

CH Activation - Model Testing

Table of Contents

Section Results Summary	1
Truncated Datasets - Union of Similarity within Groups and with Group 3	1
1st Place.....	2
2nd Place.....	7
3rd Place.....	12
4th Place.....	17
5th Place.....	22
6th Place.....	27
7th Place.....	32
8th Place.....	37
9th Place.....	42
10th Place.....	47

Section Results Summary

	Training.Accuracy	Test.Accuracy
1. 1st Place	85.42	75.0
2. 2nd Place	85.42	75.0
3. 3rd Place	85.42	91.7
4. 4th Place	83.33	75.0
5. 5th Place	85.42	75.0
6. 6th Place	83.33	75.0
7. 7th Place	83.33	75.0
8. 8th Place	83.33	75.0
9. 9th Place	85.42	58.3
10. 10th Place	83.33	75.0

Truncated Datasets - Union of Similarity within Groups and with Group 3

formula	McFadden R2
class ~ fr_NH0 + fr_aldehyde + fr_amide + fr_para_hydroxylation	0.766
class ~ fr_Ar_N + fr_aldehyde + fr_amide + fr_para_hydroxylation	0.764
class ~ fr_NH0 + fr_aldehyde + fr_amide + fr_pyridine	0.762
class ~ fr_aldehyde + fr_amide + fr_para_hydroxylation + fr_pyridine	0.755
class ~ fr_Ar_N + fr_NH2 + fr_aldehyde + fr_amide	0.738

formula	McFadden R2
class ~ fr_C_O + fr_NH0 + fr_NH1 + fr_aldehyde	0.737
class ~ fr_C_O_noCOO + fr_NH0 + fr_NH1 + fr_aldehyde	0.737
class ~ fr_NH0 + fr_NH1 + fr_aldehyde + fr_amide	0.737
class ~ fr_NH0 + fr_NH1 + fr_aldehyde + fr_pyridine	0.737
class ~ fr_NH0 + fr_NH2 + fr_aldehyde + fr_amide	0.737

1st Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
85.42%	0.766

Model Coefficients

	(Inter cept)	fr_N H01	fr_NH 02	fr_NH 05	fr_alde hyde1	fr_ami de1	fr_am ide2	fr_para_hydr oxylation1	fr_para_hydr oxylation2
2	89.74 685	- 91.9 592	- 89.31 897	- 11.24 9980	0.8923 468	- 52.15 421	- 15.56 946	146.9118	-15.56946
3	32.64 020	- 205. 1309	- 59.22 859	- 4.392 311	- 205.31 17375	172.7 7838	91.75 199	126.8486	91.75199
4	126.4 2535	- 204. 5713	- 126.1 8874	31.56 6324	- 220.61 62132	77.74 052	- 60.00 557	126.8486	-60.00557

Training Set

Confusion Matrix




1	8 (80%)	1 (10%)	1 (10%)	0 (0%)	10 (20.8%)
2	1 (7.1%)	13 (92.9%)	0 (0%)	0 (0%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	1 (7.1%)	3 (21.4%)	10 (71.4%)	14 (29.2%)
total	9 (88.9%)	15 (86.7%)	14 (71.4%)	10 (100%)	48 (85.4%)
	1	2	3	4	total

1. 1st Place

Legend

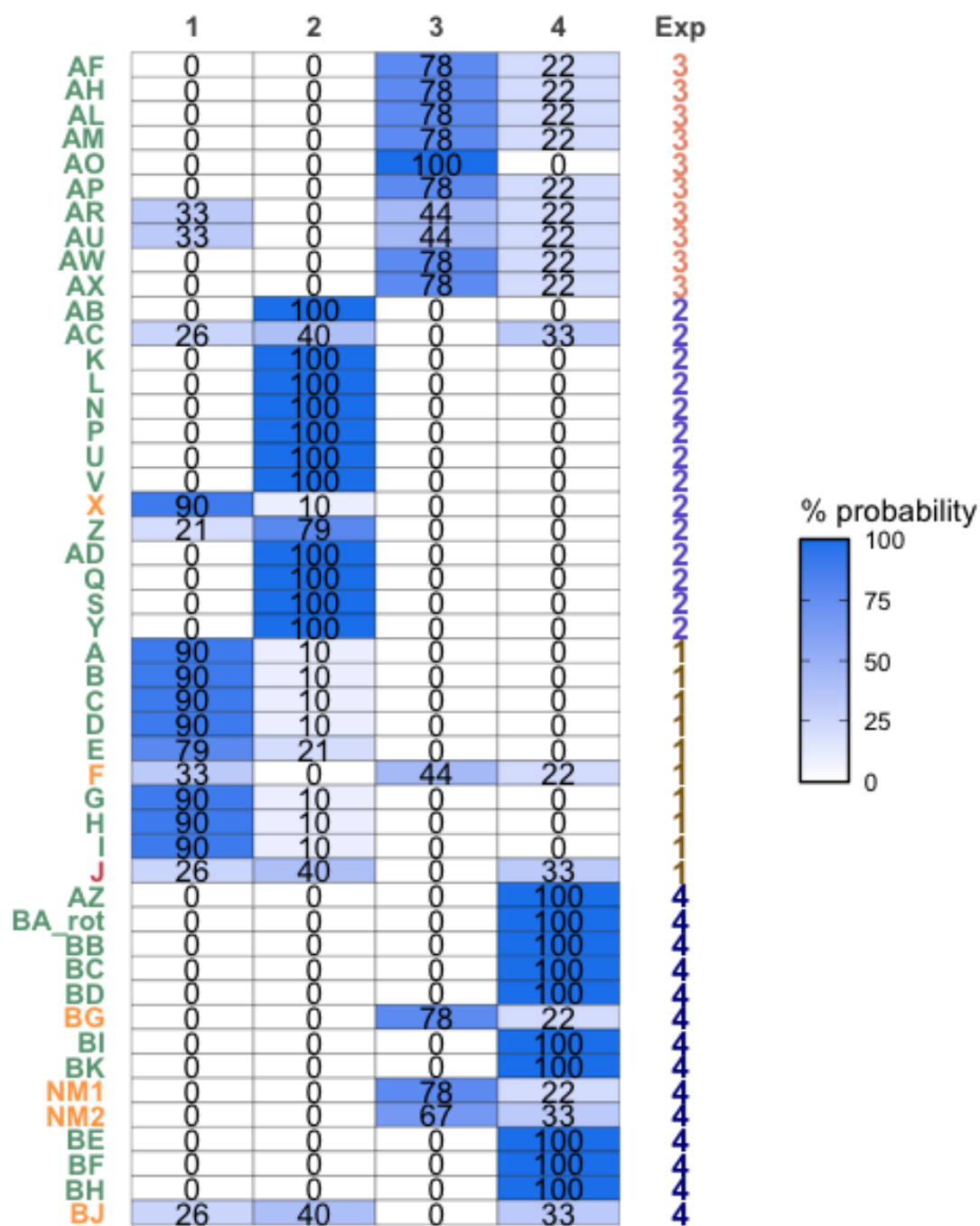
	Accuracy
	False
	Precision
	Size
	True

prop

	0
	25
	50
	75
	100

Training Set

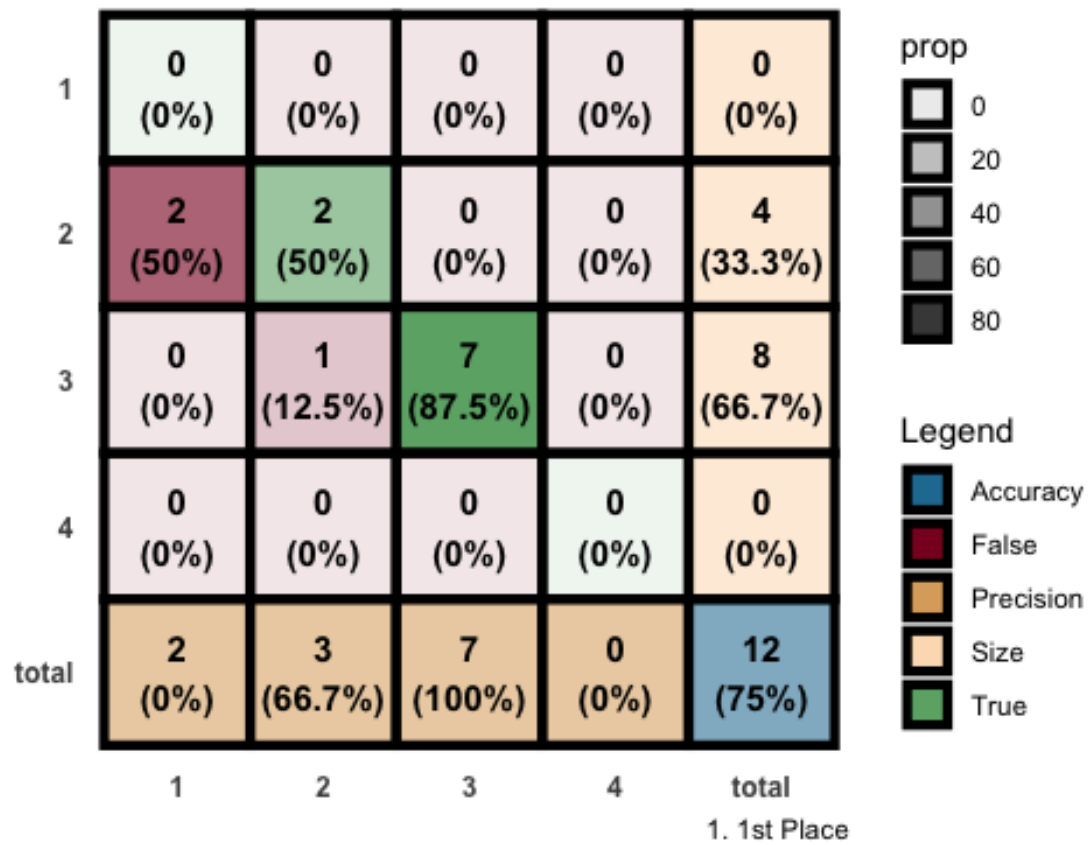
Probability Heatmap



1. 1st Place

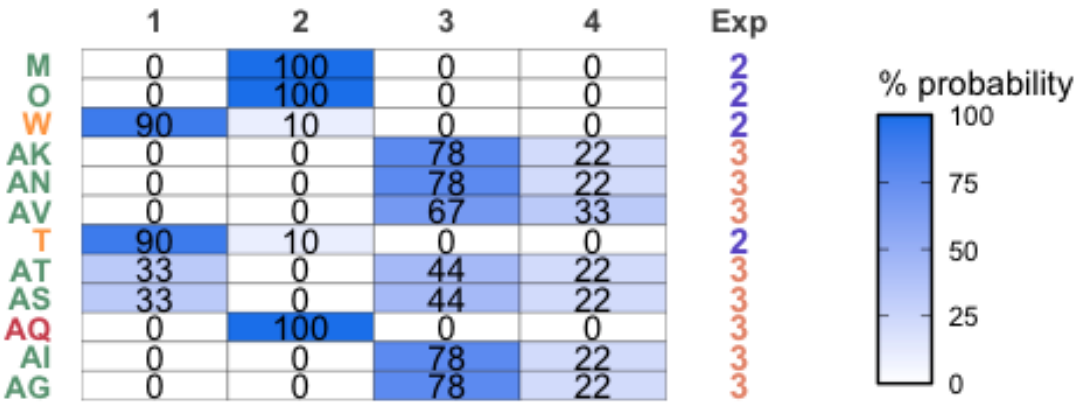
Test Set

Confusion Matrix



Test Set

Probability Heatmap



1. 1st Place

2nd Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
85.42%	0.764

Model Coefficients

	(Inter cept)	fr_Ar_ N1	fr_Ar_ _N2	fr_Ar_ N3	fr_alde hyde1	fr_am ide1	fr_am ide2	fr_para_hydr oxylation1	fr_para_hydr oxylation2
2	59.70 755	- 61.91 986	- 59.2 7963	- 7.165 543	0.8922 544	- 44.34 8597	- 9.332 006	73.95221	-9.332006
3	12.76 279	- 156.7 7406	- 45.2 3150	- 2.129 974	- 102.95 57782	72.55 6856	55.07 1413	51.74269	55.071413
4	82.61 025	- 125.2 9993	- 82.3 7361	17.94 2911	- 117.54 85908	1.610 791	- 31.50 9658	51.74267	-31.509658

Training Set



Confusion Matrix

1	9 (90%)	1 (10%)	0 (0%)	0 (0%)	10 (20.8%)
2	1 (7.1%)	12 (85.7%)	0 (0%)	1 (7.1%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	1 (7.1%)	3 (21.4%)	10 (71.4%)	14 (29.2%)
total	10 (90%)	14 (85.7%)	13 (76.9%)	11 (90.9%)	48 (85.4%)
	1	2	3	4	total

Legend

	Accuracy
	False
	Precision
	Size
	True

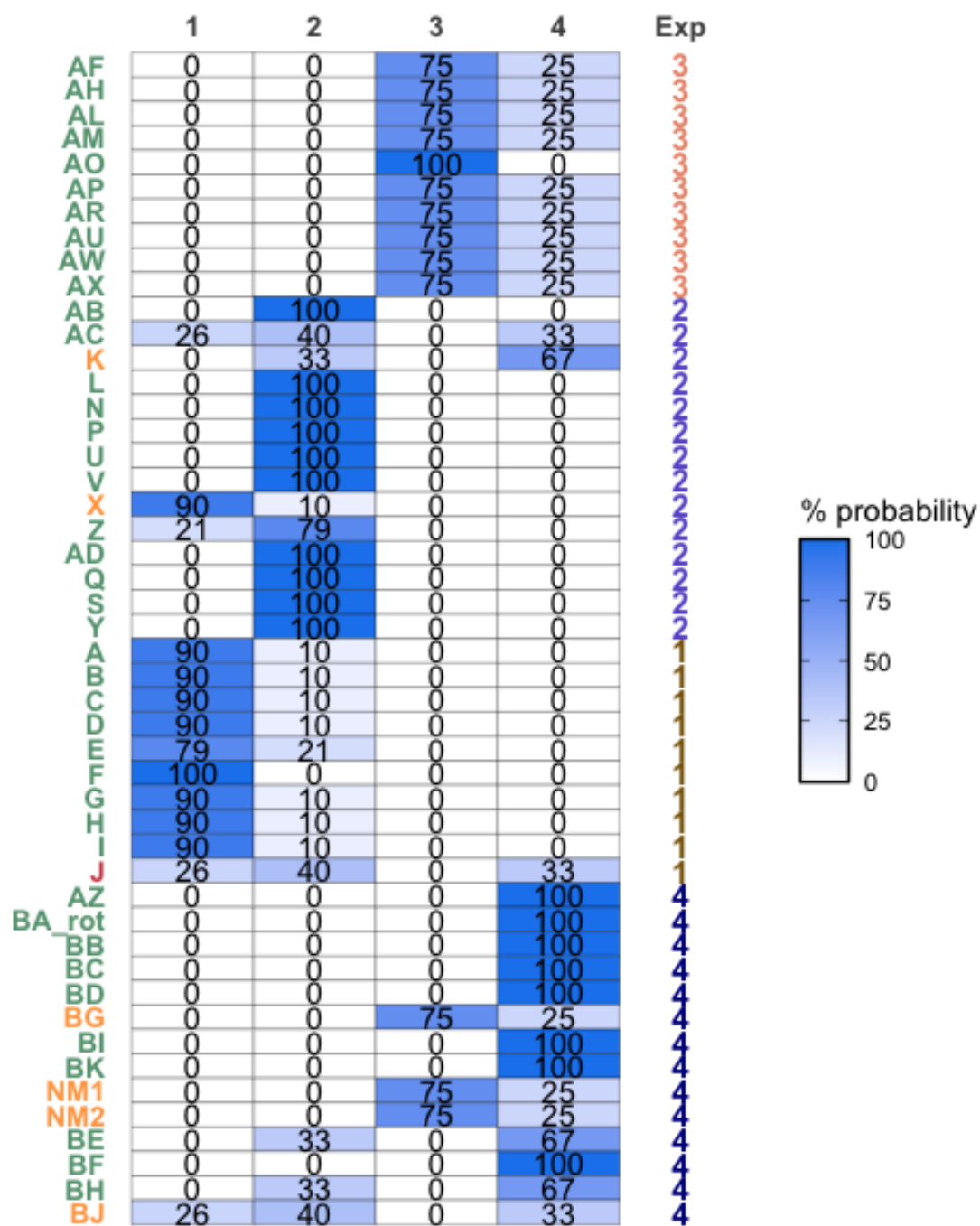
prop

	0
	25
	50
	75
	100

2. 2nd Place

Training Set

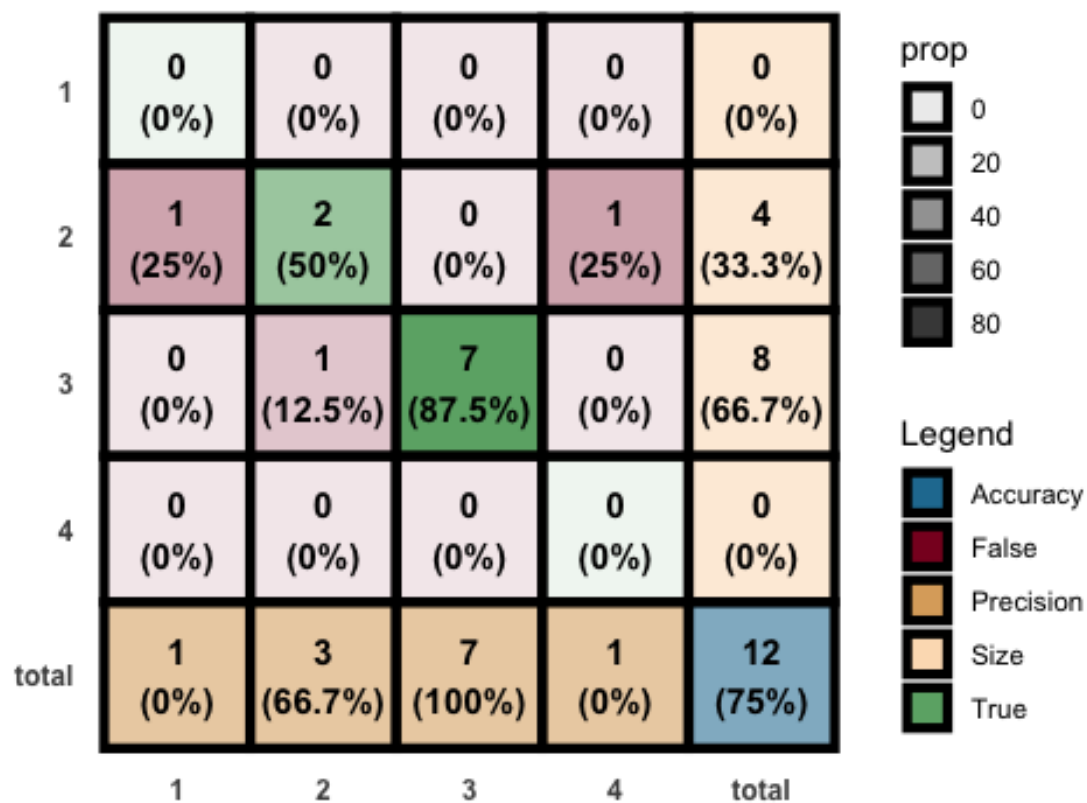
Probability Heatmap



2. 2nd Place

Test Set

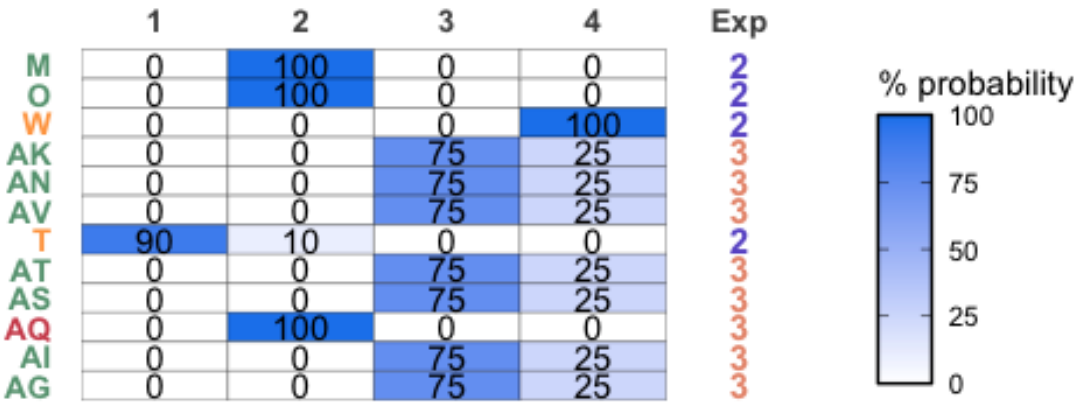
Confusion Matrix



2. 2nd Place

Test Set

Probability Heatmap



2. 2nd Place

3rd Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
85.42%	0.762

Model Coefficients

	(Interc ept)	fr_NH0 1	fr_NH 02	fr_NH0 5	fr_aldehy de1	fr_amid e1	fr_amid e2	fr_pyridi ne1	fr_pyridi ne2
2	77.146 276	7.6137 79	- 76.42 688	- 11.946 103	0.84884 33	- 56.072 04	- 20.527 82	- 86.2502 4	78.0081 71
3	8.3193 67	- 52.025 492	- 57.35 079	- 3.9100 05	- 185.047 3453	103.38 214	111.78 747	- 162.834 43	- 3.61883 8
4	95.677 264	- 51.465 887	- 95.25 421	24.534 540	- 196.787 2443	14.771 46	- 72.367 90	- 155.401 07	- 33.2751 26

Training Set

Confusion Matrix

1	9 (90%)	1 (10%)	0 (0%)	0 (0%)	10 (20.8%)
2	2 (14.3%)	12 (85.7%)	0 (0%)	0 (0%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	1 (7.1%)	3 (21.4%)	10 (71.4%)	14 (29.2%)
total	11 (81.8%)	14 (85.7%)	13 (76.9%)	10 (100%)	48 (85.4%)
	1	2	3	4	total

Legend

	Accuracy
	False
	Precision
	Size
	True

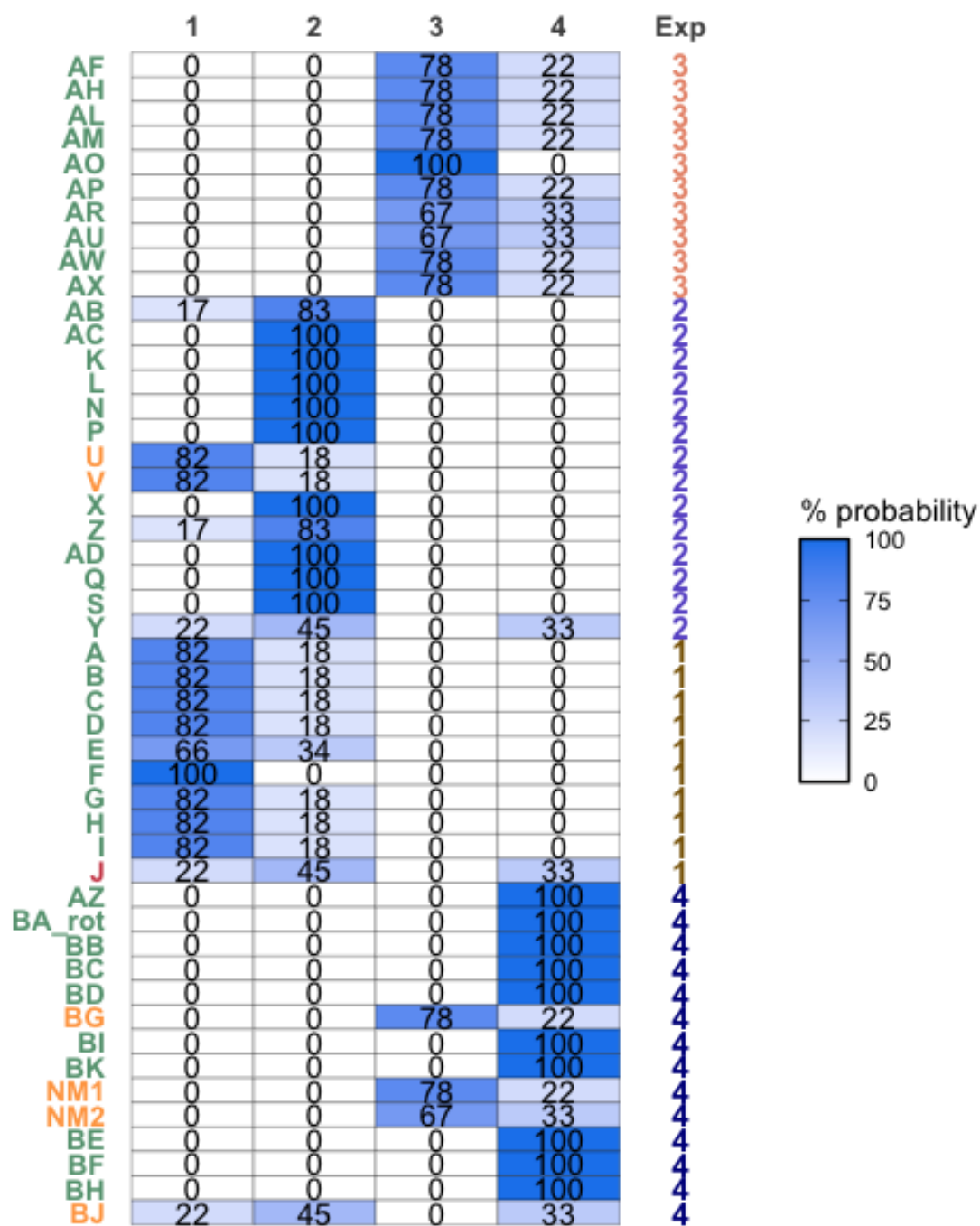
prop

	0
	25
	50
	75
	100

3. 3rd Place

Training Set

Probability Heatmap



3. 3rd Place





Test Set

Confusion Matrix



1	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
2	0 (0%)	4 (100%)	0 (0%)	0 (0%)	4 (33.3%)
3	0 (0%)	1 (12.5%)	7 (87.5%)	0 (0%)	8 (66.7%)
4	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
total	0 (0%)	5 (80%)	7 (100%)	0 (0%)	12 (91.7%)
	1	2	3	4	total

3. 3rd Place

Legend

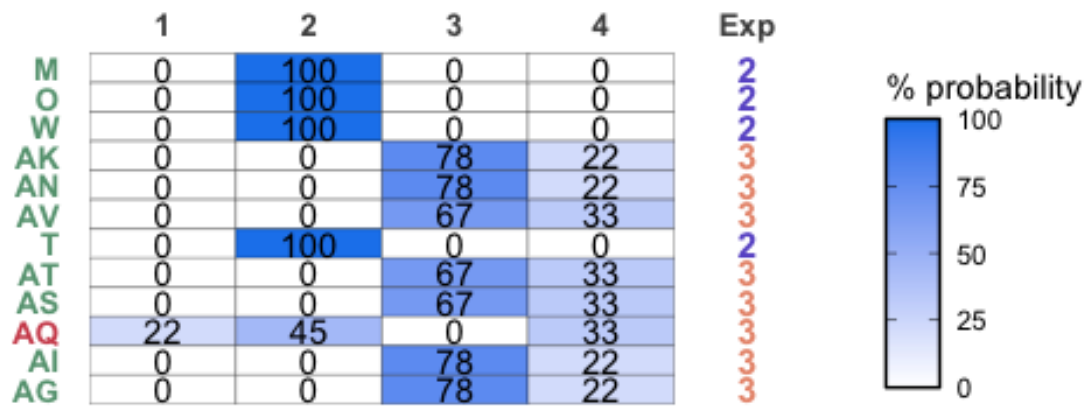
	Accuracy
	False
	Precision
	Size
	True

prop

	0
	25
	50
	75
	100

Test Set

Probability Heatmap



3. 3rd Place

4th Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
83.33%	0.755

Model Coefficients

	(Interc ept)	fr_aldeh yde1	fr_ami de1	fr_ami de2	fr_para_hydro xylation1	fr_para_hydro xylation2	fr_pyri dine1	fr_pyri dine2
2	0.0000 181	29.465 78	14.76 031	- 11.69 553	116.6129	-11.69553	- 64.068 25	71.874 72
3	- 16.320 6609	- 66.190 00	73.79 567	61.55 836	114.4155	61.55836	- 90.047 95	- 2.9087 6
4	2.1972 931	- 56.251 33	54.17 908	- 30.82 048	114.4156	-30.82048	- 104.88 701	- 46.903 15

Training Set

Confusion Matrix

1	9 (90%)	0 (0%)	0 (0%)	1 (10%)	10 (20.8%)
2	0 (0%)	12 (85.7%)	0 (0%)	2 (14.3%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	2 (14.3%)	3 (21.4%)	9 (64.3%)	14 (29.2%)
total	9 (100%)	14 (85.7%)	13 (76.9%)	12 (75%)	48 (83.3%)
	1	2	3	4	total

4. 4th Place

Legend

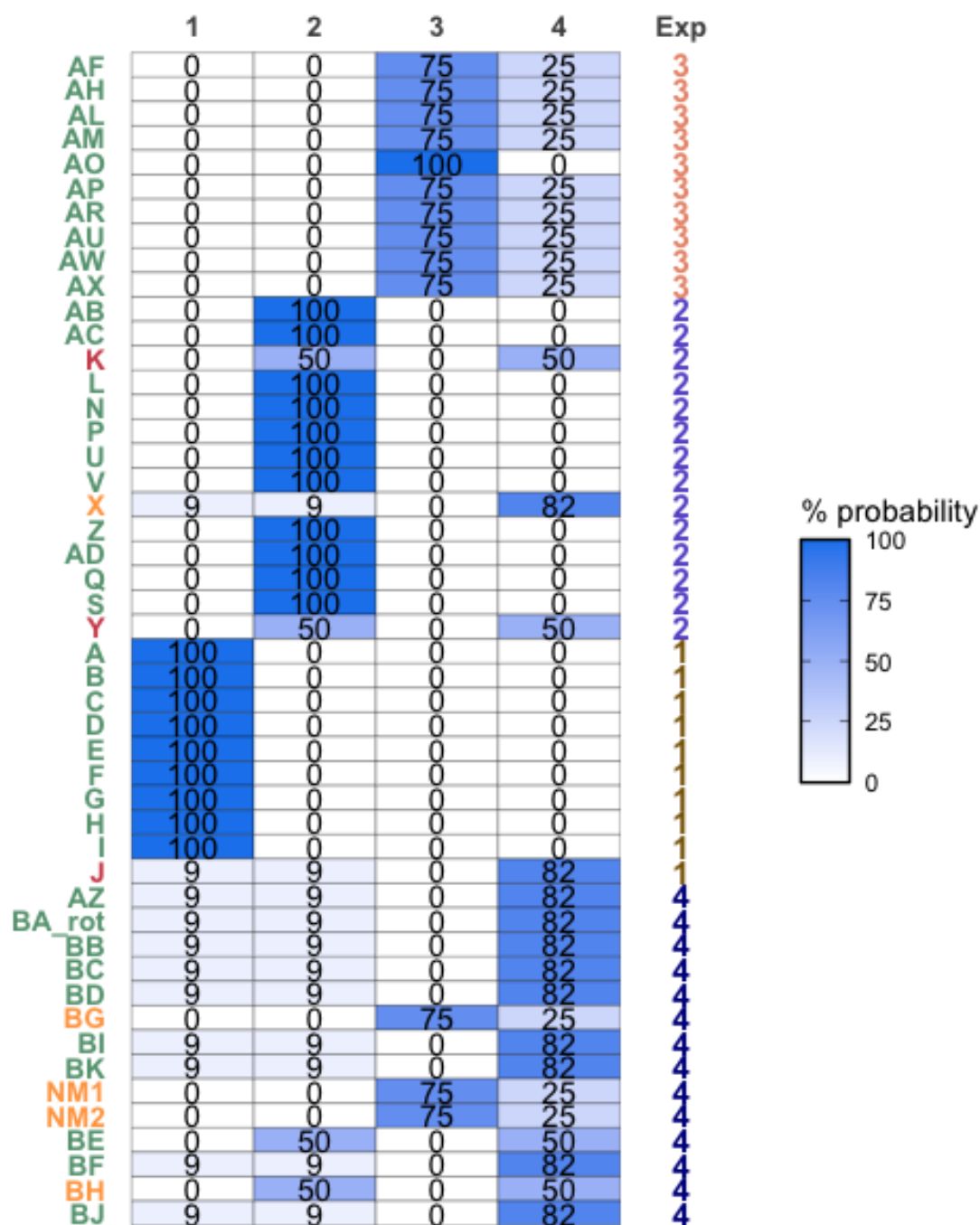
- Accuracy
- False
- Precision
- Size
- True

prop

- 0
- 25
- 50
- 75
- 100

Training Set

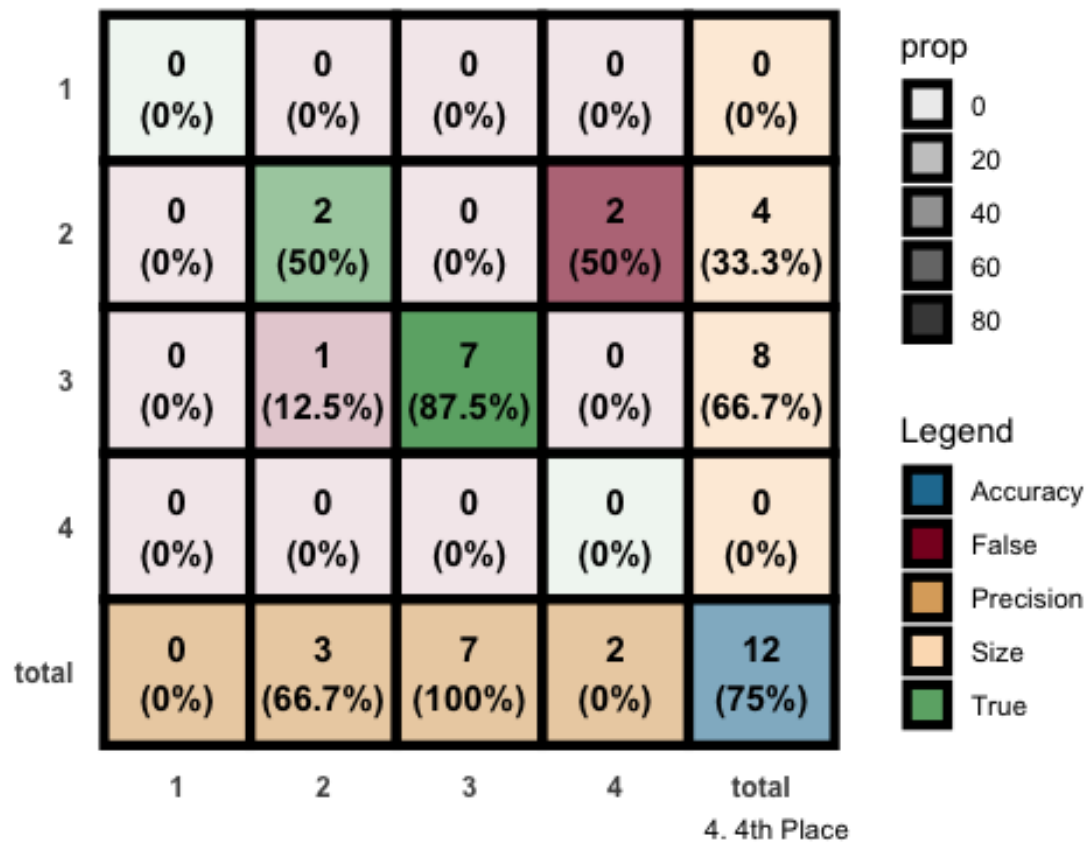
Probability Heatmap



4. 4th Place

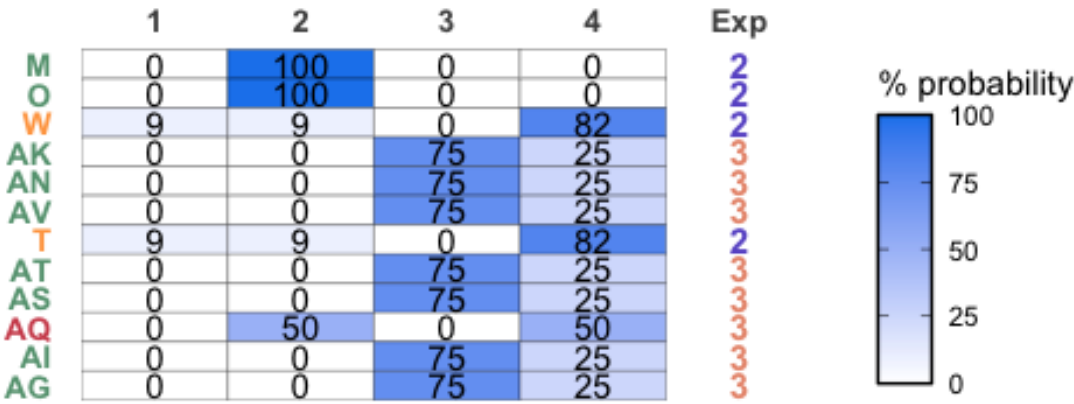
Test Set

Confusion Matrix



Test Set

Probability Heatmap



4. 4th Place

5th Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
85.42%	0.738

Model Coefficients

	(Intercept)	fr_Ar_N1	fr_Ar_N2	fr_Ar_N3	fr_NH21	fr_aldehyde1	fr_amide1	fr_amide2
2	59.85048	-	-	-	-	0.3366145	-	-
	1	60.86394	58.59069	9.573486	20.00759		36.049845	13.01705
3	8.976537	-	-	-	-	-	52.056849	73.38108
		77.20032	25.85147	2.184739	13.53670	59.3923268		
4	61.92992	-	-	20.7181	52.7230	-	-	-
	3	87.74431	85.87365	03	4	69.5332164	2.400612	42.89410

Training Set

Confusion Matrix

1	9 (90%)	1 (10%)	0 (0%)	0 (0%)	10 (20.8%)
2	3 (21.4%)	10 (71.4%)	0 (0%)	1 (7.1%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	0 (0%)	2 (14.3%)	12 (85.7%)	14 (29.2%)
total	12 (75%)	11 (90.9%)	12 (83.3%)	13 (92.3%)	48 (85.4%)
	1	2	3	4	total

5. 5th Place

Legend

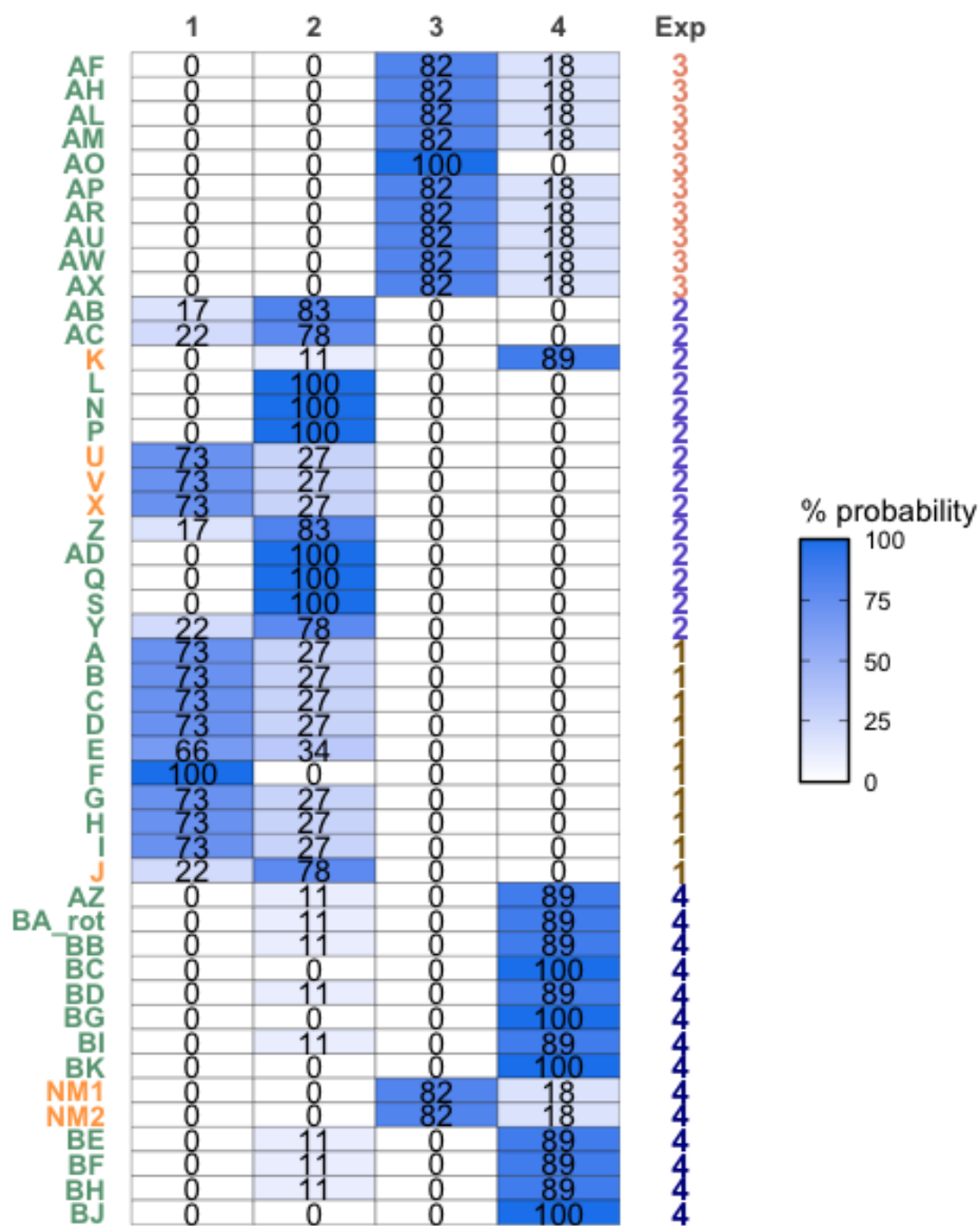
- Accuracy
- False
- Precision
- Size
- True

prop

- 0
- 25
- 50
- 75
- 100

Training Set

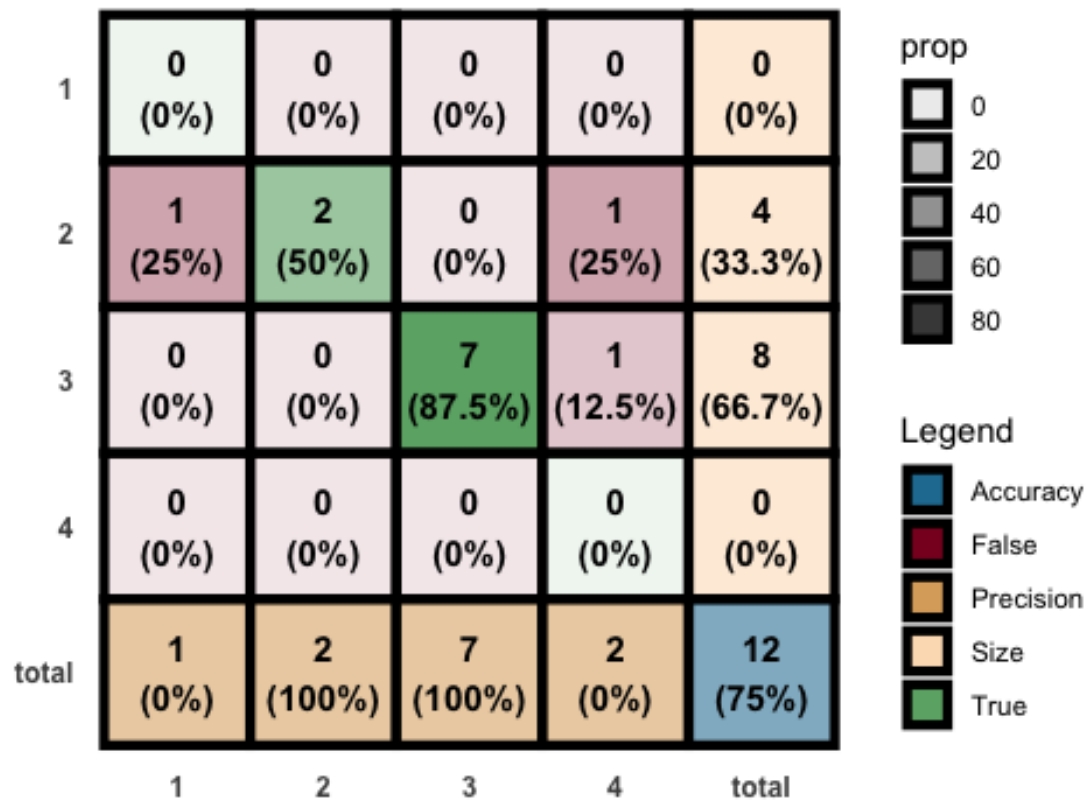
Probability Heatmap



5. 5th Place

Test Set

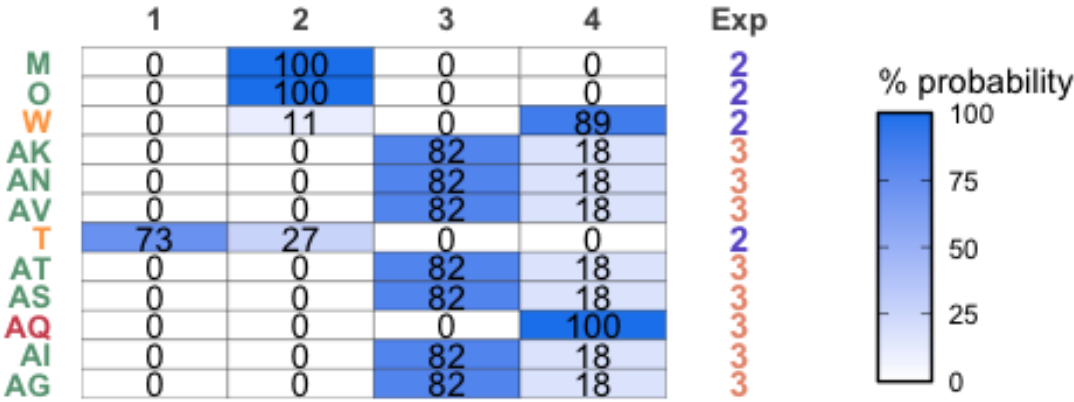
Confusion Matrix



5. 5th Place

Test Set

Probability Heatmap



5. 5th Place

6th Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
83.33%	0.737

Model Coefficients

	(Interc ept)	fr_C_O1	fr_C_O2	fr_NH0 1	fr_NH0 2	fr_NH0 5	fr_NH1 1	fr_NH1 2	fr_aldehy de1
2	70.170 34	- 52.717 75	- 12.812 964	- 70.877 79	- 68.848 94	- 12.990 123	- 15.998 35	- 3.3681 35	52.8851 5
3	22.616 67	138.16 561	- 7.8259 94	- 83.294 22	- 38.018 90	- 4.0756 28	- 115.47 094	70.560 126	- 142.601 97
4	124.86 697	125.68 492	49.508 695	- 173.75 697	- 124.40 768	31.636 503	- 207.18 644	- 61.411 756	- 233.604 56





Training Set

Confusion Matrix




1	9 (90%)	1 (10%)	0 (0%)	0 (0%)	10 (20.8%)
2	4 (28.6%)	10 (71.4%)	0 (0%)	0 (0%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	1 (7.1%)	2 (14.3%)	11 (78.6%)	14 (29.2%)
total	13 (69.2%)	12 (83.3%)	12 (83.3%)	11 (100%)	48 (83.3%)
	1	2	3	4	total

6. 6th Place

Legend

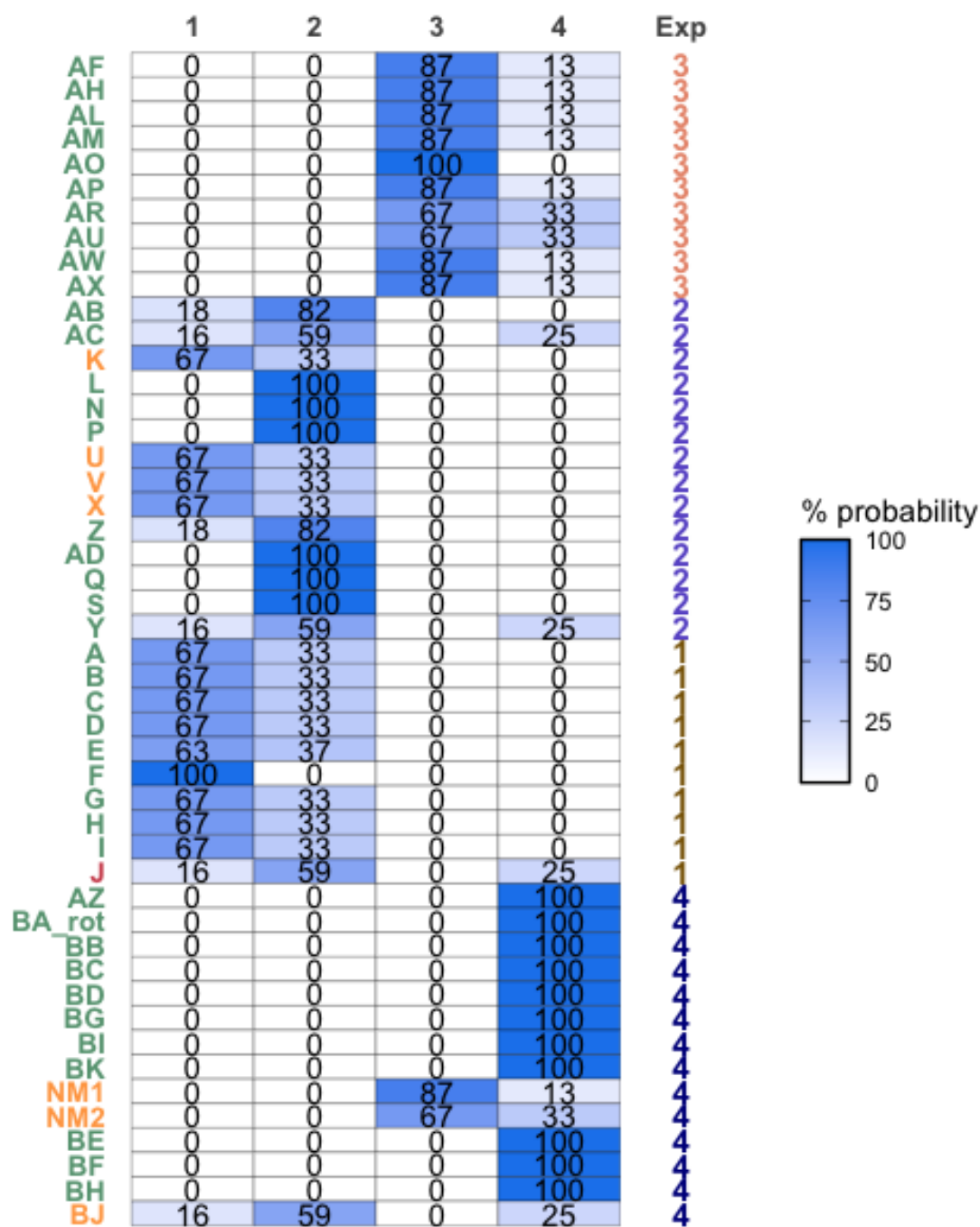
	Accuracy
	False
	Precision
	Size
	True

prop

	0
	25
	50
	75
	100

Training Set

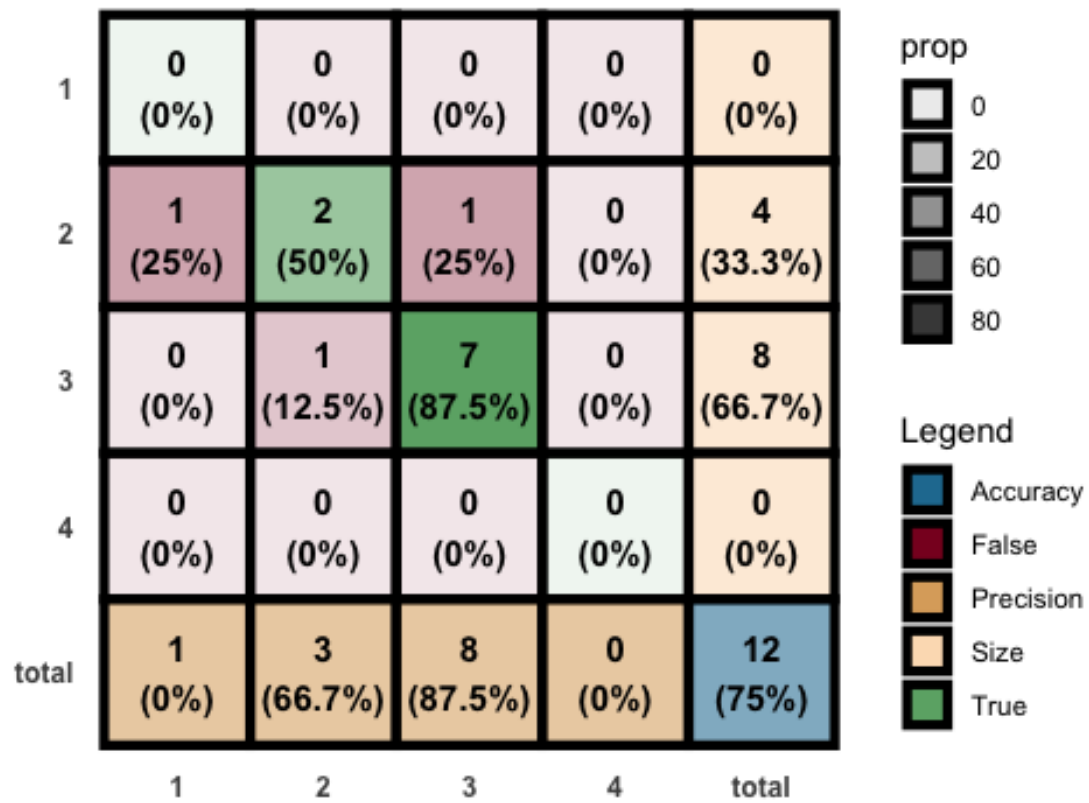
Probability Heatmap



6. 6th Place

Test Set

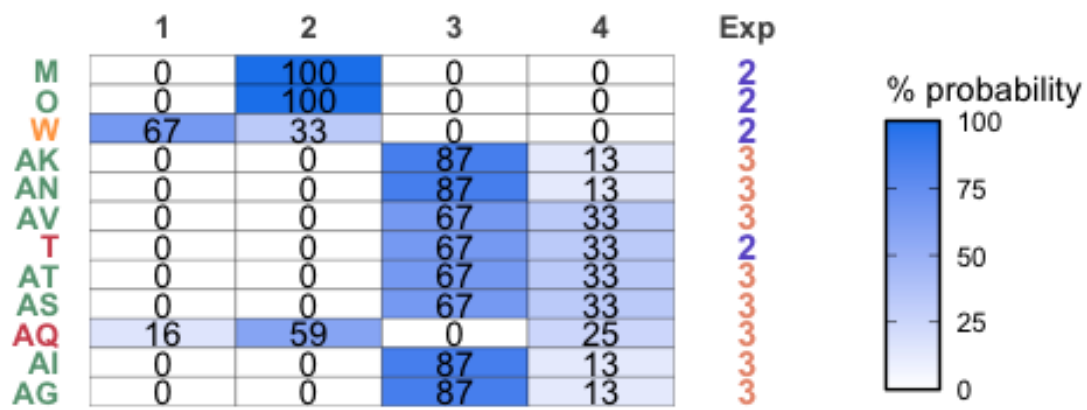
Confusion Matrix



6. 6th Place

Test Set

Probability Heatmap



6. 6th Place

7th Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
83.33%	0.737

Model Coefficients

	(Interc ept)	fr_C_O_no COO1	fr_C_O_no COO2	fr_NH0 1	fr_NH0 2	fr_NH0 5	fr_NH1 1	fr_NH 12	fr_aldeh yde1
2	54.736 286	- 44.66841	- 7.795743	- 55.443 72	- 53.414 87	- 9.8793 02	- 13.409 97	- 2.704 21	44.8358 2
3	9.1139 25	104.5817 8	- 6.811975	- 59.283 24	- 24.743 41	- 3.5230 18	- 83.052 31	65.30 208	- 117.930 56
4	105.67 3656	104.1224 4	38.19880 8	- 156.07 682	- 105.21 437	24.612 613	- 181.09 863	- 57.61 870	- 178.982 33

Training Set

Confusion Matrix

1	9 (90%)	1 (10%)	0 (0%)	0 (0%)	10 (20.8%)
2	4 (28.6%)	10 (71.4%)	0 (0%)	0 (0%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	1 (7.1%)	2 (14.3%)	11 (78.6%)	14 (29.2%)
total	13 (69.2%)	12 (83.3%)	12 (83.3%)	11 (100%)	48 (83.3%)
	1	2	3	4	total

7. 7th Place

Legend

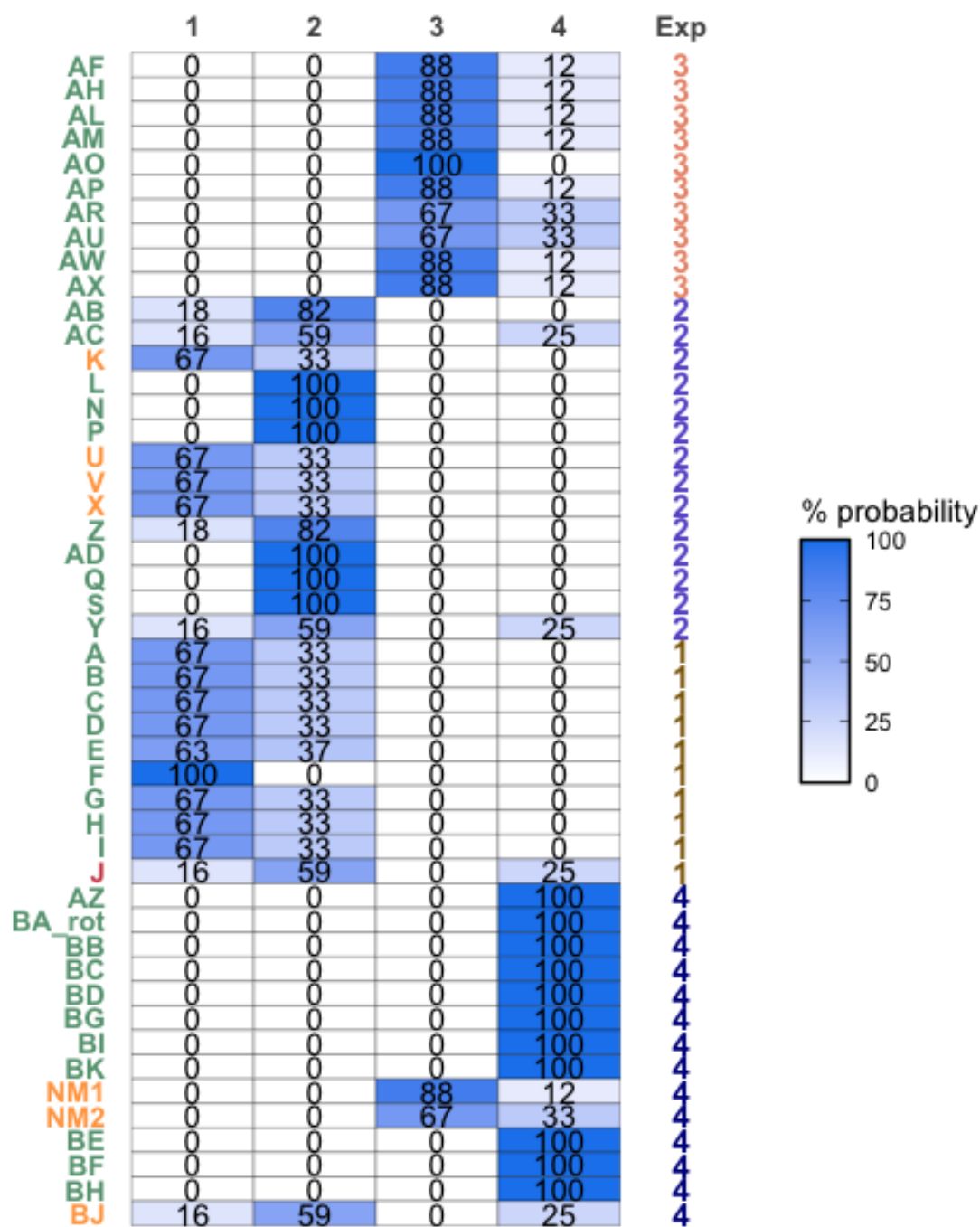
- Accuracy
- False
- Precision
- Size
- True

prop

- 0
- 25
- 50
- 75
- 100

Training Set

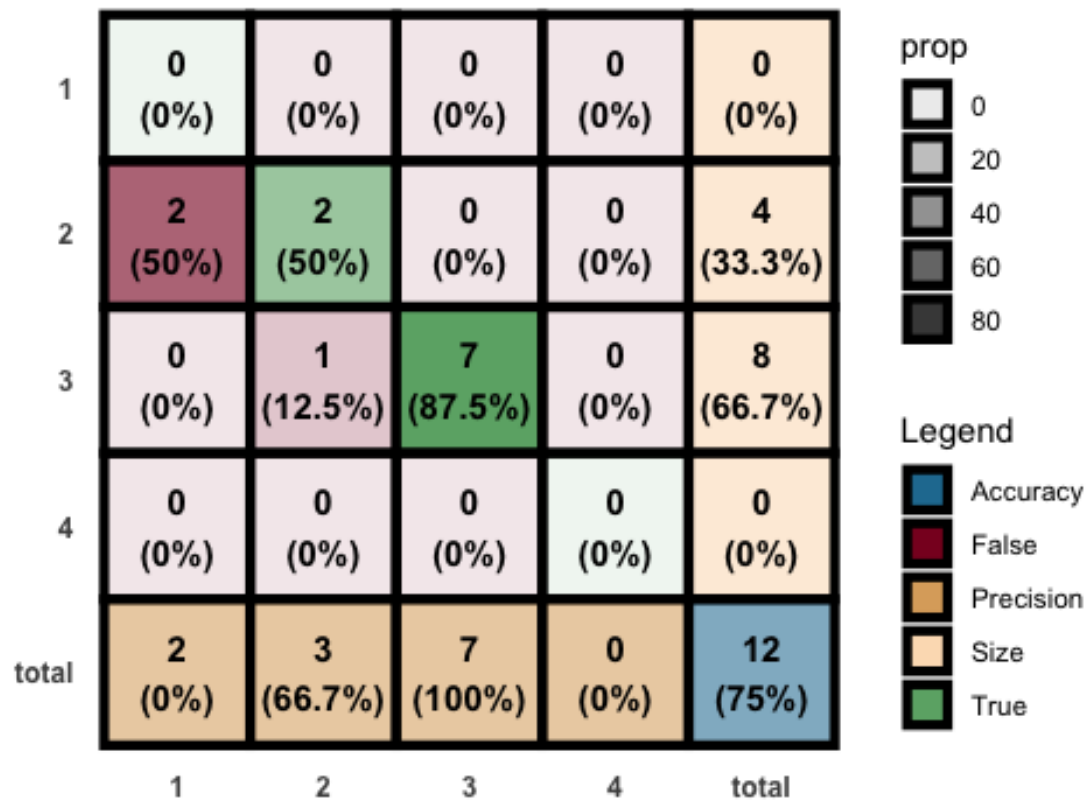
Probability Heatmap



7. 7th Place

Test Set

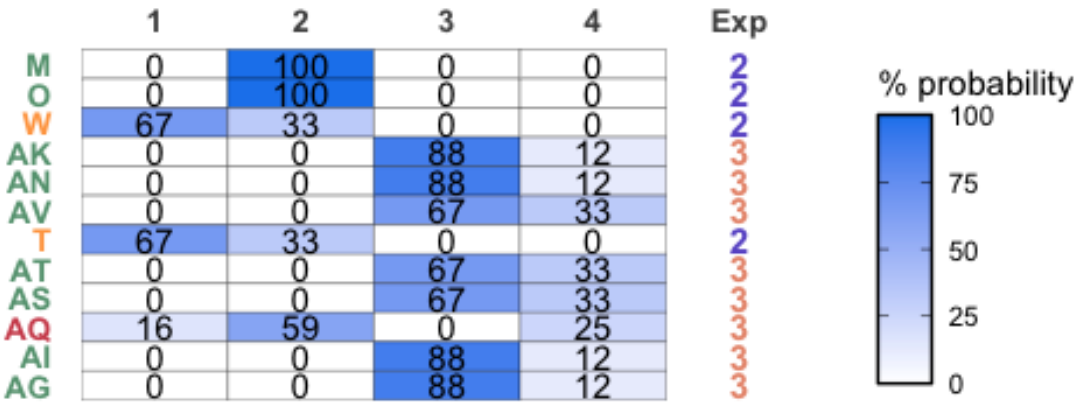
Confusion Matrix



7. 7th Place

Test Set

Probability Heatmap



7. 7th Place

8th Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
83.33%	0.737

Model Coefficients

	(Interc ept)	fr_NH0 1	fr_NH 02	fr_NH0 5	fr_NH11	fr_NH1 2	fr_aldehy de1	fr_amid e1	fr_amid e2
2	60.704	-	-	-	3.12318	-	0.167619	-	-
	80	61.412	59.38	6.8880	1	9.9963	2	41.448	9.9963
		25	359	58		72		92	72
3	27.872	-	-	-	-	57.808	-	119.18	57.808
	99	84.468	43.58	2.4827	104.292	035	63.85601	717	035
		99	631	07	953		58		
4	90.293	-	-	17.764	-	-	-	100.25	-
	37	128.64	89.83	335	149.723	32.920	123.5222	195	32.920
		724	429		994	594	002		594

Training Set

Confusion Matrix

1	9 (90%)	1 (10%)	0 (0%)	0 (0%)	10 (20.8%)
2	4 (28.6%)	10 (71.4%)	0 (0%)	0 (0%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	1 (7.1%)	2 (14.3%)	11 (78.6%)	14 (29.2%)
total	13 (69.2%)	12 (83.3%)	12 (83.3%)	11 (100%)	48 (83.3%)
	1	2	3	4	total

8. 8th Place

Legend

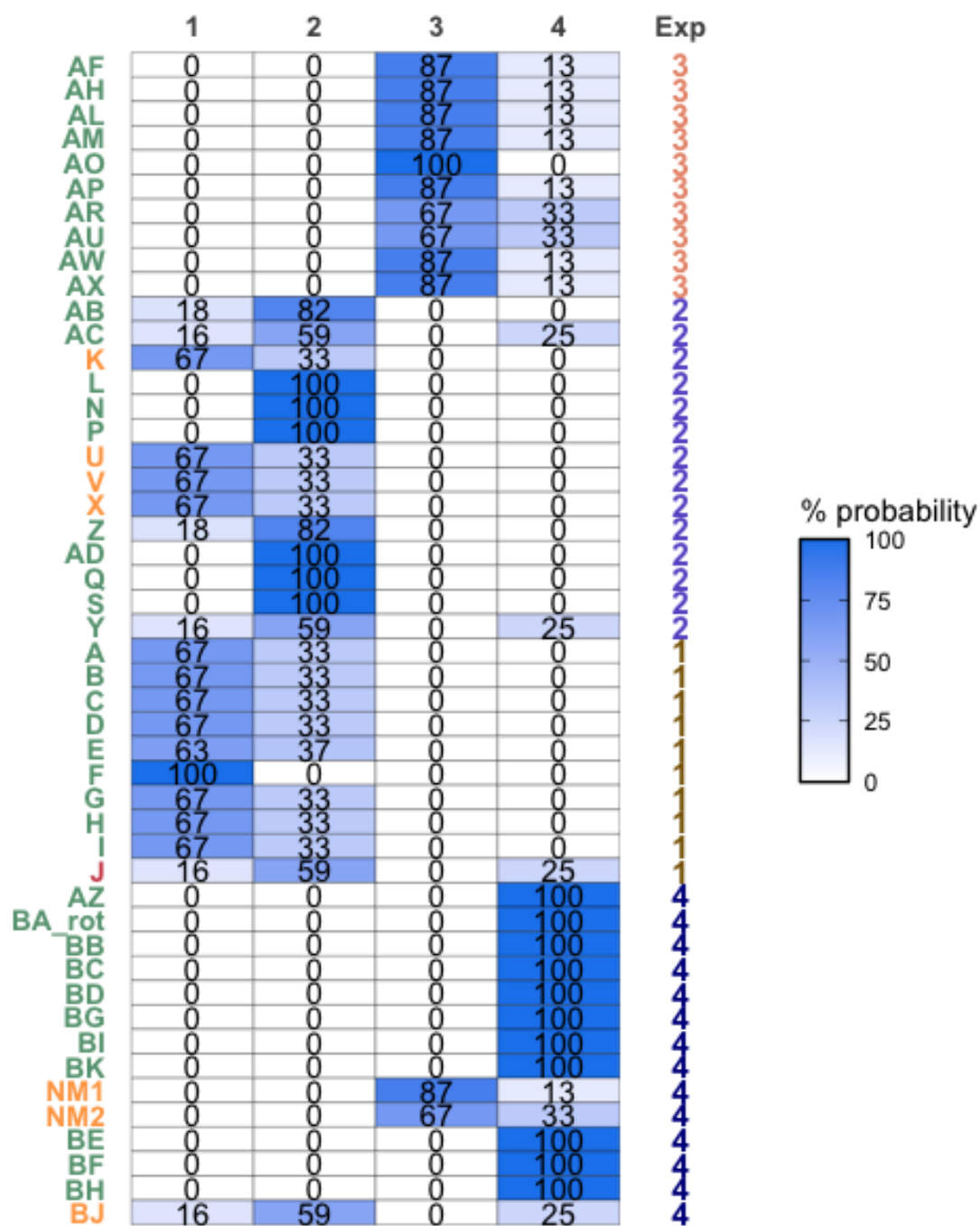
- Accuracy
- False
- Precision
- Size
- True

prop

- 0
- 25
- 50
- 75
- 100

Training Set

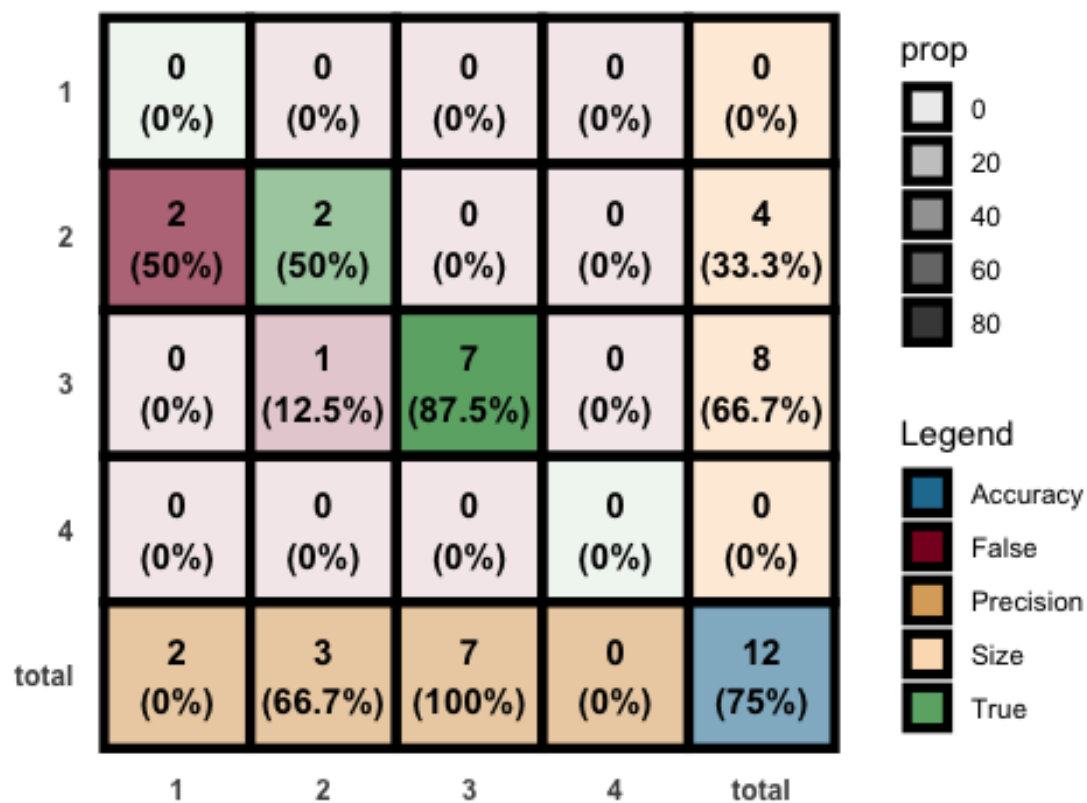
Probability Heatmap



8. 8th Place

Test Set

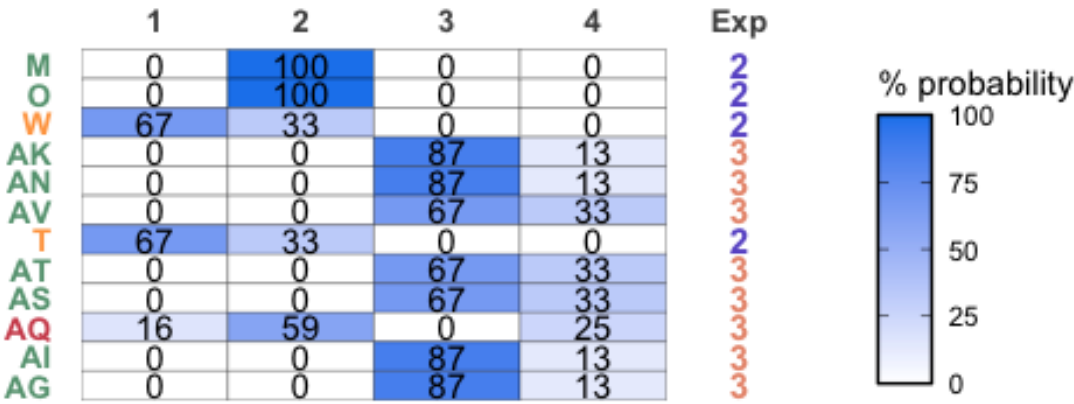
Confusion Matrix



8. 8th Place

Test Set

Probability Heatmap



8. 8th Place

9th Place

Table:

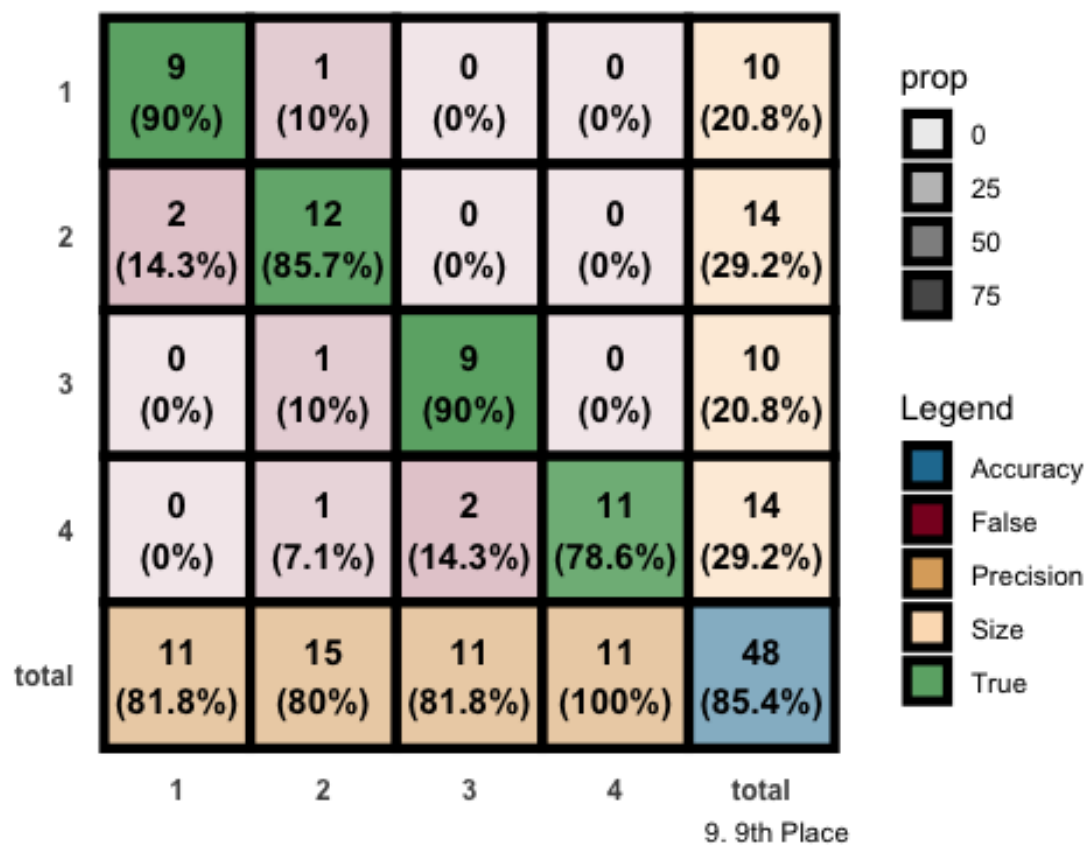
Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
85.42%	0.737

Model Coefficients

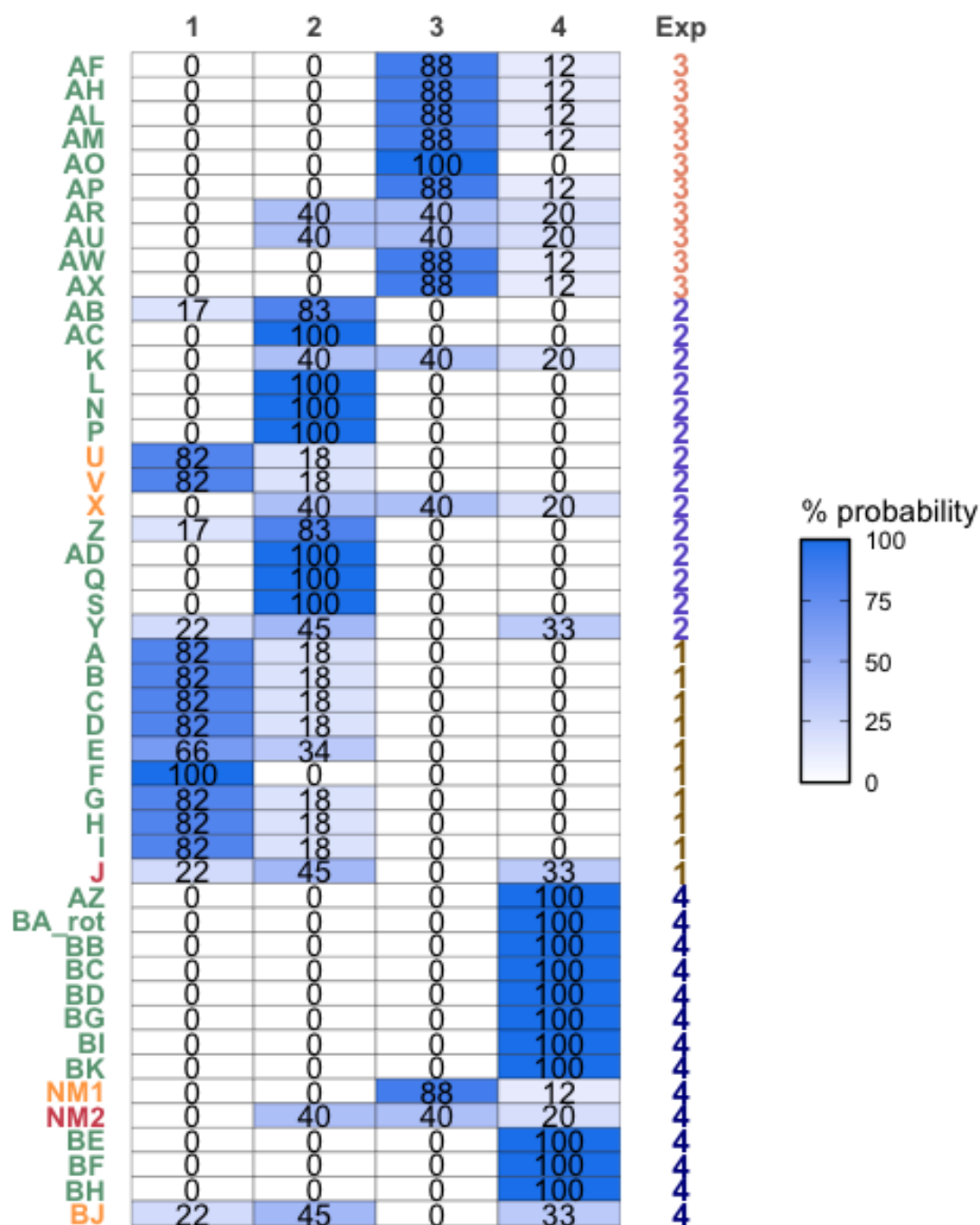
	(Interc ept)	fr_NH0 1	fr_NH 02	fr_NH0 5	fr_NH 11	fr_NH1 2	fr_aldehy de1	fr_pyridi ne1	fr_pyridi ne2
2	38.899 29	5.1817 97	- 38.18 001	- 4.4482 89	- 23.86 882	- 7.4126 37	0.84899 31	- 45.5712 5	39.1562 40
3	13.240 98	30.840 104	- 44.97 720	- 3.3424 63	34.62 293	45.949 696	- 55.6646 292	- 98.2244 9	- 2.99315 3
4	72.322 53	- 28.934 631	- 71.89 964	11.003 202	- 26.40 458	- 33.256 233	- 102.630 9141	- 84.2656 5	- 17.4785 19

Training Set Confusion Matrix



Training Set

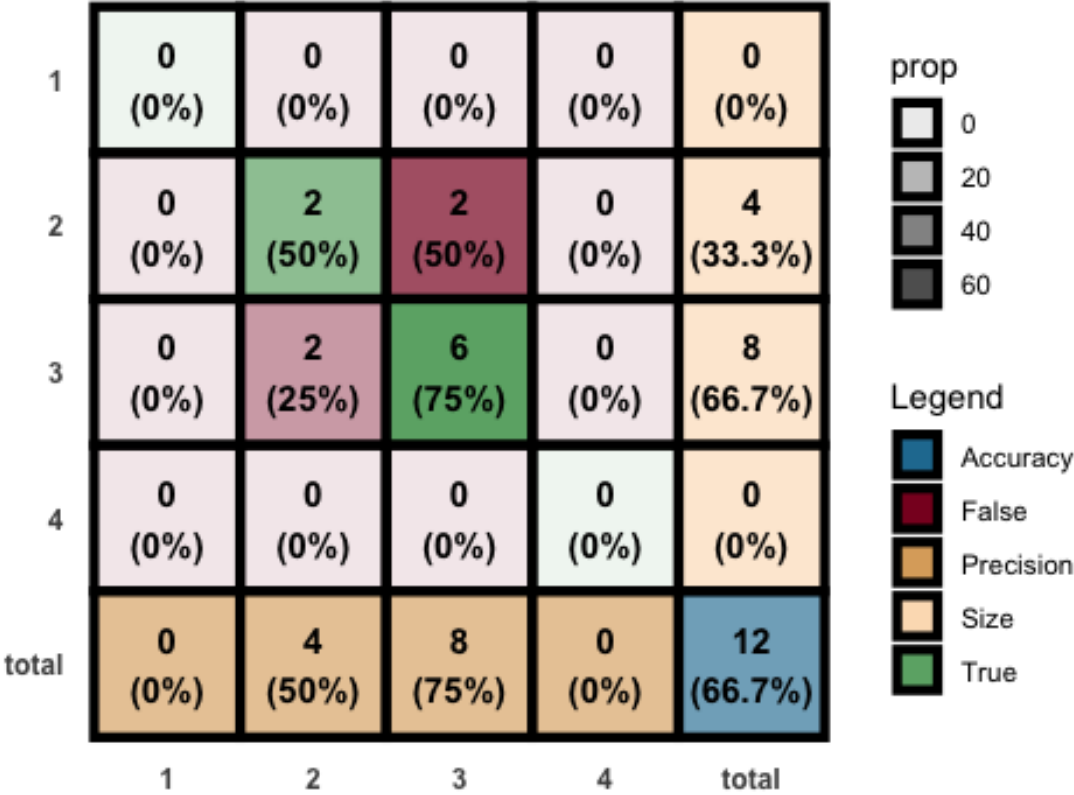
Probability Heatmap



9. 9th Place

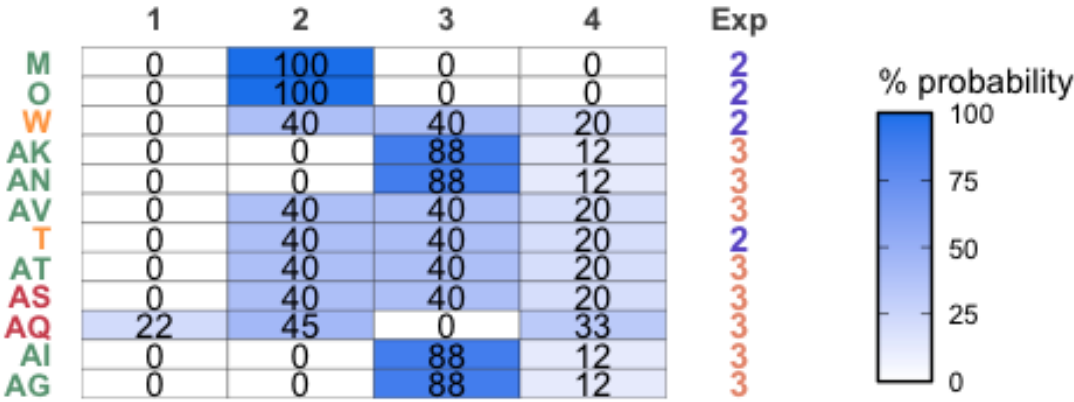
Test Set

Confusion Matrix



Test Set

Probability Heatmap



9. 9th Place

10th Place

Table:

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
83.33%	0.737

Model Coefficients

	(Intercept)	fr_NH01	fr_NH02	fr_NH05	fr_NH21	fr_aldehyde1	fr_amide1	fr_amide2
2	42.46057	-	-	-	-	0.167515	-	-
	2	43.1680	41.1392	6.91136	23.4208		21.4344	10.5610
		2	7	4	6		7	9
3	9.360216	-	-	-	-	-	84.9465	65.7533
		93.6136	25.1656	3.07083	15.4858	103.84517	7	4
		4	8	2	7	1		
4	63.64494	-	-	19.2137	63.3172	-	28.7159	-
	0	92.3608	80.4889	11	4	106.46367	3	41.8190
		5	5			7		1

Training Set

Confusion Matrix

1	8 (80%)	1 (10%)	1 (10%)	0 (0%)	10 (20.8%)
2	4 (28.6%)	10 (71.4%)	0 (0%)	0 (0%)	14 (29.2%)
3	0 (0%)	0 (0%)	10 (100%)	0 (0%)	10 (20.8%)
4	0 (0%)	0 (0%)	2 (14.3%)	12 (85.7%)	14 (29.2%)
total	12 (66.7%)	11 (90.9%)	13 (76.9%)	12 (100%)	48 (83.3%)
	1	2	3	4	total

Legend

	Accuracy
	False
	Precision
	Size
	True

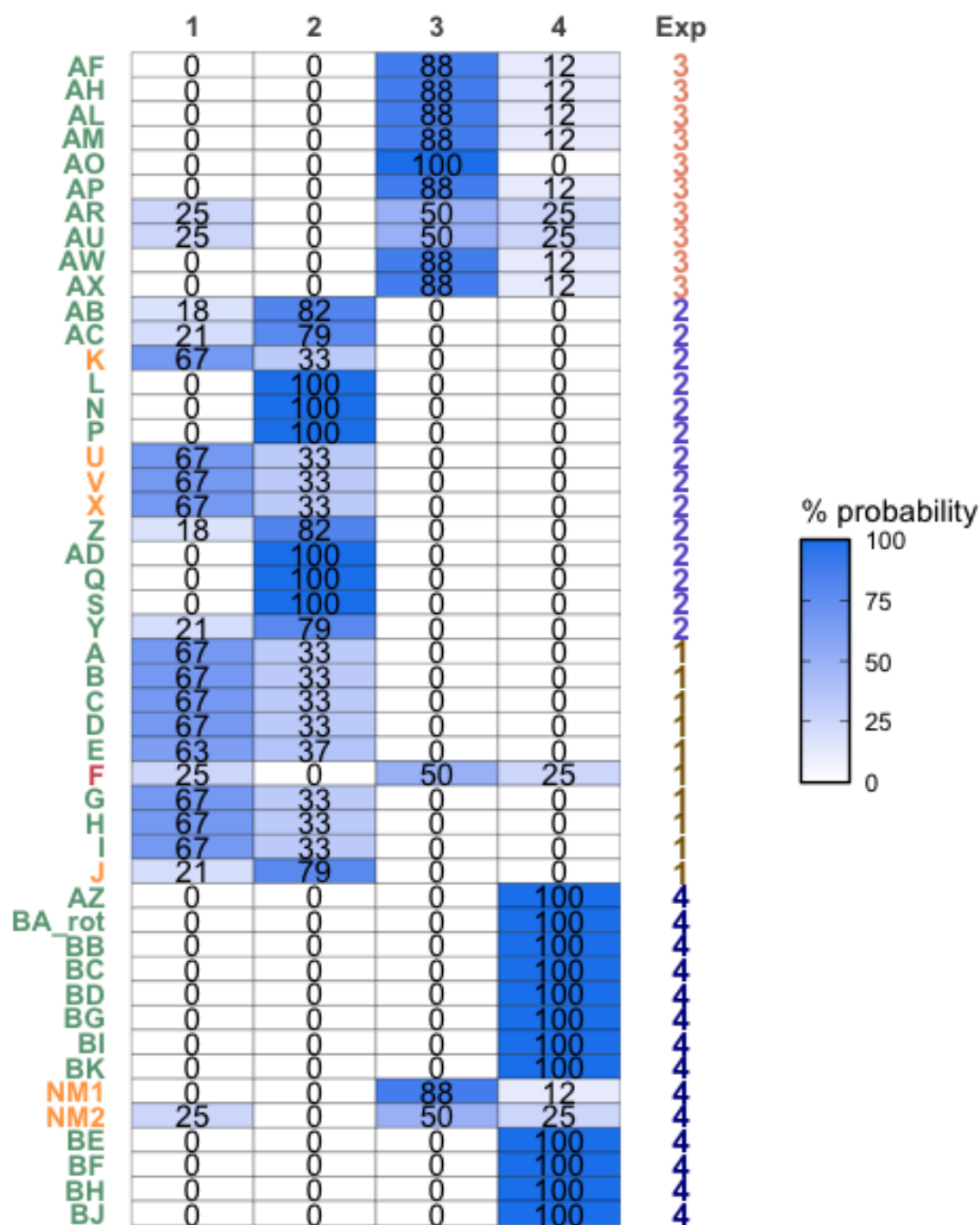
prop

	0
	25
	50
	75
	100

10. 10th Place

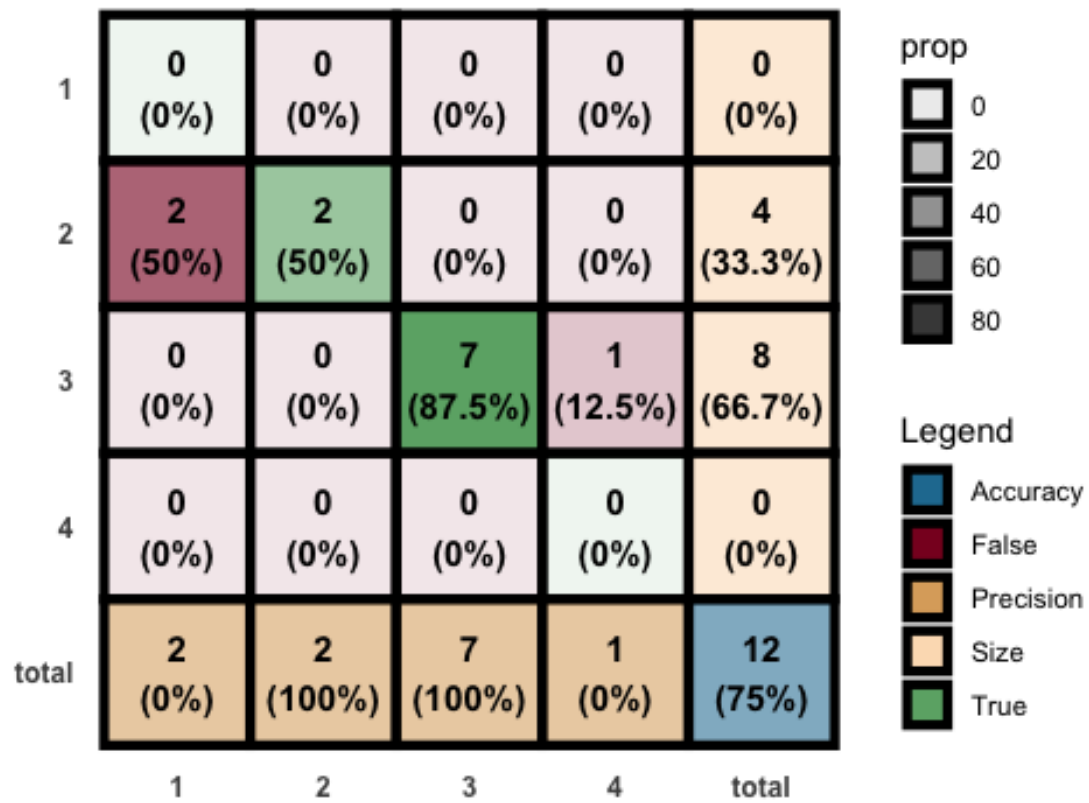
Training Set

Probability Heatmap



Test Set

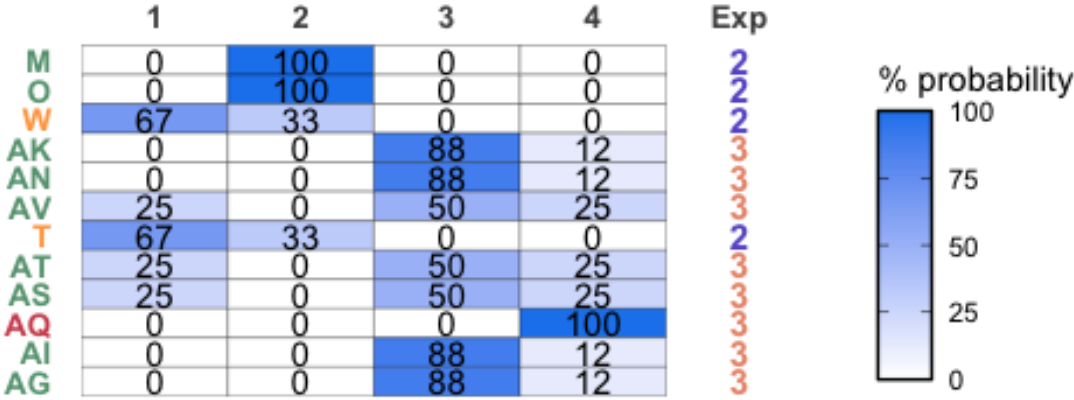
Confusion Matrix



10. 10th Place

Test Set

Probability Heatmap



10. 10th Place