## **Aldehyde Deuteration - Model Search**

### **Table of Contents**

Model Search	1
Section Results Summary	1
1st Place	2
2nd Place	9
3rd Place	
4th Place	23
5th Place	30
6th Place	37
7th Place	44
8th Place	51
9th Place	58
10th Place	65

#### **Model Search**

formula	McFadden R2
class $\sim -2-3- + dip_y + Dist(2, 7) + NPA_1$	0.577
class $\sim$ -2-3- + dip_y + NPA_1 + NPA_3	0.577
class $\sim$ dip_y + NPA_3 + diff 1 2 + L	0.570
class $\sim \text{dip}_y + \text{Dist}(2, 7) + \text{NPA}_3 + \text{diff } 1 2$	0.567
class $\sim$ -2-3- + dip_y + NPA_1 + NPA_2	0.557
class $\sim -2-3- + dip_y + Dist(2, 7) + diff 1 2$	0.555
class $\sim$ -2-7- + dip_y + NPA_1 + NPA_3	0.555
class $\sim$ dip_y + NPA_1 + NPA_3 + diff 2 7	0.553
class $\sim$ dip_y + NPA_3 + NPA_7 + diff 1 2	0.551
class $\sim$ -2-3- + Total + NPA_1 + NPA_3	0.549

#### **Section Results Summary**

	Training Accuracy	Test Accuracy
1. 1st Place	90.48	92.3
2. 2nd Place	85.71	100.0
3. 3rd Place	78.57	92.3
4. 4th Place	85.71	100.0
5.5th Place	80.95	92.3
6. 6th Place	85.71	92.3
7.7th Place	83.33	100.0
8. 8th Place	83.33	100.0

	Training Accuracy	Test Accuracy
9. 9th Place	83.33	100.0
10. 10th Place	85.71	84.6

#### 1st Place

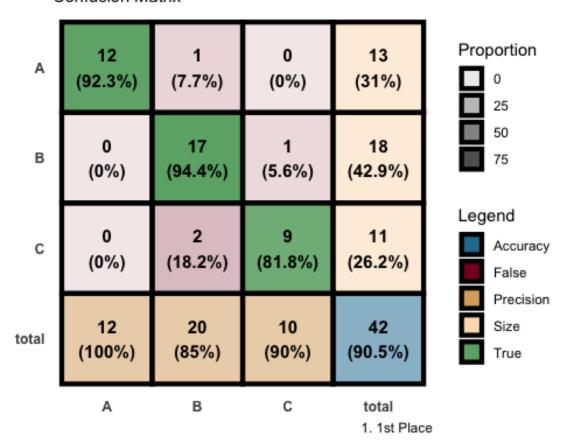
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

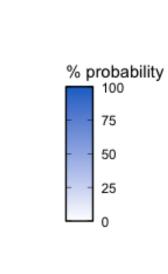
Accuracy	McFadden_R2
90.48%	0.577

	X
-2-3-	-0.3801322
dip_y	-1.5680528
Dist(2, 7)	295.8494658
NPA_1	47.5647077

Training Set Confusion Matrix



	Α	В	С
2A	57	42	1
2G	98	2 2 11	0
2H	98	2	0
2L	88	11	0
2M	89	11	0
20	100	0	0
2R	13 71	80	6
2S	71	28	0
2T	74	25	0
2U	99 82	1 18	0
2V	82	18	0
2J	87	13	0
2P	98	2	0
2Y	98 2 2 1	68	30
2Z	2	65	33
2AA	1	51	48
2AE	10	81	8
2AF	46	53	1
2AG	18	77	4
2AH	16	79	5
2AJ	16 2 6	77 79 62	36
2AK	6	81	13
2AM	7	81	13
2AO	23 27 16	74	
2W	27	70	3 3 5
2X	16	79	
2AB	3	71	26
2AC	5	79	15
2AP	20 22	76 75	4
2AQ	22	75	4
2AR	0	24	76
2AS	1	44	55
2AT	1	44	55
2AU	38	60	2
2AV	1	46	54
2AW	5	79	16
2AX	0	2	98
2AY	0	23	77
2AZ	0	2	98
2BA	0	0	100
2BB	0	0	100
2BC	0	0	100



С

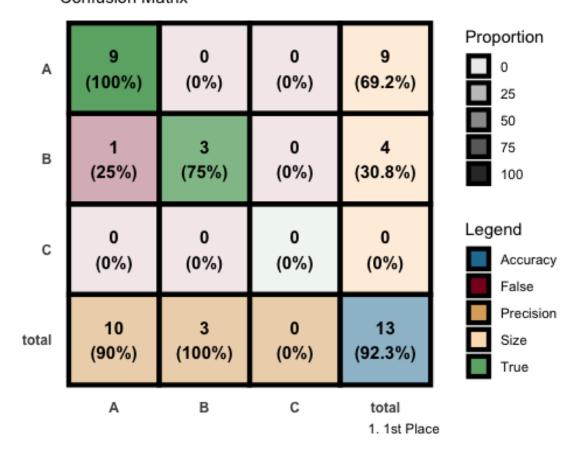
Exp

В

В

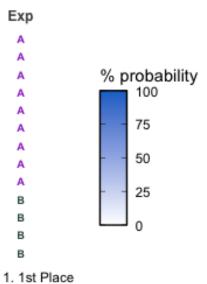
B B B B

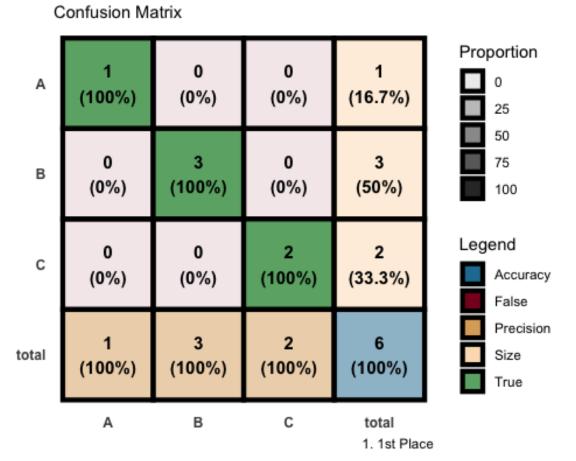
Test Set Confusion Matrix



Test Set Probability Heatmap

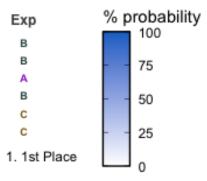
	A	В	C
2B	86	14	0
2C	85	15	0
2D	100	0	0
2E	83	17	0
2F	81	19	0
21	92	8	0
2K	100	0	0
2N	77	23	0
2Q	98	2	0
2AD	29	69	2
2AI	5	78	17
2AL	67	33	1
2AN	10	81	8





## External Validation Probability Heatmap

	A	В	С
2BD	29	69	3
2BE	5	78	17
2BF	74	25	0
2BG	3	69	28
2BH	0	0	100
2BI	1	34	66



#### **2nd Place**

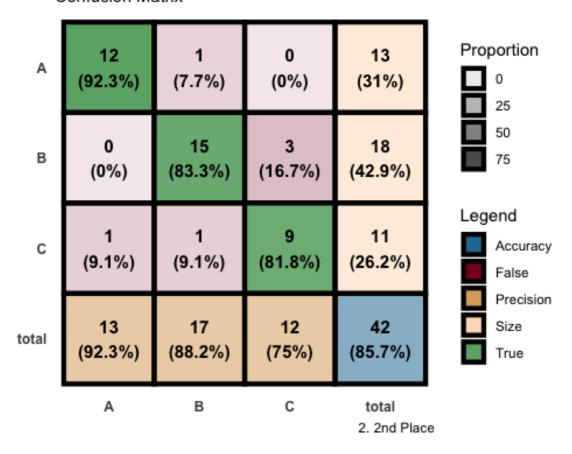
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

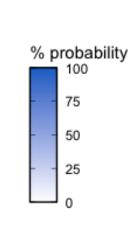
Accuracy	McFadden_R2
85.71%	0.577

	X
-2-3-	-0.1316956
dip_y	-1.2386416
NPA_1	39.1279581
NPA_3	-227.3707397

Training Set Confusion Matrix



	Α	В	С
2A	55	45	1
2G	92	7	0
2H	96	4	0
2L	66	33	0
2M	83	17	0
20	99	1	0
2R	10	17 1 83	6
2S	89	11	0
2T	98	2	0
2U	95	5	0
2V	98	2	0
2J	79	11 2 5 2 21	0
2P	98	2	0
2Y	3	78	18
2Z	1	57	42
2AA	7	41	59
2AE		84	9
2AF	34	65	2
2AG	15	80	4
2AH	15 11	83	6
2AJ	1	40	59
2AK	3	79	18
2AM	11	83	6
2AO	18	79	3
2W	38	61	1
2X	24	74	2
2AB	3	78	18
2AC	11	83	6
2AP	26 17	71	2
2AQ	17	79	4
2AR	1	44	56
2AS	1	49	50
2AT	0	35	65
2AU	54	46	1
2AV	0	34	66
2AW	1	63	35
2AX	0	6	94
2AY	0	37	63
2AZ	0	2	98
2BA	0	0	100
2BB	0	0	100
2BC	0	0	100



С

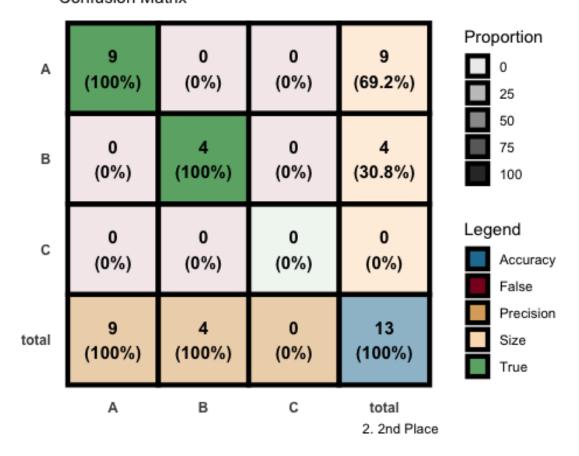
Exp

В

В

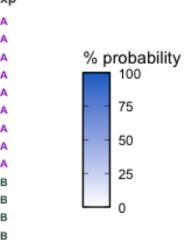
B B B

Test Set Confusion Matrix

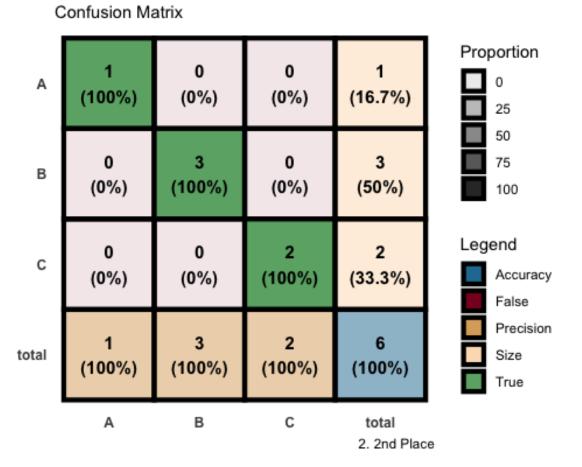


Test Set Probability Heatmap

	Α	В	С	Exp
2B	83	17	0	Α
2C	81	19	0	A
2D	100	0	0	A
2E	78	22 21	0	Α
2F 2I	79	21	0	Α
21	86	14	0	Α
2K	100	0	0	Α
2N	75	25	0	Α
2Q	98	25 2	0	Α
2AD	34	65	2	В
2AI	3	76	21	В
2AL	49	51	1	В
2AN	8	84	8	В



2. 2nd Place



## External Validation Probability Heatmap

2BD

2BE

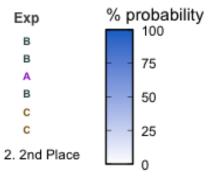
2BF

2BG

2BH

2BI

Α	В	С
19	78	3
2	69	29 1
52 2	47	
2	72	26
0	0	100
0	26	73



#### **3rd Place**

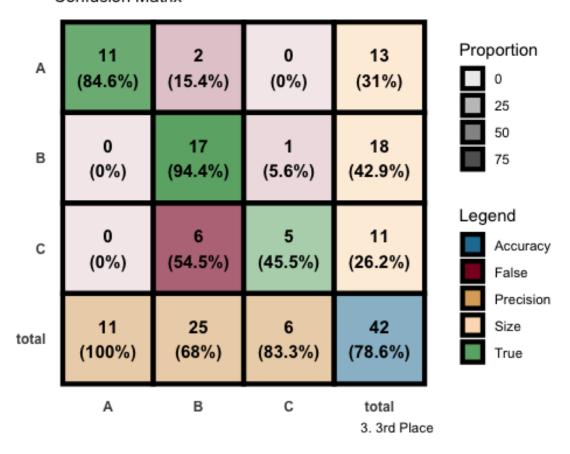
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

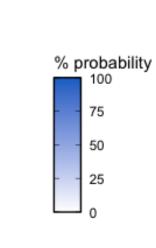
Accuracy	McFadden_R2
78.57%	0.57

	X
dip_y	-1.9662387
NPA_3	-203.6733689
diff 1 2	-39.1768689
L	-0.4500423

Training Set Confusion Matrix



	Α	В	С
2A	32	66	2
2G	90	9	0
2H	90	10	0
2L	70	29	0
2M	88	12	0
20	100	0	0
2R	35	63	2
2S	90	10	0
2T	98	1	0
2U	97	3	0
2V	97	3	0
2J	86	14	0
2P	95	5	0
2Y	1	62	37
2Z	1	51	48
2AA	1 2	50	49
2AE	2	69	28
2AF	16	80	4
2AG	9	83	9 5
2AH	15	80	5
2AJ	1 2	44	55
2AK	2	65	33
2AM	13 43	81	5
2AO	43	56	1 1
2W	21 13	76	3 6
2X	13	81	
2AB	4 2	80	16
2AC	2	64	34
2AP	23	74	3
2AQ	43	56	11
2AR	1	55	44
2AS	11	57	41
2AT	1	51	48
2AU	36	63	2
2AV	1	53	46
2AW	1 2 0	63	36
2AX		2	98
2AY	2	63	35
2AZ	0	2	98
2BA	0	1	99
2BB	0	0	100
2BC	0	1	99



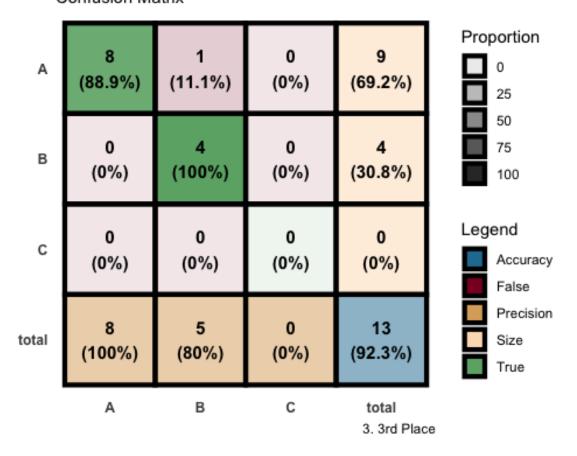
С

Exp

B B

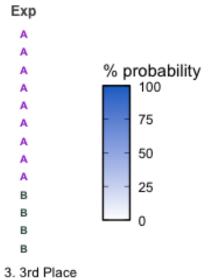
B B B

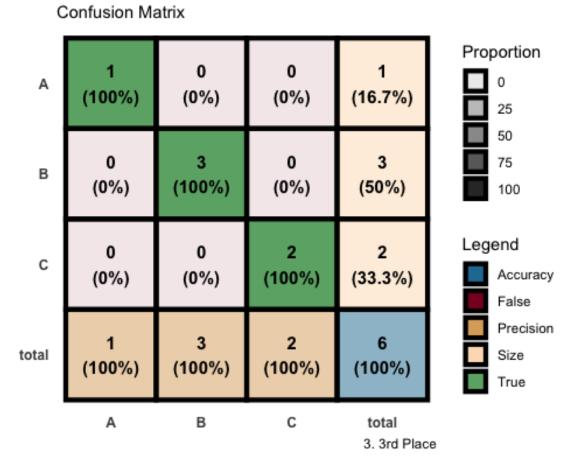
Test Set Confusion Matrix



Test Set Probability Heatmap

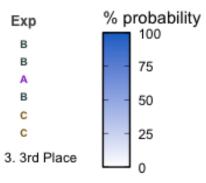
	Α	В	С
2B	57	42	1
2C	52	47	1
2D	99	1	0
2E	43	56	1
2F	43 54	45	1
21	90	9	0
2K	100	0	0
2N	69	30	0
2Q	93	7	0
2AD	19	78	4
2AI	2	71	26
2AL	47	52	1
2AN	6	82	12





Probability Heatmap

	Α	В	С
2BD	6	82	12
2BE	6	82	12
2BF	70	29	0
2BG	4	77	20
2BH	0	0	100
2BI	0	9	91



#### 4th Place

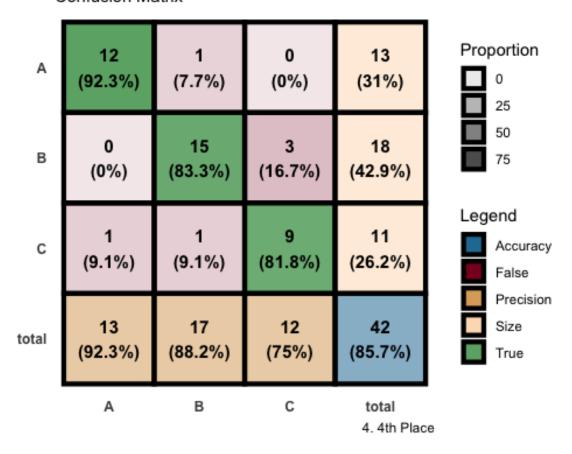
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

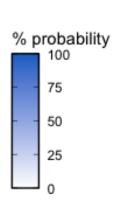
Accuracy McFadden\_R2 85.71% 0.567

	X
dip_y	-1.212812
Dist(2, 7)	-205.902948
NPA_3	-298.027882
diff 1 2	-37.601198

Training Set Confusion Matrix



	Α	В	С
2A	50	49	1
2G	90	10	0
2H	95	5	0
2L	68	32	0
2M	84	15	0
20	99	1	0
2R	10	82	8
2S	88	12	0
2T	99	1	0
2U	94	6	0
2V	99	1	0
2J	82	18	0
2P	99	1	0
2Y	3	73	24
2Z	1	51	48
2AA	1	36	63
2AE	10	82	8
2AF	38	60	2 5
2AG	16	79	5
2AH	13	81	6
2AJ	11	45	54
2AK	5	79	16
2AM	14	80	6
2AO	18	78	4
	36	62	2
	22		3
	5		
			2
	18		50
	1		
	2		
	0	12	
2BC			
2W 2X 2AB 2AC 2AP 2AQ 2AR 2AS 2AT 2AU 2AV 2AV 2AX 2AY 2AZ 2BA 2BB	36 22 5 7 34 18 1 1 0 56 0 2 0 1 0 0	76 62 74 79 81 64 78 48 45 34 44 33 66 12 38 3 0	2 3 16 12 2 4 52 54 66 1 67 31 88 62 97 100 99



С

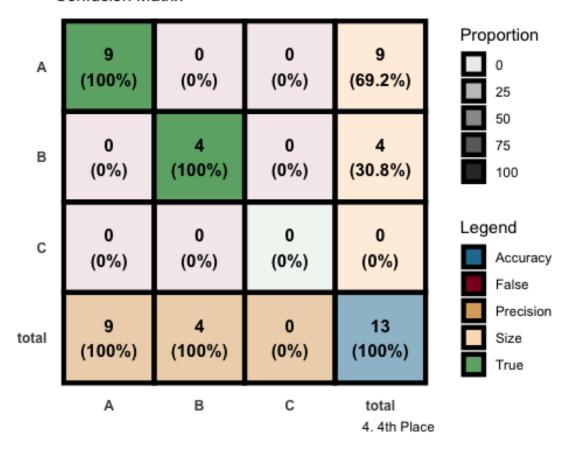
Exp

В

В

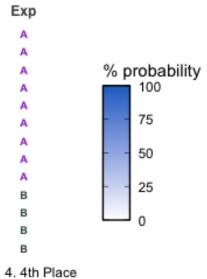
B B B

Test Set Confusion Matrix

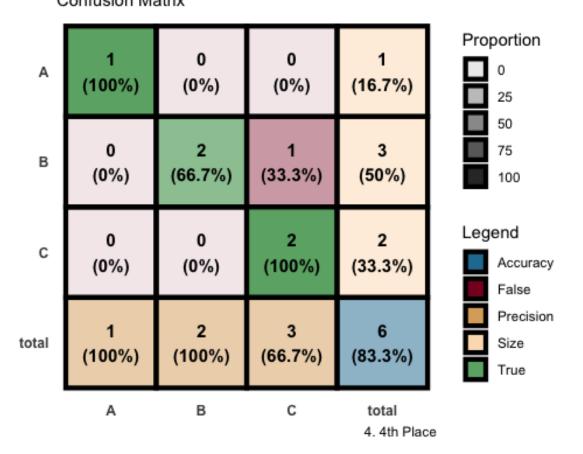


Test Set Probability Heatmap

	Α	В	С	
2B	83	17	0	
2C	81	19	0	
2D	99	1	0	
2E	75	24	0	
2F 2I	80	20	0	
21	87	13	0	
2K	100	0	0	
2N	74	25	0	
2Q	98	2	0	
2AD	29	69	2	
2AI	4	78	18	
2AL	47	52	1	
2AN	9	82	9	

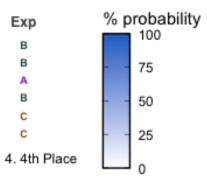


# External Validation Confusion Matrix



## External Validation Probability Heatmap

	A	В	C
2BD	21	75	3
2BE	1	40	59
2BF	54	45	1
2BG	4	77	19
2BH	0	0	100
2BI	0	30	70



#### **5th Place**

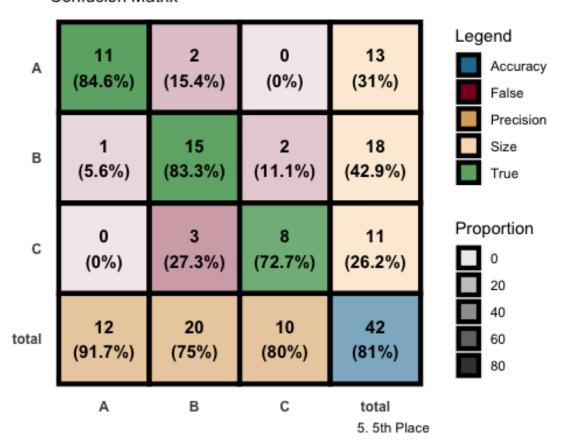
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

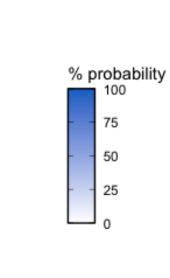
Accuracy	McFadden_R2
80.95%	0.557

	X
-2-3-	-0.2656046
dip_y	-1.5218006
NPA_1	37.0038197
NPA_2	58.0760567

Training Set Confusion Matrix



	Α	В	С
2A	34	63	3
2G	97	3	0
2H	96	4	0
2L	89	11	0
2M	88	12	0
20	99	1	0
2R	11	1 78	10
2S	62	37	1
2T	87	13	0
2U	98	2 7	0
2V	93	7	0
2J	88	12	0
2P	95	5	0
2Y	4	68	28
2Z	2	56	42
2AA	1	46	53 7
2AE	16	77	7
2AF	54	44	1
2AG	19	75	6
2AH	19 21 5 11 7	44 75 74 73 78	5
2AJ	5	73	22 10
2AK	11	78	10
2AM	7	77	17
2AO	23 13	72	5 9
2W	13	78	9
2X	9	78	13
2AB	6	76	18
2AC	3	64	33
2AP	22	73	5
2AQ	23	72	5
2AR	1	30	69
2AS	3	62	35
2AT	1	43	56
2AU	20	74	5
2AV	11	44	55
2AW	12	78	10
2AX	0	5	95
2AY	0	25	75
2AZ	0	10	90
2BA	0	0	100
2BB	0	2	98
2BC	0	1	99



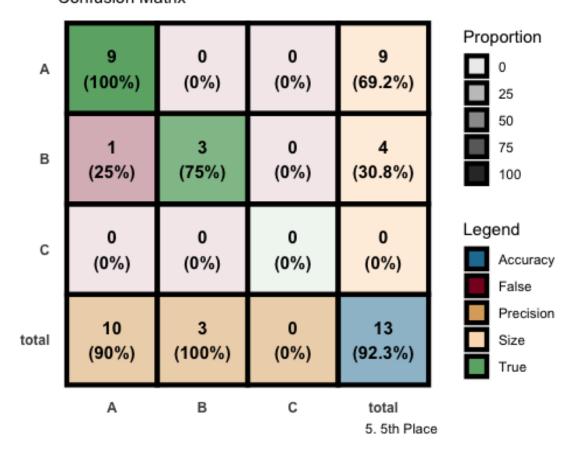
С

Exp

В

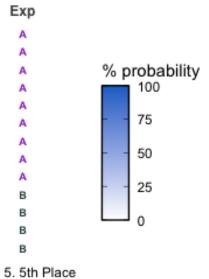
B B B

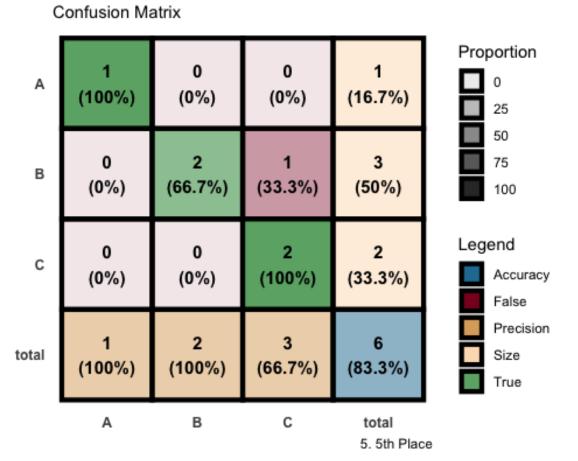
Test Set Confusion Matrix



Test Set Probability Heatmap

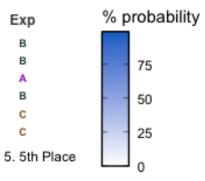
	Α	В	C
2B	81	19	0
2C	81	19	0
2D	99	1	0
2E	79	21	0
2F	76	24 8	0
21	92	8	0
2K	100	0	0
2N	75	25 5	0
2Q	95	5	0
2AD	15	78	8
2AI	9	78	13
2AL	61	38	1
2AN	9	78	13





Probability Heatmap

	A	В	C
2BD	34	63	3
2BE	1	29	70
2BF	75	24	0
2BG	8	78	15
2BH	0	1	99
2BI	1	45	54



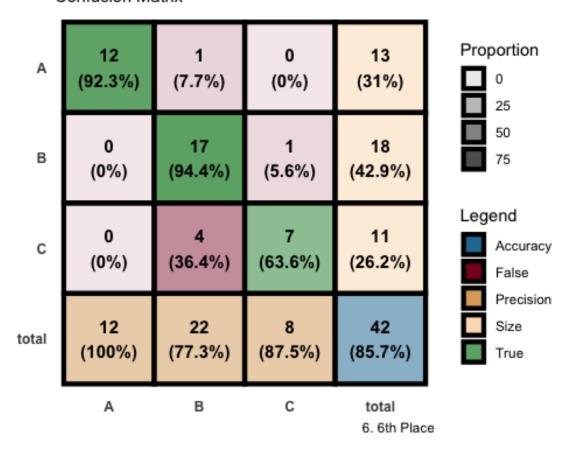
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

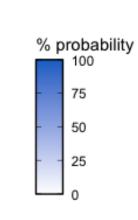
Accuracy	McFadden_R2
85.71%	0.555

	X
-2-3-	-0.298887
dip_y	-1.568930
Dist(2, 7)	243.537174
diff 1 2	-42.690753

Training Set Confusion Matrix



	Α	В	С
2A	53	46	1
2G	98		0
2H	96	3	0
2L	89	11	0
2M	88	12	0
20	99	1	0
2R	12	2 3 11 12 1 77	10
2S	69	30	1
2T	64	35	1
2U	99 76	35 1 24	0
2V 2J	76	24	1
2J	87	12	0
2P	97	3 62 64 56	0
2Y	97 3 3 2 13	62	35
2Z	3	64	33
2AA	2	56	42
2AE	13	77	10 2 7 7
2AF	49	49 75 75 64 77 77 77 73 69	2
2AG	17 18	75	7
2AH	18	75	7
2AJ	3 9	64	33
2AK	9	77	14
2AM	11 22 27 16	77	12
2AO	22	73	6
2W	27	69	4
2X	16	76	8
2AB	5	71 68	24
2AC	4	68	28
2AP	25	71	5
2AQ	21	73	6
2AR	21 1 1	31 45	68
2AS	1	45	53
2AT	2	51	47
2AU	39	59	3
2AV	2	54	44
2AW	9	77	14
2AX	0	3	97
2AY	1	29	70
2AZ	0	4	96
2BA	0	0	100
2BB	0	0	100
2BC	0	0	100



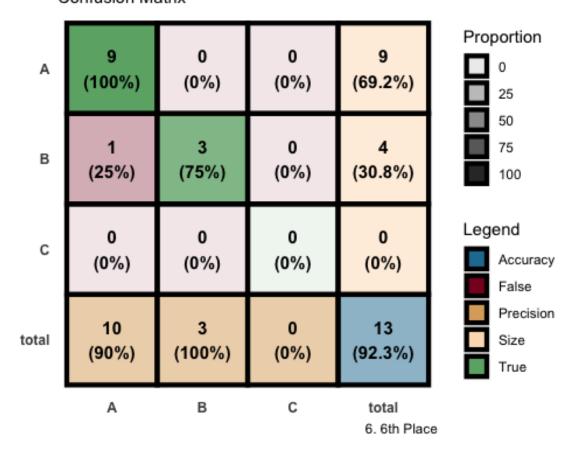
6. 6th Place

C C

В

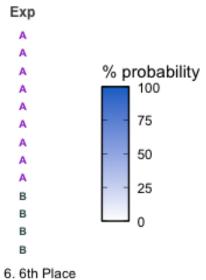
Exp

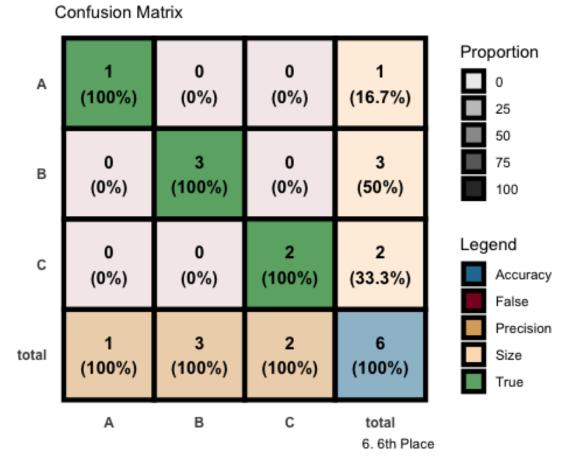
Test Set Confusion Matrix



Test Set Probability Heatmap

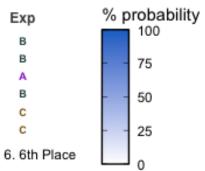
	Α	В	С
2B	83	16	0
2C	82	18	0
2D	100	0	0
2E	78	21	0
2F	78	21	0
21	92	8	0
2K	100	0	0
2N	71	28	1
2Q	96	4	0
2AD	27	68	4
2AI	7	75	18
2AL	64	36	1
2AN	10	77	13





### Probability Heatmap

2BD 31 66 4 2BE 7 75 18 2BF 75 25 1 2BG 4 69 26
2BF <b>75</b> 25 1
380 4 60 36
2BG 4 69 26
2BH 0 0 <b>100</b>
2BI 1 36 63



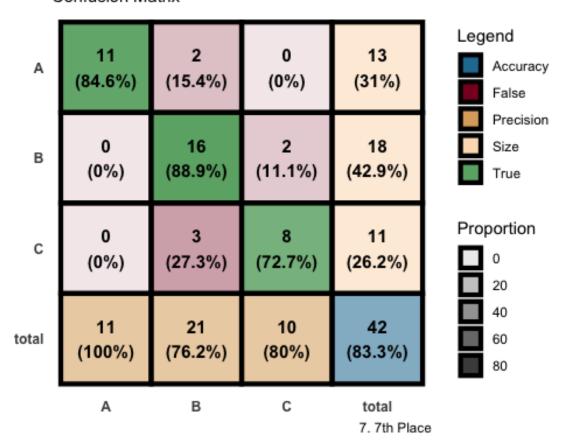
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

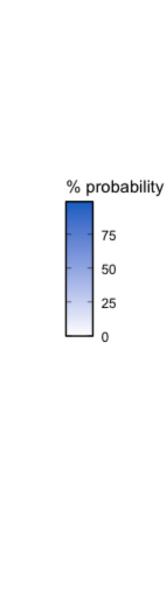
Accuracy	McFadden_R2
83.33%	0.555

	X
-2-7-	0.0196432
dip_y	-1.2374931
NPA_1	34.2137827
NPA 3	-247.0244428

Training Set Confusion Matrix



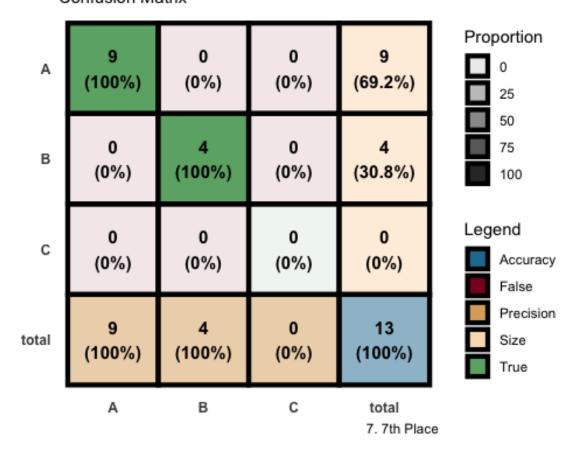
	Α	В	С
2A	47	51	2
2G	88	12	0
2H	94	6	0
2L	71	29	1
2M	81	18	0
20	99	1	0
2R	8	1 78	14
2S	85	14	0
2T	99	14 1 7 1	0
2U	93	7	0
2V	99	1	0
2J	81	18 3	0
2P	97	3	0
2Y	4	68	28
2Z	2	51	47
2AA	1	37 78	62
2AE	10	78	11
2AF	40	58 78	62 11 2 9
2AG	13	78	9
2AH	13 1	78	9
2AJ	1	78 47 75 77 77 66 75 77	52
2AK	6	75	19
2AM	15	77	8
2AO	16	77	7
2W	31	66	3 6
2X	19	75	6
2AB	8	77	16
2AC	6	74	20
2AP	35	63	3 7
2AQ	16 2 2	77	7
2AR	2	49	49
2AS		53	46
2AT	1	39	60
2AU	48	51	2
2AV	11	38	61
2AW	4	68	28
2AX	0	13	87
2AY	1	41	58
2AZ	0	4	95
2BA	0	2 7	98
2BB	0	/	93
2BC	0	2	98



Exp

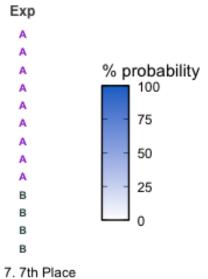
С

Test Set Confusion Matrix

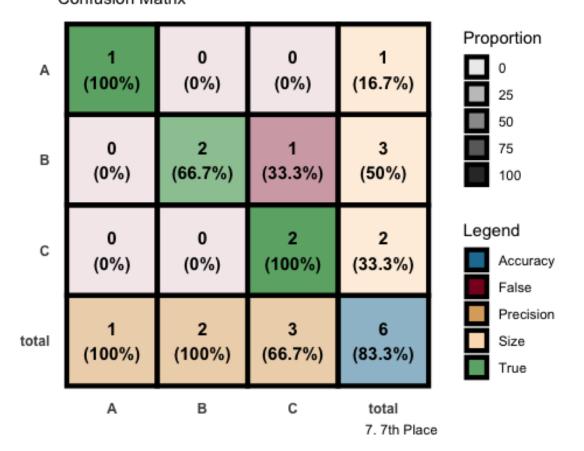


Test Set Probability Heatmap

	Α	В	С	
2B	78	21	0	
2C	75	24	0	
2D	99	1	0	
2E	70	30	1	
2F	74	26	1	
21	85	15	0	
2K	100	0 31	0	
2N	69	31	1	
2Q	96	4	0	
2AD	28	68	4	
2AI	5 45	73	22	
2AL	45	54	4 22 2	
2AN	7	77	16	

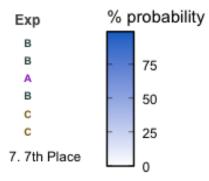


# External Validation Confusion Matrix



### External Validation Probability Heatmap

	Α	В	C
2BD	23	73	5
2BE	1	33	66
2BF	54	45	1
2BG	6	75	19
2BH	0	1	99
2BI	1	28	72



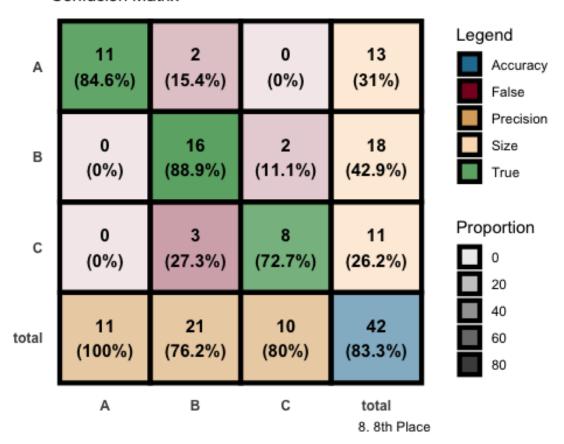
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

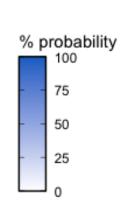
Accuracy	McFadden_R2
83.33%	0.553

	X
dip_y	-1.267705
NPA_1	37.769555
NPA_3	-256.395854
diff 2 7	-60.498122

Training Set Confusion Matrix



	Α	В	С
2A	41	57	2
2G	89	10	0
2H	93	7	Ö
2L	71	29	Ö
2M	83	17	0
20	99	1	0
2R	8	80	13
2S	89	11	0
2T	99	1	0
2U	94	6	0
2V	99	1	0
2J	81	18	0
2P	97	3	0
2Y	1	54	44
2Z	1	46	53
2AA	1	37	62
2AE	11	80	9
2AF	41	58	2
2AG	14	79	7
2AH	14	79	2 7 7
2AJ	1	52 78	46
2AK	6	78	16
2AM	28	69	3
2AO	17	78	6
2W	28	69	3 6
2X	17	77	6
2AB	8	80	12
2AC	4	72	24
2AP	28	69	3
2AQ	17	78	6
2AR	3	66	32
2AS	1	36	64
2AT	1	35	64
2AU	47	52	1
2AV	1	37	62
2AW	4	37 72	25
2AX	0	16	84
2AY	2	59	40
2AZ	0	5	95
2BA	0	0	100
2BB	0	1	99
2BC	0	1	99



8. 8th Place

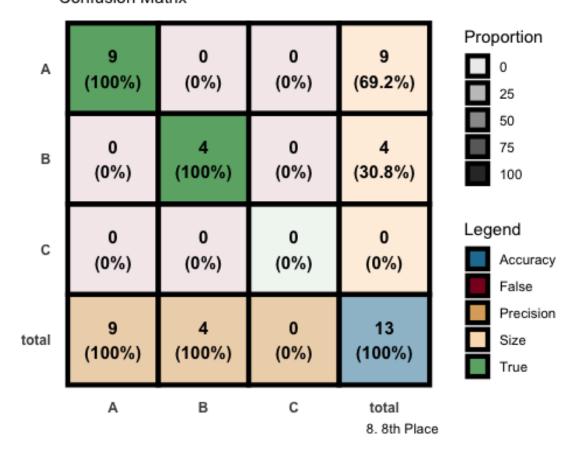
С

Exp

В

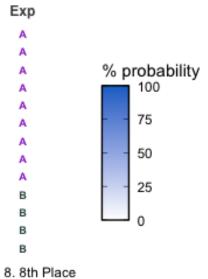
B B B

Test Set Confusion Matrix

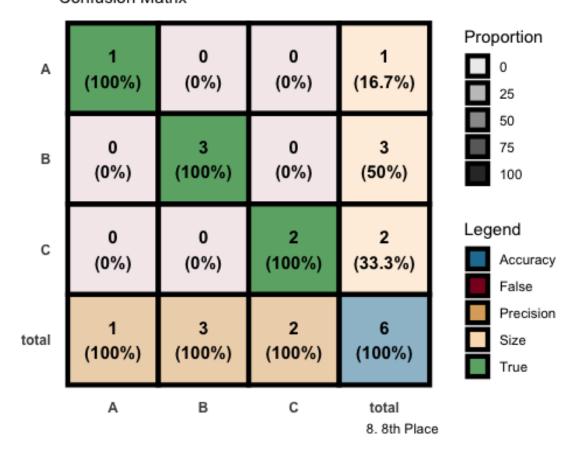


Test Set Probability Heatmap

	Α	В	С
2B	78	22	0
2C	75	22 25	0
2D	99	1	0
2E	69	30 25 14	1
2F	74	25	0
21	86	14	0
2K	100	0 33	0
2N	67	33	1
2Q	96	4	0
2AD	21	75	4
2AI	21 5	77	18
2AL	44	55	2
2AN	7	79	14

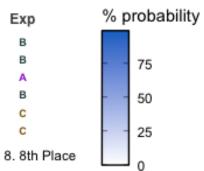


# External Validation Confusion Matrix



Probability	Heatmap
-------------	---------

	A	В	C	Exp
2BD	23	73	4	В
2BE	2	65	33	В
2BF	55	44	1	Α
2BG	5	77	18	В
2BH	0	1	99	С
2BI	1	33	67	С



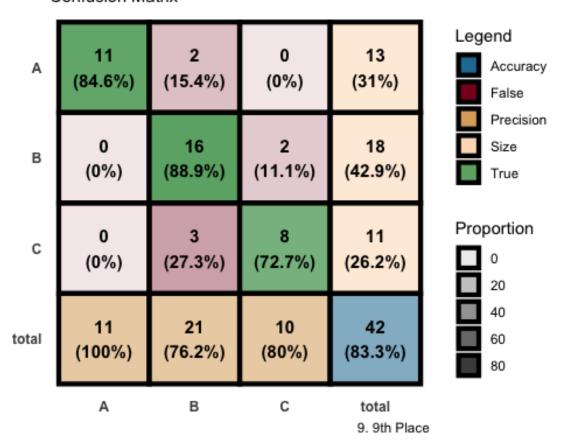
#### Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

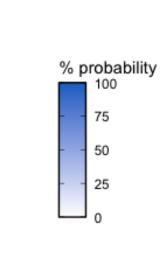
Accuracy	McFadden_R2	
83.33%	0.551	

	X
dip_y	-1.311817
NPA_3	-250.122843
NPA_7	57.898207
diff 1 2	-37.003103

Training Set Confusion Matrix



	Α	В	С
2A	42	56	2
2G	89	10	0
2H	93	7	0
2L	71	29	1
2M	83	17 1	0
20	99	1	0
2R	99 7	78	15
2S	91	9	0
2T	99	1	0
2U	94	6	0
2V	99	1	0
2J	82	18	0
2P	97	3	0
2Y	2	57	41
2Z	1	47	52
2AA		39	60
2AE	10	80	10
2AF	39	59	2 8
2AG	13	79	8
2AH	13 13 1	79 79 49	8
2AJ	1	49	49
2AK	6	77	18
2AM	29	68	7
2AO	15 29	78	7
2W	29	68	3 5
2X	18	76	5
2AB	9	79	12
2AC	4	73	22
2AP	29	68	3 7
2AQ	15	78	/
2AR	3	67	31
2AS	1	42	57
2AT	1	37	62
2AU	47	51	1
2AV	1	39	60
2AW	3	70	27
2AX	0	14	86
2AY	2	59	39
2AZ	0	4	96
2BA	0	0	100
2BB	0	1	99
2BC	0	1	99



С

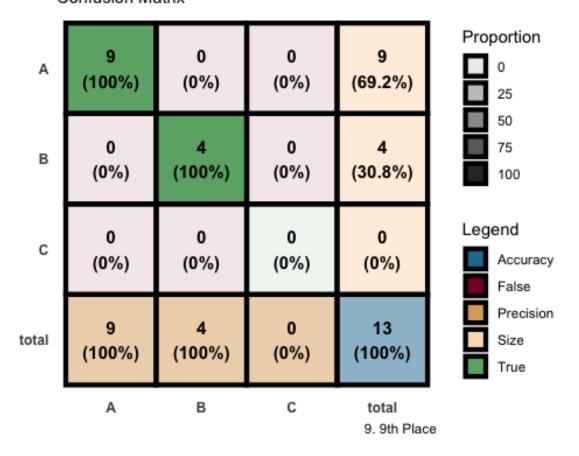
Exp

В

В

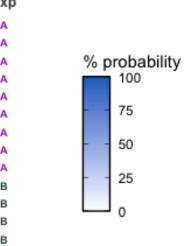
B B B B B

Test Set Confusion Matrix

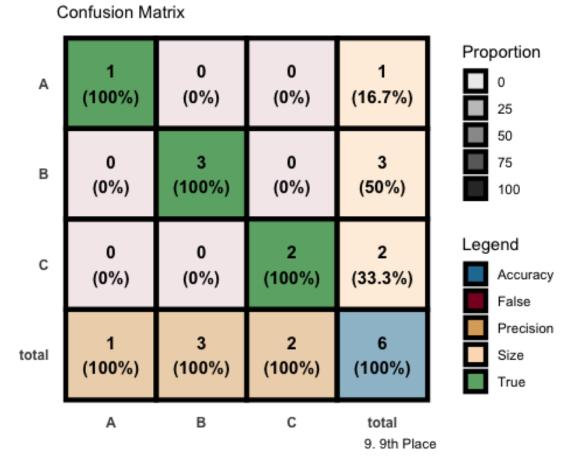


Test Set Probability Heatmap

	Α	В	С	Exp
2B	77	23	0	A
2C	74	25	0	Α
2D	99	1	0	Α
2E	69	30	1	Α
2F	73	26	0	Α
21	86	14	0	Α
2K	100	0	0	Α
2N	67	33	1	Α
2Q	96	4	0	Α
2AD	22	73	4	В
2AI	5	75	20 2	В
2AL	42	56	2	В
2AN	6	77	17	В



9. 9th Place



### External Validation Probability Heatmap

	^	ь	0
2BD	22	74	4
2BE	3	68	29
2BF	55	44	1
2BG	5	76	19
2BH	0	1	99
2BI	1	29	71

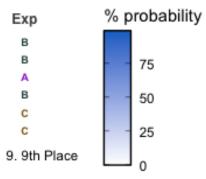


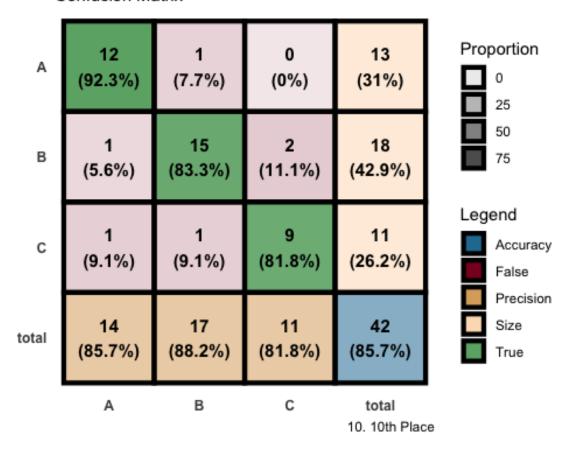
Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

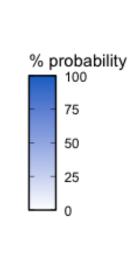
Accuracy McFadden\_R2 85.71% 0.549

	X
-2-3-	-0.1368705
Total	0.9140378
NPA_1	32.9947500
NPA 3	-222.4549516

Training Set Confusion Matrix



	Α	В	С
2A	68	32	0
2G	82	18	0
2H	97	3	0
2L	55	45	1
2M	76	23	Ó
20	96	23 4	0
2R	20	76	4
2S	69	31	0
2T	98	2	0
2U	90	10	0
2V	96	4	0
2J	69	31	0
2P	98	2	0
2Y	7	81	12
2Z	1 0	54	45 71 7
2AA	0	29	71
2AE	11	82	7
2AF	23	74	3
2AG	20	76	4
2AH	20 22	75	3
2AJ	7	76 75 58	40
2AK	7	81	12
2AM	10 17	82	8
2AO	17	79	5
2W	57	42	1
2X	41	57	1
2AB	3	74	22
2AC	32	66	2
2AP	20 11	76	7
2AQ	11	82	7
2AR	1	47	52
2AS	1	42	57
2AT	0	21	79
2AU	69	31	0
2AV	0	18	82
2AW	3	73	24
2AX	0	8	92
2AY	11	48	52
2AZ	0	6	94
2BA	0	0	100
2BB	0	0	100
2BC	0	0	100



С

Exp

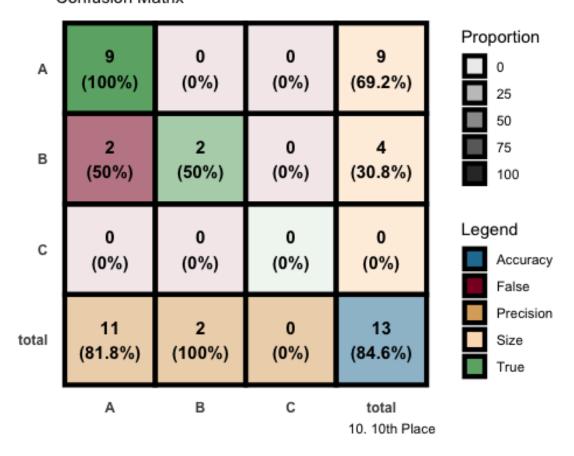
В

В

В

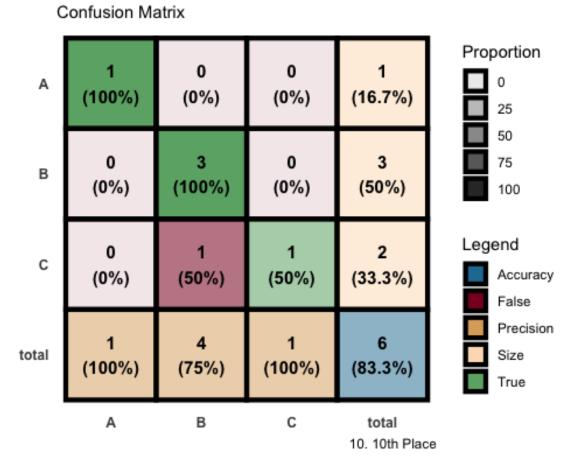
B B B

Test Set Confusion Matrix



Test Set Probability Heatmap

	Α	В	С	Exp	
2B	90	9	0	A	
2C	89	10	0	A	
2D	98	2	0	A	% probability
2E	88	12	0	A	
2F	88	12	0	A	
21	77	22	0	A	75
2K	96	4	0	A	50
2N	87	13	0	A	50
2Q	97	3	0	A	- 25
2AD	52	47	1	В	23
2AI	6	80	14	В	
2AL	68	32	0	В	v
2AN	26	72	3	В	
				10. 10th Place	



### External Validation Probability Heatmap

	Α	В	С	Exp	% probability
2BD	26	71	3	В	100
2BE	2	66	32	В	- 75
2BF	61	39	1	A	
2BG	2	62	37	В	50
2BH	0	0	100	С	
2BI	1	52	47	С	25
				10. 10th Place	