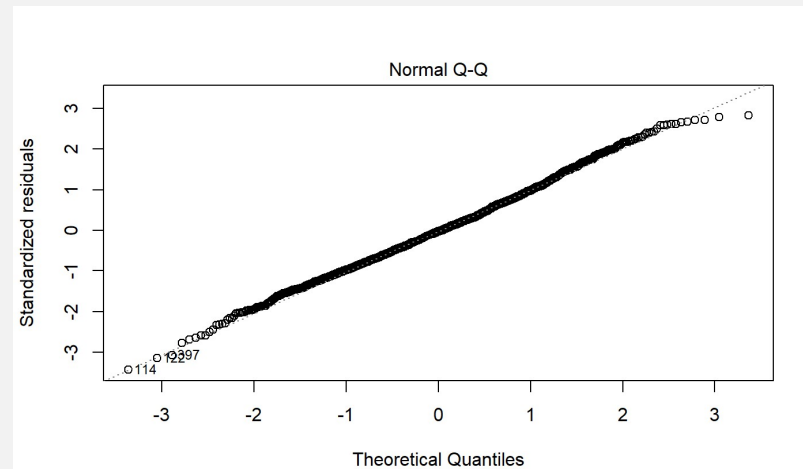
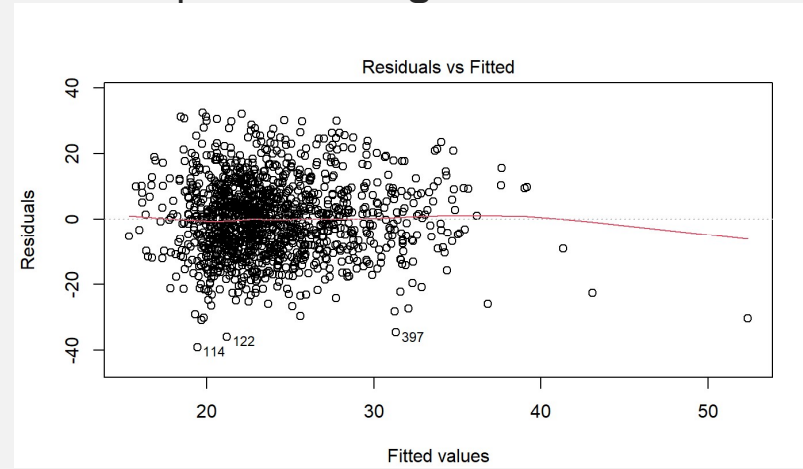
Spin Direction	Strikes	Balls	Total	Strike Percentage
1	12	19	31	39
2	4	11	15	27
3	1	4	5	20
4	6	7	13	46
5	5	5	10	50
6	60	132	192	31
7	209	480	689	30
8	72	166	238	30
9	5	24	29	17
10	8	15	23	35
11	19	24	43	44
12	10	11	21	48

MODELING RESULTS DROPBALL I (MULTIVARIATE)

- Assumptions for **Side**



- Assumptions for **Height**



MODELING RESULTS DROPBALL 2 (MULTIVARIATE)

- None of the models using multivariate analysis were well fit for inference and predictions. (Only Showing results for **Dropball**)
- **Release Height** impacting **Strike Zone Height (Changeup, Dropball, Fastball, and Riser)**

Multivariate Model Dropball CSU Softball Side

response	term	estimate	std.error	statistic	p.value
S. Zone Side	Intercept	-5.49	4.02	-1.37	0.172
S. Zone Side	Spin Efficiency	0.01	0.02	0.62	0.534
S. Zone Side	Release Height	2.18	1.66	1.31	0.190
S. Zone Side	Spin Direction 2	6.41	3.93	1.63	0.103
S. Zone Side	Spin Direction 3	5.85	6.01	0.97	0.331
S. Zone Side	Spin Direction 4	5.53	4.11	1.35	0.179
S. Zone Side	Spin Direction 5	5.55	4.53	1.23	0.220
S. Zone Side	Spin Direction 6	4.60	2.41	1.91	0.057
S. Zone Side	Spin Direction 7	0.51	2.30	0.22	0.824
S. Zone Side	Spin Direction 8	-0.13	2.40	-0.05	0.957
S. Zone Side	Spin Direction 9	-0.59	3.23	-0.18	0.856
S. Zone Side	Spin Direction 10	-3.19	3.45	-0.93	0.355
S. Zone Side	Spin Direction 11	-0.23	2.95	-0.08	0.939
S. Zone Side	Spin Direction 12	-1.91	3.52	-0.54	0.587

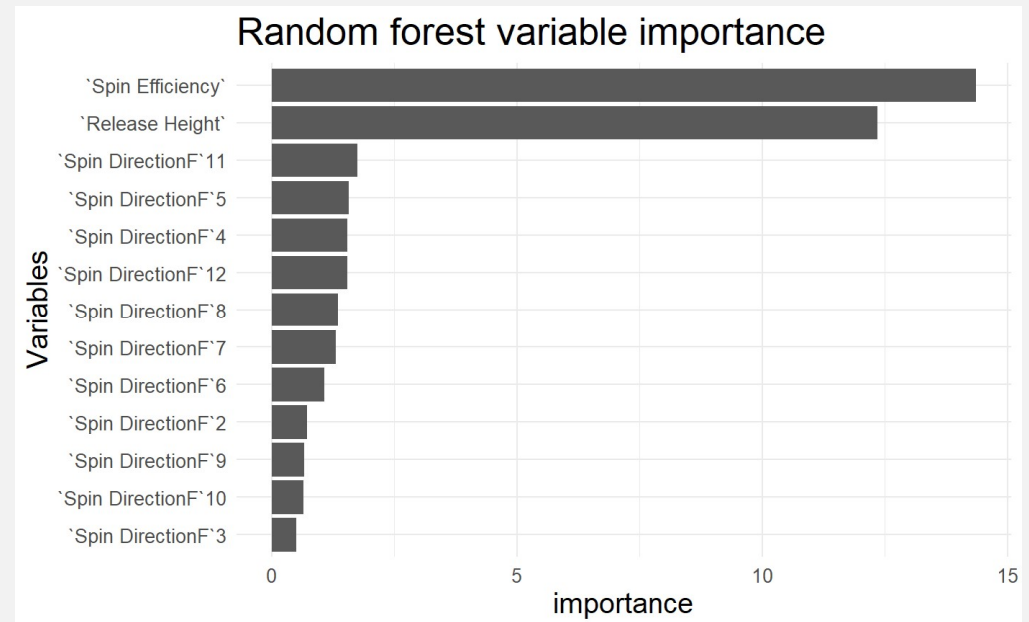
Multivariate Model Dropball CSU Softball Height

response	term	estimate	std.error	statistic	p.value
S. Zone Height	Intercept	56.60	3.71	15.24	<0.001
S. Zone Height	Spin Efficiency	-0.01	0.02	-0.33	0.743
S. Zone Height	Release Height	-13.45	1.53	-8.78	<0.001
S. Zone Height	Spin Direction 2	3.31	3.63	0.91	0.362
S. Zone Height	Spin Direction 3	-0.20	5.55	-0.04	0.971
S. Zone Height	Spin Direction 4	-9.82	3.80	-2.59	0.010
S. Zone Height	Spin Direction 5	-1.46	4.18	-0.35	0.727
S. Zone Height	Spin Direction 6	-6.56	2.23	-2.94	0.003
S. Zone Height	Spin Direction 7	-8.25	2.12	-3.89	<0.001
S. Zone Height	Spin Direction 8	-8.72	2.21	-3.94	<0.001
S. Zone Height	Spin Direction 9	-2.95	2.99	-0.99	0.324
S. Zone Height	Spin Direction 10	-8.07	3.19	-2.53	0.011
S. Zone Height	Spin Direction 11	-2.68	2.73	-0.98	0.327
S. Zone Height	Spin Direction 12	-1.10	3.25	-0.34	0.735

MODELING DROPBALL (RANDOM FOREST)

Process:

- Set.seed (2023)
- Used 70% data as training and 30% data as testing
- Repeated Cross Validation with number of folds being 10 and repeated 3 times.



MODELING RESULTS DROPBALL (RANDOM FOREST)

- Not an accurate way to predict pitches
- Only 67.09% Accuracy (test error of 31.73%)
- Correctly predicted **626 balls**, but did not correctly predict **ANY** strikes
- Incorrectly predicted **3 balls (actual result was a strike)** and incorrectly predicted **288 strikes (actual result was a ball)**

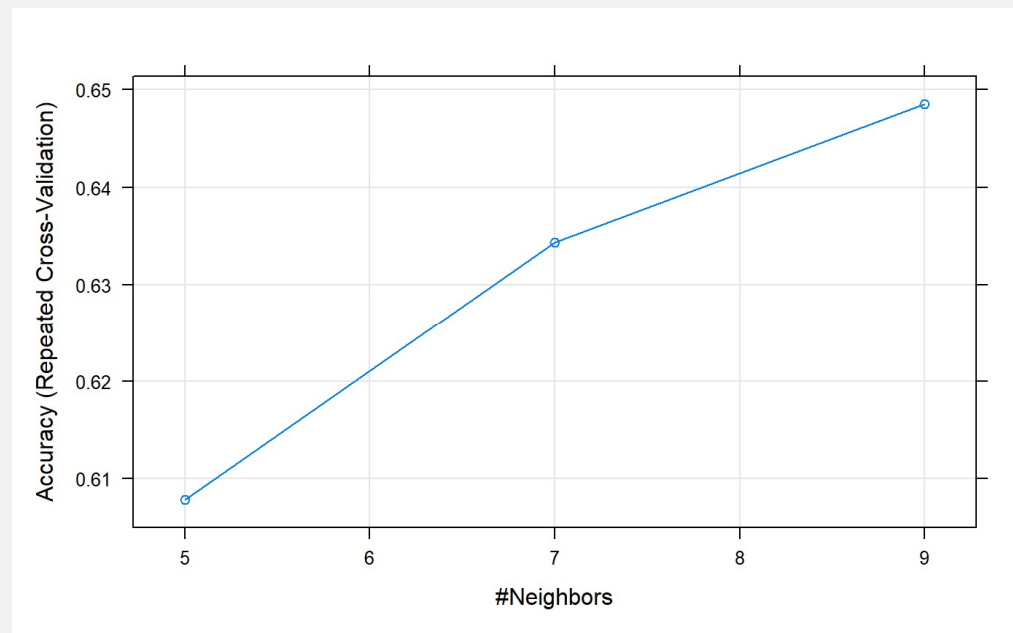
MODELING DROPBALL (KNN)

Process:

- Set.seed (2023)
- Used 70% data as training and 30% data as testing
- Repeated Cross Validation with number of folds being 10 and repeated 3 times.

Results:

- Used $k=9$
 - Not really a good fit to predict
- Note: Positive class is "N"



MODELING RESULTS DROPBALL (KNN)

Dropball KNN Model CSU Softball

	Accuracy	Sensitivity	Specificity
Accuracy	0.63	0.86	0.12

Dropball KNN
Model CSU
Softball

	N	Y
N	231	108
Y	38	15

CONCLUSIONS/CONCERNS I

- The models created were used for interpretability and “easiness” to understand, but they were not good fits
- There could be a trade off between interpretability and the accuracy of the model
- The main goal was not to predict, but was also inference and to see how these three variables impacted the end result of a pitch (Release Height)
- The data was also practice data, could be more accurate if we had live game data.
- Could be discrepancies in the data due to rapsodo. We saw in practice how the data might be incorrectly measured.

CONCLUSIONS 2

CSU Softball Ideal Pitch Values

Pitch Type	Release Height	Spin Efficiency	Spin Direction
Dropball	1.5-2	65-100	6-8
Curveball	1.8-2.2	35-100	8-11
Riser	1.5-2	35-100	1-2, 9-10
Changeup	1.5-2	65-100	5-8
Fastball	1.5-2	50-100	6-8

FUTURE RESEARCH

- Left-handed pitching (Since it is Reagan's first year)
- Release Side impact on `Strike Zone Side` (since `Strike Zone Height` was impacted on most pitch types from Release Height)

BEYOND SOFTBALL

- Even though the models created were not the best fit, there is a lot to take away
- The same methods can be transferred to baseball, especially with rapsodo data.
- MLB data is more accurate(professional level, more money) and has bigger sample size, so these methods may be able to be used at a professional level

RESOURCES

- <https://certified.rapsodo.com/courses/softball-pitching/modules/sbp-glossary/lessons/sbp-spin-direction/> (Picture of Spin Direction)
- <https://certified.rapsodo.com/courses/softball-pitching/> (Rapsodo)