Table of Contents

Published Model **1**
3rd Place 1
Model Analysis - Variable Impact **7**
1. fr_NH0 Removed 7
2. fr_aldehyde Removed 12
3. fr_amide Removed 17
Section Results Summary **27**

CH Activation - Best Model Feature Analysis

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
`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

Published Model

3rd Place

Table:

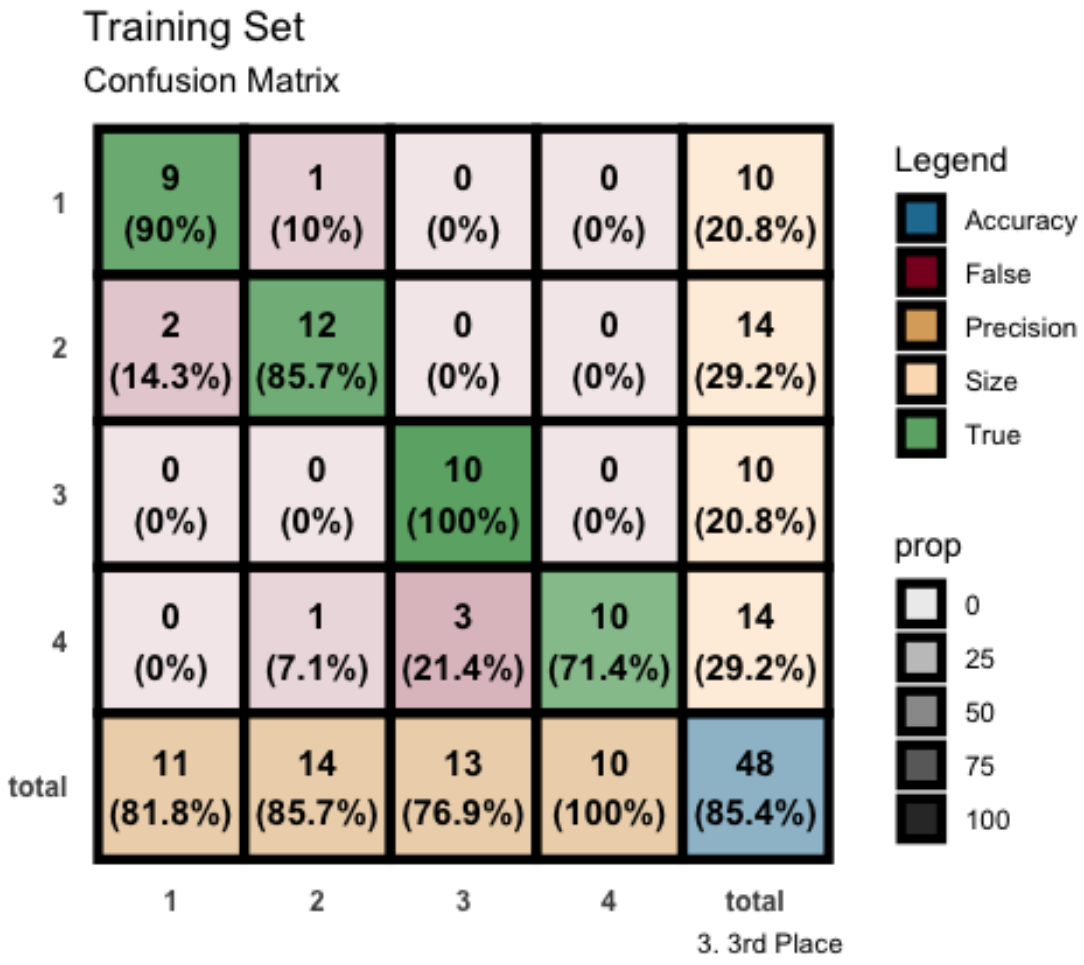
Full Model Stats - Overall Accuracy and Pseudo-R2

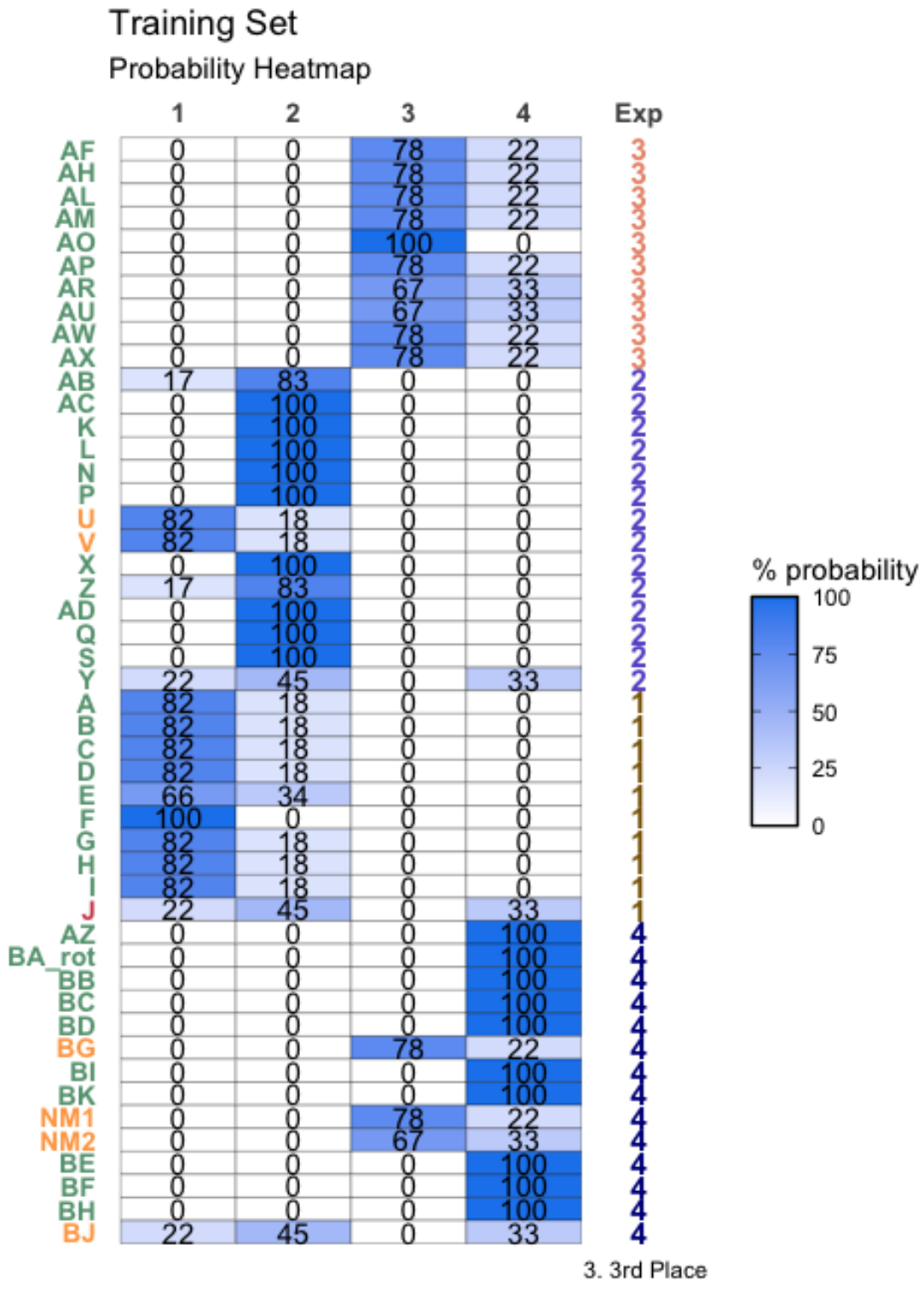
Accuracy	McFadden_R2
85.42%	0.762

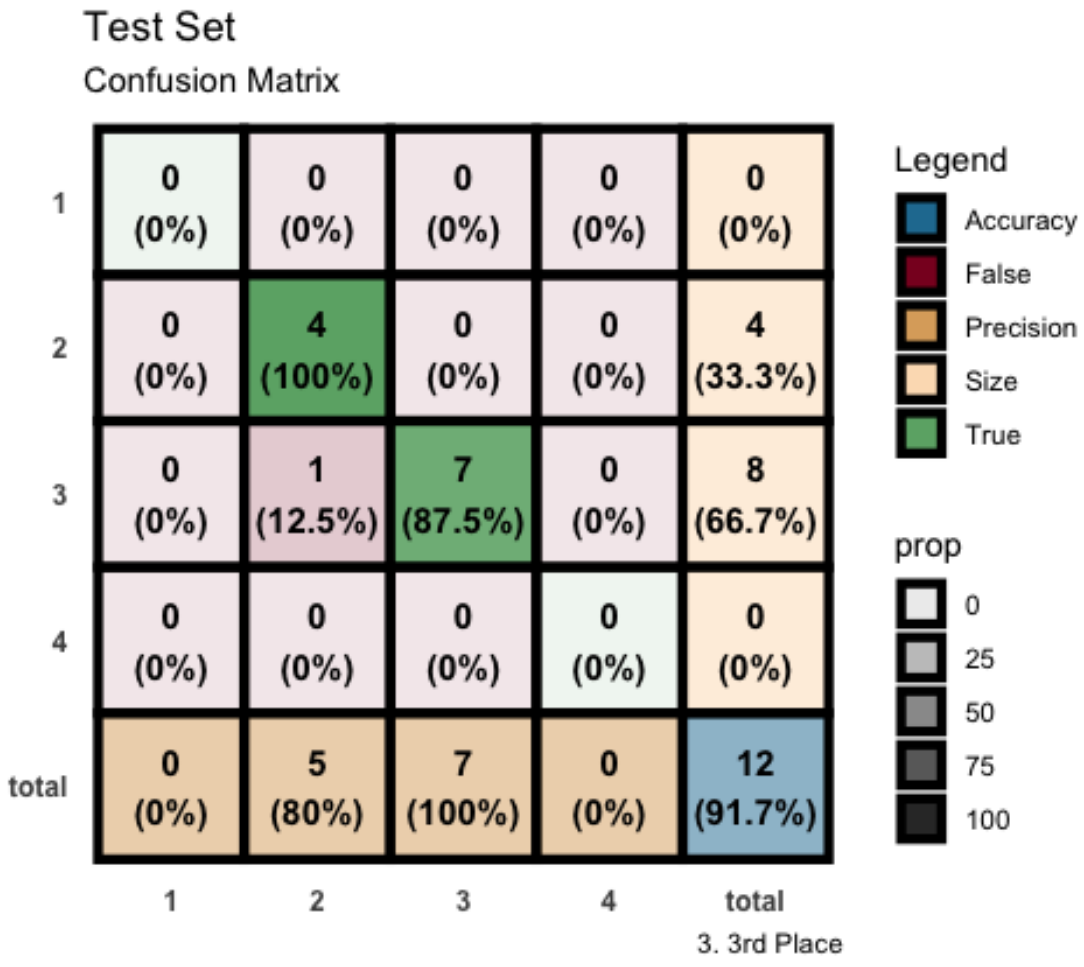
Model Coefficients

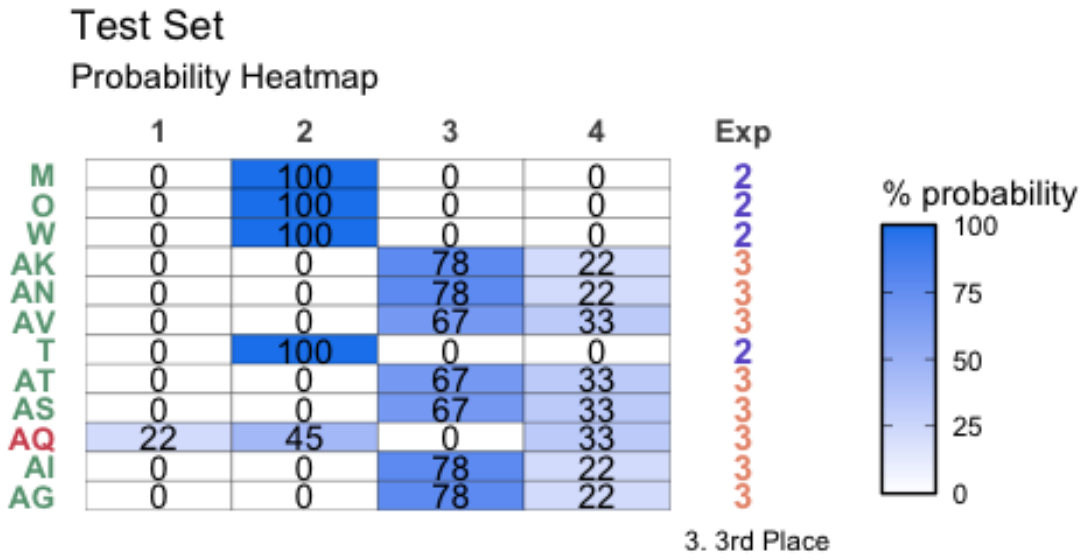
(Intercept)	fr_NH01	fr_NH02	fr_NH05	fr_aldehyde1	fr_amide1	fr_amide2	fr_pyridine1	fr_pyridine2

	(Intercept)	fr_NH01	fr_NH02	fr_NH05	fr_aldehyde1	fr_amide1	fr_amide2	fr_pyridine1	fr_pyridine2
2	77.146276	7.613779	-76.42688	-11.946103	0.8488433	-56.07204	-20.52782	-86.25024	78.008171
3	8.319367	-52.025492	-57.35079	-3.910005	-185.0473453	103.38214	111.78747	-162.83443	-3.618838
4	95.677264	-51.465887	-95.25421	24.534540	-196.7872443	-14.77146	-72.36790	-155.40107	-33.275126


Training Set
Confusion Matrix
1
2
3
4
total
9 (90%)
1 (10%)
0 (0%)
0 (0%)
10 (20.8%)
2 (14.3%)
12 (85.7%)
0 (0%)
0 (0%)
14 (29.2%)
0 (0%)
0 (0%)
10 (100%)
0 (0%)
10 (20.8%)
0 (0%)
1 (7.1%)
3 (21.4%)
10 (71.4%)
14 (29.2%)
11 (81.8%)
14 (85.7%)
13 (76.9%)
10 (100%)
48 (85.4%)
1
2
3
4
total
3. 3rd Place
Legend
Accuracy
False
Precision
Size
True
prop
0
25
50
75
100


Training Set
Probability Heatmap
1
2
3
4
Exp
AF 0 0 78 22 3
AH 0 0 78 22 3
AL 0 0 78 22 3
AM 0 0 78 22 3
AO 0 0 100 0 3
AP 0 0 78 22 3
AR 0 0 67 33 3
AU 0 0 67 33 3
AW 0 0 78 22 3
AX 0 0 78 22 3
AB 17 83 0 0 2
AC 0 100 0 0 2
K 0 100 0 0 2
L 0 100 0 0 2
N 0 100 0 0 2
P 0 100 0 0 2
U 82 18 0 0 2
V 82 18 0 0 2
X 0 100 0 0 2
Z 17 83 0 0 2
AD 0 100 0 0 2
Q 0 100 0 0 2
S 0 100 0 0 2
Y 22 45 0 33 2
A 82 18 0 0 1
B 82 18 0 0 1
C 82 18 0 0 1
D 82 18 0 0 1
E 66 34 0 0 1
F 100 0 0 0 1
G 82 18 0 0 1
H 82 18 0 0 1
I 82 18 0 0 1
J 22 45 0 33 1
AZ 0 0 0 100 4
BA_rot 0 0 0 100 4
BB 0 0 0 100 4
BC 0 0 0 100 4
BD 0 0 0 100 4
BG 0 0 78 22 4
BI 0 0 0 100 4
BK 0 0 0 100 4
NM1 0 0 78 22 4
NM2 0 0 67 33 4
BE 0 0 0 100 4
BF 0 0 0 100 4
BH 0 0 0 100 4
BJ 22 45 0 33 4
% probability
100
75
50
25
0
3. 3rd Place


Test Set
Confusion Matrix
1
2
3
4
total
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
4 (100%)
0 (0%)
0 (0%)
4 (33.3%)
0 (0%)
1 (12.5%)
7 (87.5%)
0 (0%)
8 (66.7%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
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
1
2
3
4
Exp
M 0 100 0 0 2
O 0 100 0 0 2
W 0 100 0 0 2
AK 0 0 78 22 3
AN 0 0 78 22 3
AV 0 0 67 33 3
T 0 100 0 0 2
AT 0 0 67 33 3
AS 0 0 67 33 3
AQ 22 45 0 33 3
AI 0 0 78 22 3
AG 0 0 78 22 3
% probability
100
75
50
25
0
3. 3rd Place

	1	2	3	4	predicted_class
p1	0.00000	0.00000	0	100.00000	4
p2	100.00000	0.00000	0	0.00000	1
p3	21.83496	44.83128	0	33.33376	2
p4	0.00000	100.00000	0	0.00000	2
p5	21.83496	44.83128	0	33.33376	2
p6	21.83496	44.83128	0	33.33376	2
p7	21.83496	44.83128	0	33.33376	2
p8	21.83496	44.83128	0	33.33376	2
p9	21.83496	44.83128	0	33.33376	2
p10	0.00000	100.00000	0	0.00000	2

Model Analysis - Variable Impact

1. fr_NH0 Removed

formula

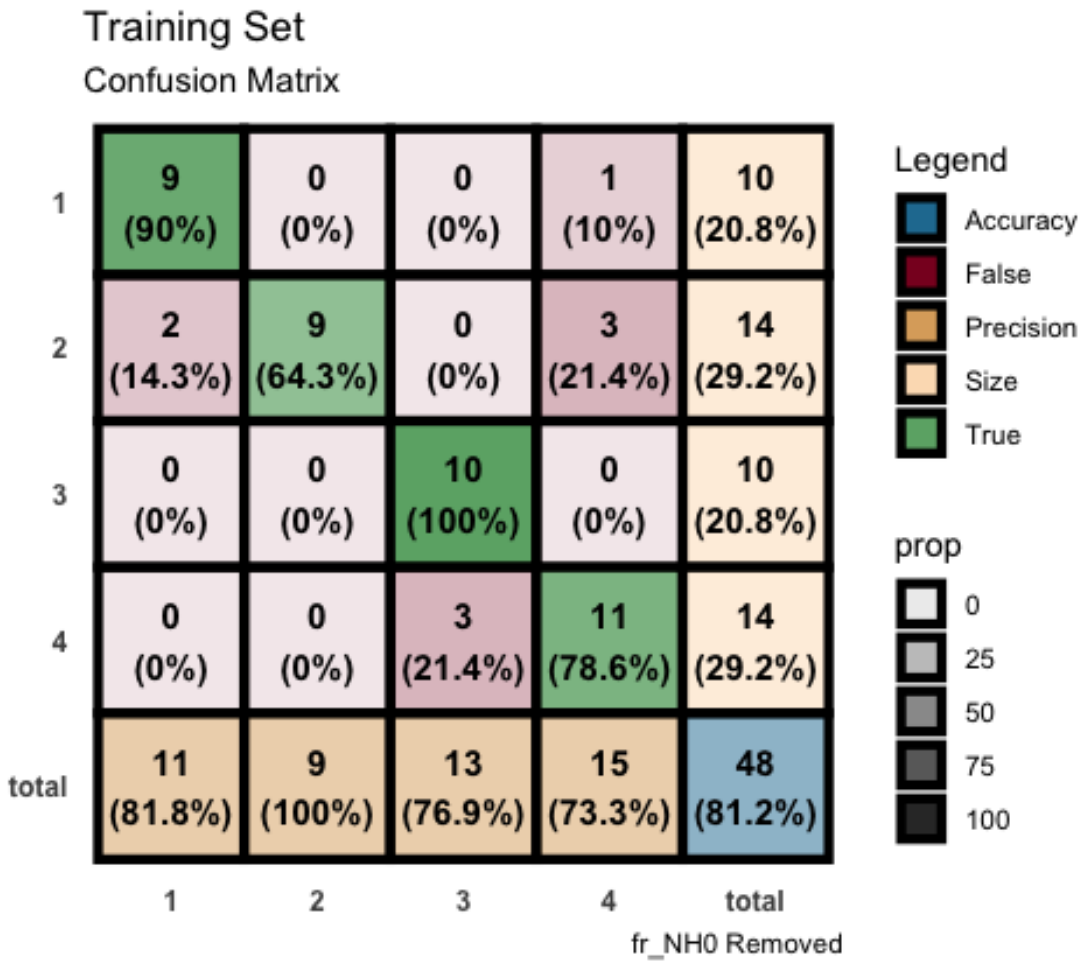
`class ~ fr_aldehyde + fr_amide + fr_pyridine`

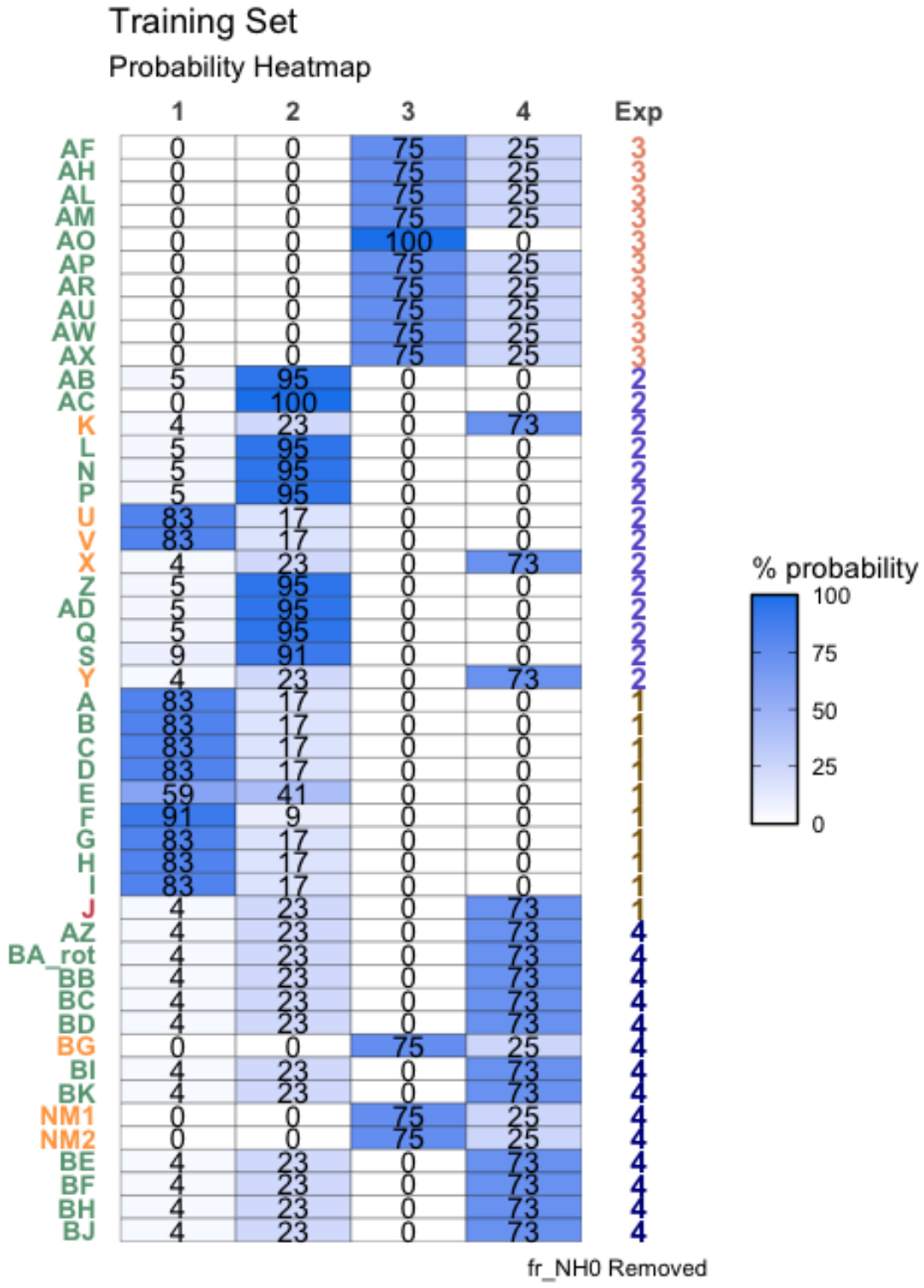
Full Model Stats - Overall Accuracy and Pseudo-R2

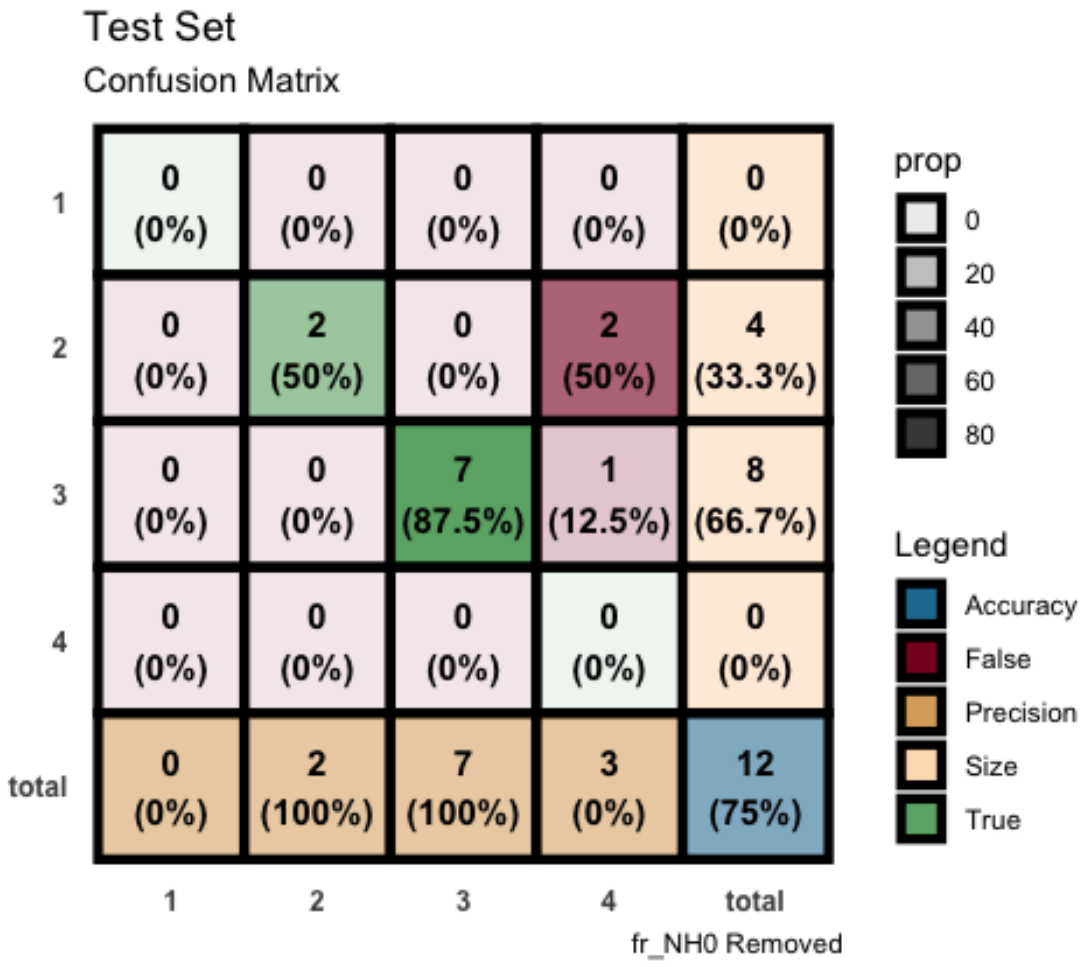
Accuracy	McFadden_R2
81.25%	0.639

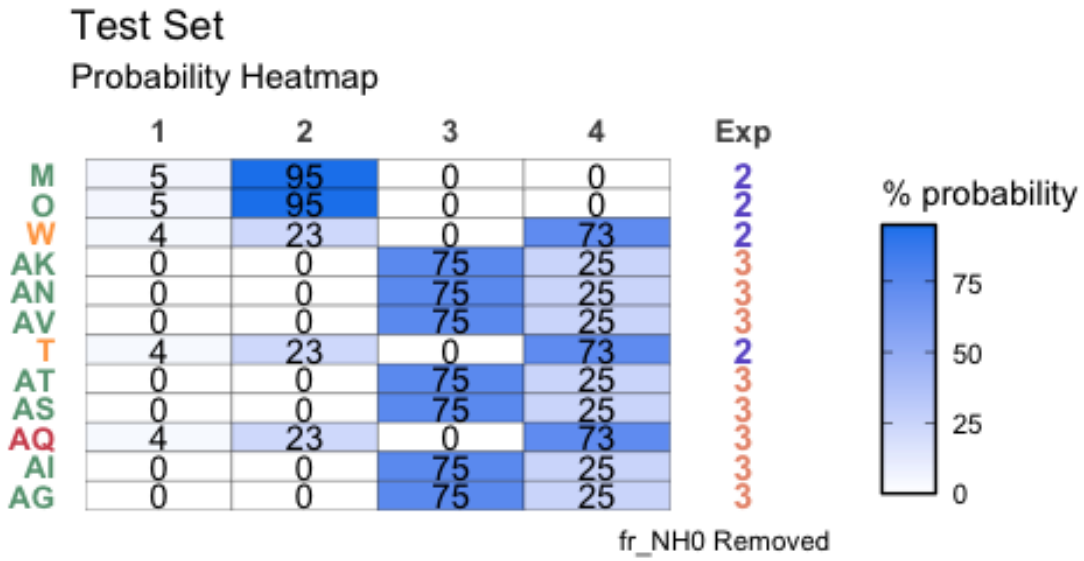
Model Coefficients

	(Intercept)	fr_aldehyde1	fr_amide1	fr_amide2	fr_pyridine1	fr_pyridine2
2	1.761409	1.263614	-0.7059781	-14.79323	-3.373458	23.695461
3	-16.420132	-56.180766	51.2368906	47.97627	-59.021896	-1.446856
4	2.931482	-65.434124	30.7865976	-21.31971	-46.467270	-15.186553


Training Set
Confusion Matrix
1
2
3
4
total
9 (90%)
0 (0%)
0 (0%)
1 (10%)
10 (20.8%)
2 (14.3%)
9 (64.3%)
0 (0%)
3 (21.4%)
14 (29.2%)
0 (0%)
0 (0%)
10 (100%)
0 (0%)
10 (20.8%)
0 (0%)
0 (0%)
3 (21.4%)
11 (78.6%)
14 (29.2%)
11 (81.8%)
9 (100%)
13 (76.9%)
15 (73.3%)
48 (81.2%)
1
2
3
4
total
fr_NH0 Removed
Legend
Accuracy
False
Precision
Size
True
prop
0
25
50
75
100


Training Set
Probability Heatmap
1 2 3 4 Exp
AF 0 0 75 25 3
AH 0 0 75 25 3
AL 0 0 75 25 3
AM 0 0 75 25 3
AO 0 0 100 0 3
AP 0 0 75 25 3
AR 0 0 75 25 3
AU 0 0 75 25 3
AW 0 0 75 25 3
AX 0 0 75 25 3
AB 5 95 0 0 2
AC 0 100 0 0 2
K 4 23 0 73 2
L 5 95 0 0 2
N 5 95 0 0 2
P 5 95 0 0 2
U 83 17 0 0 2
V 83 17 0 0 2
X 4 23 0 73 2
Z 5 95 0 0 2
AD 5 95 0 0 2
Q 5 95 0 0 2
S 9 91 0 0 2
Y 4 23 0 73 2
A 83 17 0 0 1
B 83 17 0 0 1
C 83 17 0 0 1
D 83 17 0 0 1
E 59 41 0 0 1
F 91 9 0 0 1
G 83 17 0 0 1
H 83 17 0 0 1
I 83 17 0 0 1
J 4 23 0 73 1
AZ 4 23 0 73 4
BA_rot 4 23 0 73 4
BB 4 23 0 73 4
BC 4 23 0 73 4
BD 4 23 0 73 4
BG 0 0 75 25 4
BI 4 23 0 73 4
BK 4 23 0 73 4
NM1 0 0 75 25 4
NM2 0 0 75 25 4
BE 4 23 0 73 4
BF 4 23 0 73 4
BH 4 23 0 73 4
BJ 4 23 0 73 4
% probability
100
75
50
25
0
fr_NH0 Removed


Test Set
Confusion Matrix
1
2
3
4
total
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
2 (50%)
0 (0%)
2 (50%)
4 (33.3%)
0 (0%)
0 (0%)
7 (87.5%)
1 (12.5%)
8 (66.7%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
2 (100%)
7 (100%)
3 (0%)
12 (75%)
1
2
3
4
total
fr_NH0 Removed
prop
0
20
40
60
80
Legend
Accuracy
False
Precision
Size
True


Test Set
Probability Heatmap
1
2
3
4
Exp
M 5 95 0 0 2
O 5 95 0 0 2
W 4 23 0 73 2
AK 0 0 75 25 3
AN 0 0 75 25 3
AV 0 0 75 25 3
T 4 23 0 73 2
AT 0 0 75 25 3
AS 0 0 75 25 3
AQ 4 23 0 73 3
AI 0 0 75 25 3
AG 0 0 75 25 3
% probability
75
50
25
0
fr_NH0 Removed

2. fr_aldehyde Removed

formula

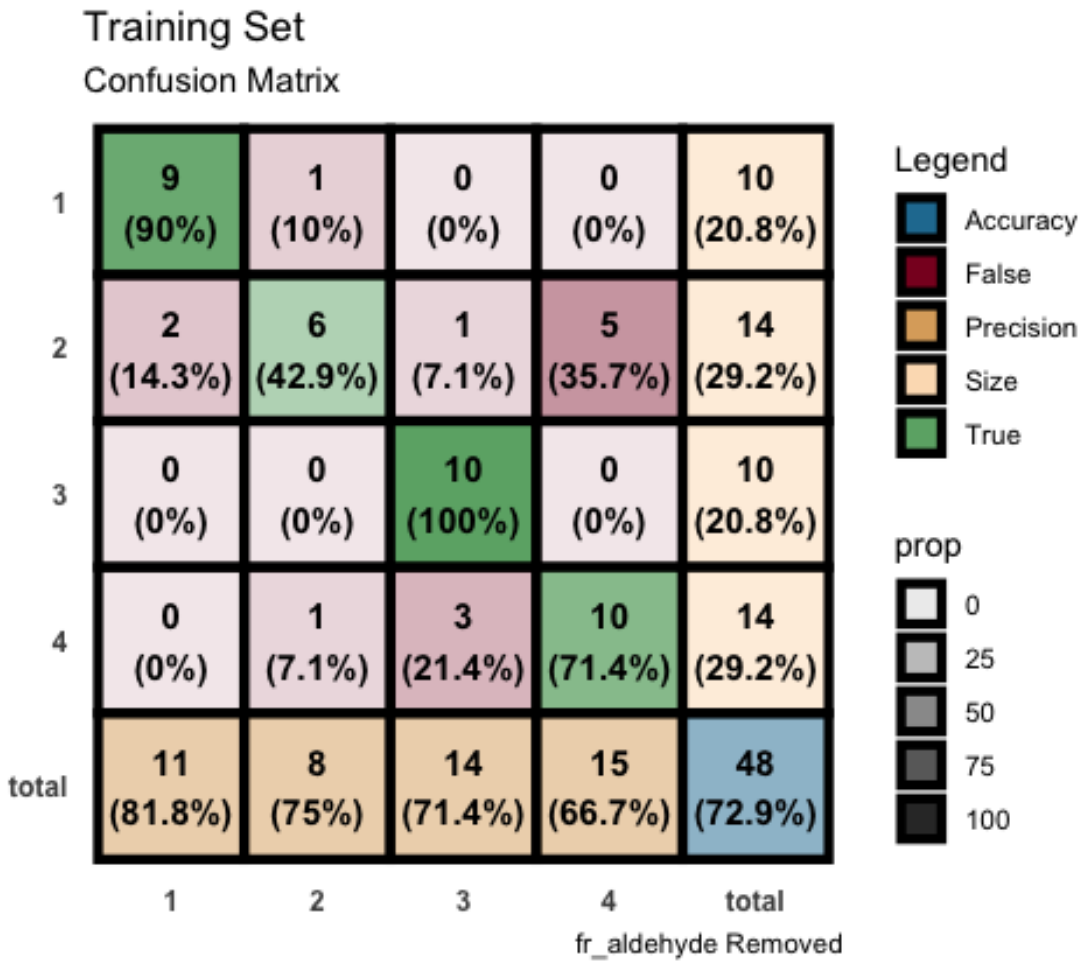
```
class ~ fr_NH0 + fr_amide + fr_pyridine
```

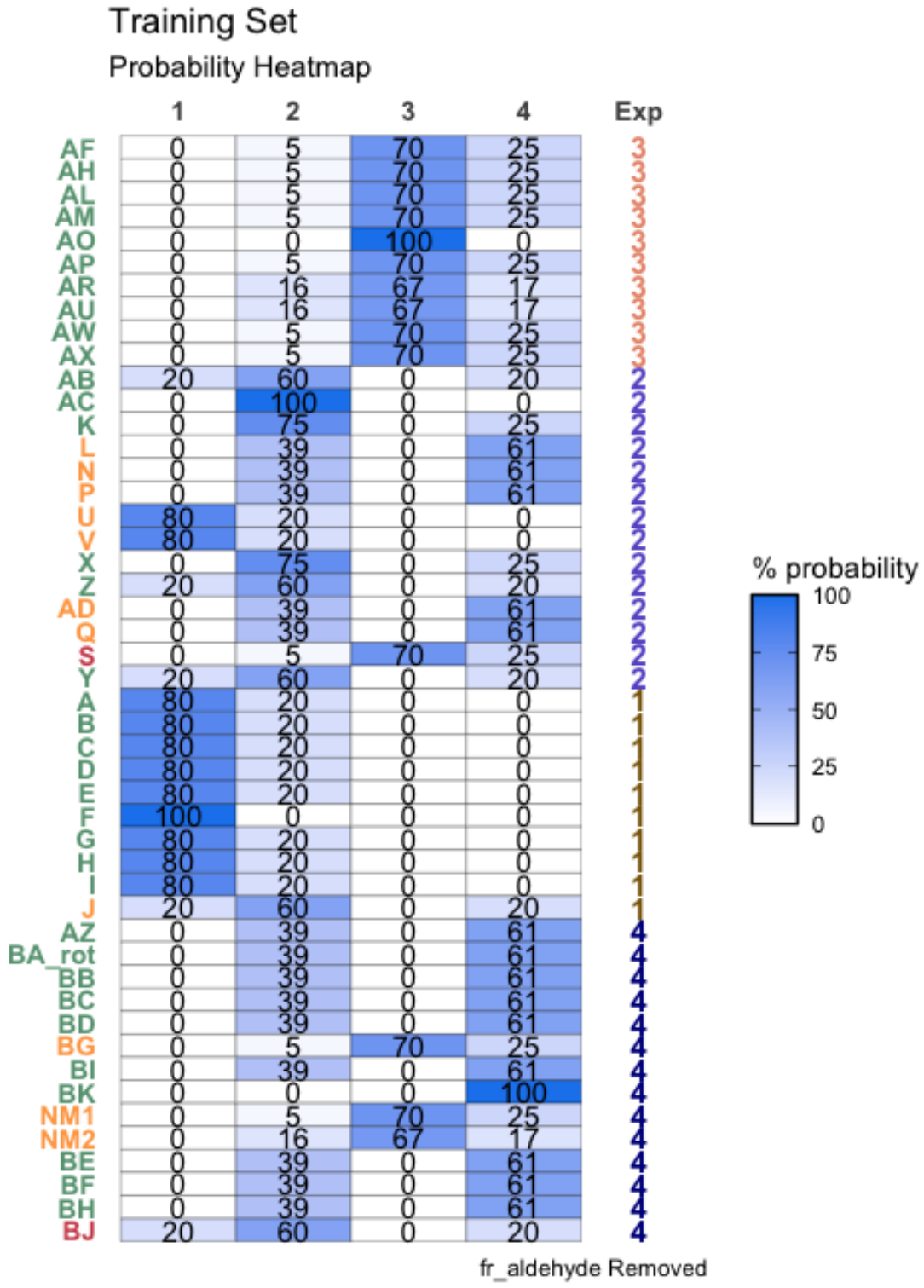
Full Model Stats - Overall Accuracy and Pseudo-R2

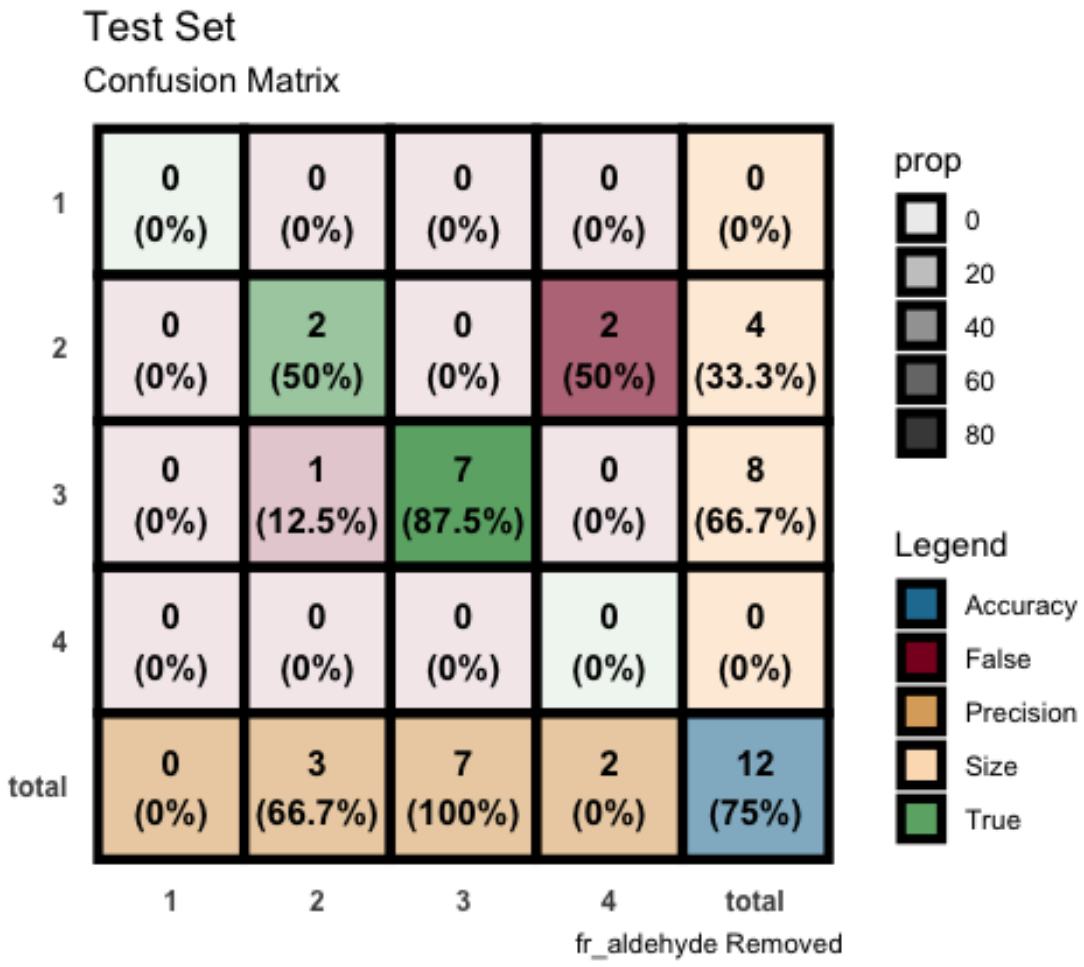
Accuracy	McFadden_R2
72.92%	0.54

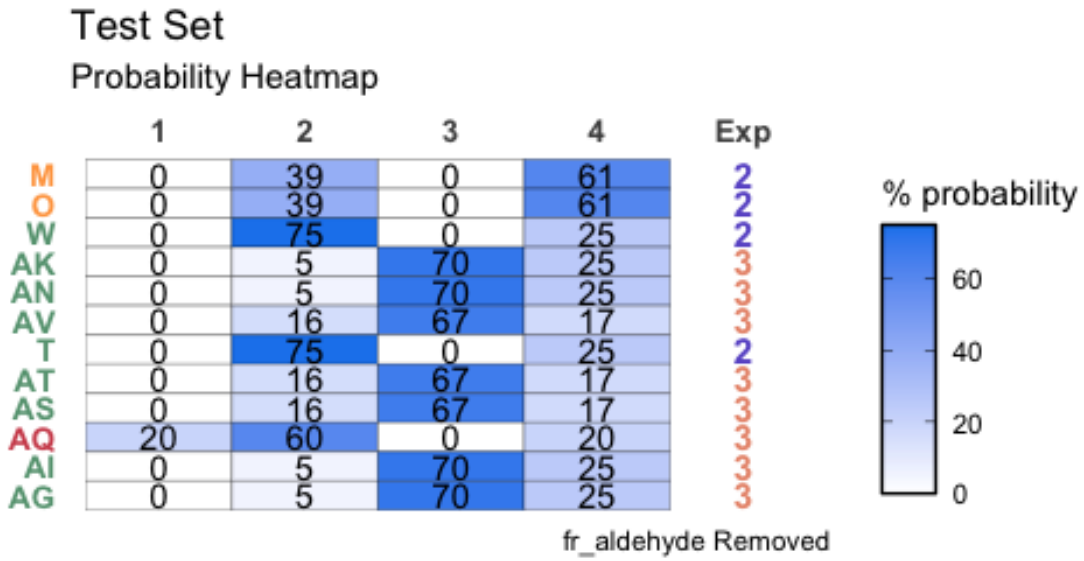
Model Coefficients

	(Intercept)	fr_NH01	fr_NH02	fr_NH05	fr_amide1	fr_amide2	fr_pyridine1	fr_pyridine2
2	34.562388	-5.124864	-33.46378	-13.301597	-14.27077	-18.96640	-30.82382	22.846609
3	-4.246795	-6.345393	-15.30382	-1.365502	27.16148	47.31291	-45.92722	-1.138872
4	35.000120	-6.682834	-35.00013	18.090948	-13.11828	-21.08784	-52.78198	-10.537996


Training Set
Confusion Matrix
9 (90%)
1 (10%)
0 (0%)
0 (0%)
10 (20.8%)
2 (14.3%)
6 (42.9%)
1 (7.1%)
5 (35.7%)
14 (29.2%)
0 (0%)
0 (0%)
10 (100%)
0 (0%)
10 (20.8%)
0 (0%)
1 (7.1%)
3 (21.4%)
10 (71.4%)
14 (29.2%)
11 (81.8%)
8 (75%)
14 (71.4%)
15 (66.7%)
48 (72.9%)
1
2
3
4
total
fr_aldehyde Removed
Legend
Accuracy
False
Precision
Size
True
prop
0
25
50
75
100


Training Set
Probability Heatmap
1
2
3
4
Exp
AF 0 5 70 25 3
AH 0 5 70 25 3
AL 0 5 70 25 3
AM 0 5 70 25 3
AO 0 0 100 0 3
AP 0 5 70 25 3
AR 0 16 67 17 3
AU 0 16 67 17 3
AW 0 5 70 25 3
AX 0 5 70 25 3
AB 20 60 0 20 2
AC 0 100 0 0 2
K 0 75 0 25 2
L 0 39 0 61 2
N 0 39 0 61 2
P 0 39 0 61 2
U 80 20 0 0 2
V 80 20 0 0 2
X 0 75 0 25 2
Z 20 60 0 20 2
AD 0 39 0 61 2
Q 0 39 0 61 2
S 0 5 70 25 2
Y 20 60 0 20 2
A 80 20 0 0 1
B 80 20 0 0 1
C 80 20 0 0 1
D 80 20 0 0 1
E 80 20 0 0 1
F 100 0 0 0 1
G 80 20 0 0 1
H 80 20 0 0 1
I 80 20 0 0 1
J 20 60 0 20 1
AZ 0 39 0 61 4
BA_rot 0 39 0 61 4
BB 0 39 0 61 4
BC 0 39 0 61 4
BD 0 39 0 61 4
BG 0 5 70 25 4
BI 0 39 0 61 4
BK 0 0 0 100 4
NM1 0 5 70 25 4
NM2 0 16 67 17 4
BE 0 39 0 61 4
BF 0 39 0 61 4
BH 0 39 0 61 4
BJ 20 60 0 20 4
% probability
100
75
50
25
0
fr_aldehyde Removed


Test Set
Confusion Matrix
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
2 (50%)
0 (0%)
2 (50%)
4 (33.3%)
0 (0%)
1 (12.5%)
7 (87.5%)
0 (0%)
8 (66.7%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
3 (66.7%)
7 (100%)
2 (0%)
12 (75%)
1
2
3
4
total
fr_aldehyde Removed
prop
0
20
40
60
80
Legend
Accuracy
False
Precision
Size
True


Test Set
Probability Heatmap
1
2
3
4
Exp
M 0 39 0 61 2
O 0 39 0 61 2
W 0 75 0 25 2
AK 0 5 70 25 3
AN 0 5 70 25 3
AV 0 16 67 17 3
T 0 75 0 25 2
AT 0 16 67 17 3
AS 0 16 67 17 3
AQ 20 60 0 20 3
AI 0 5 70 25 3
AG 0 5 70 25 3
% probability
60
40
20
0
fr_aldehyde Removed

3. fr_amide Removed

formula

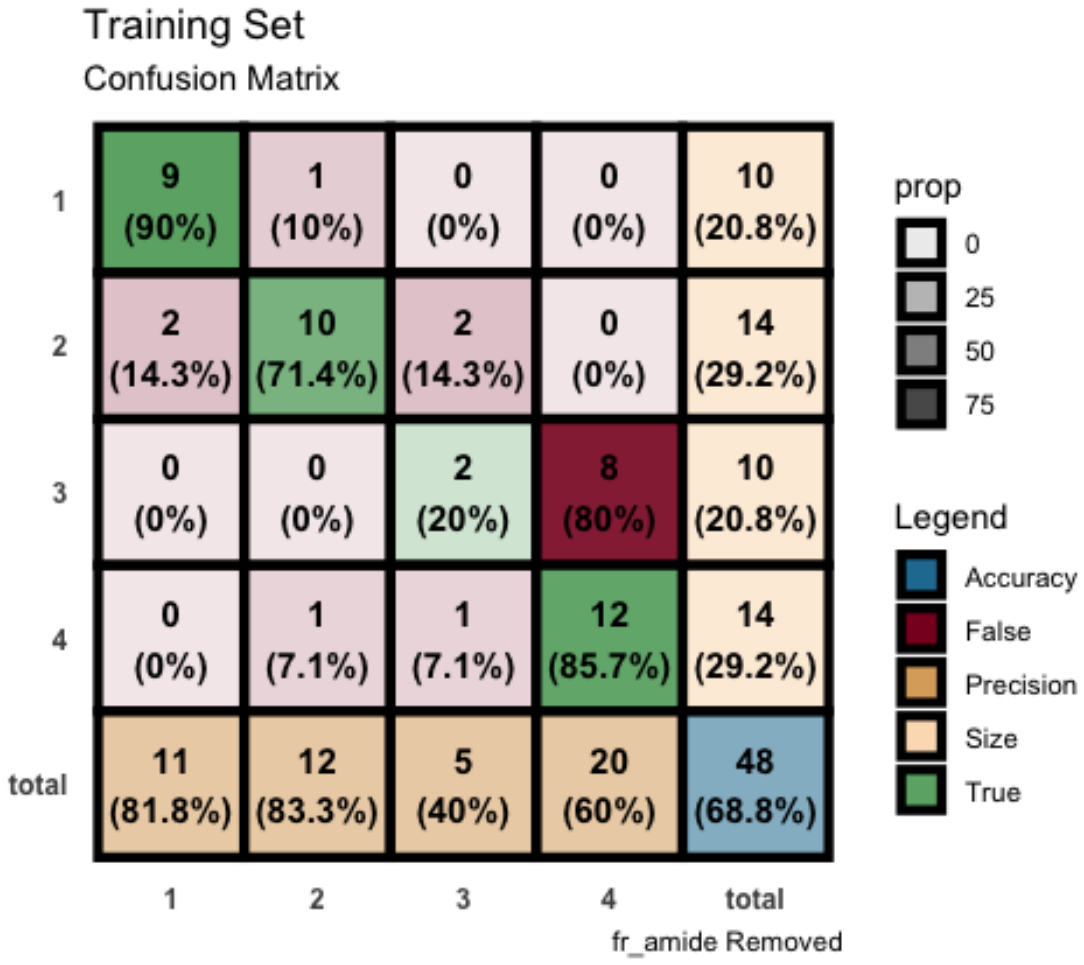
`class ~ fr_NH0 + fr_aldehyde + fr_pyridine`

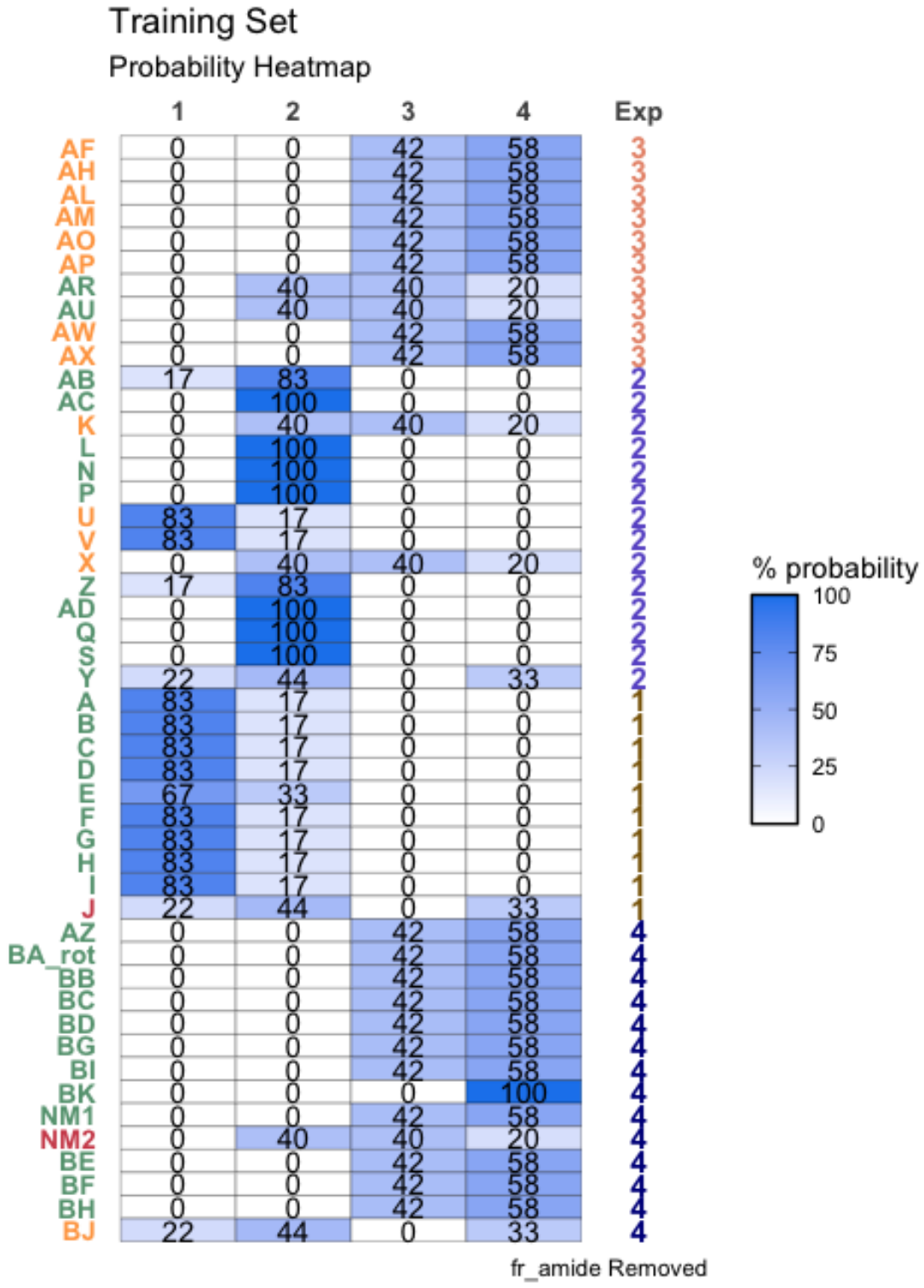
Full Model Stats - Overall Accuracy and Pseudo-R2

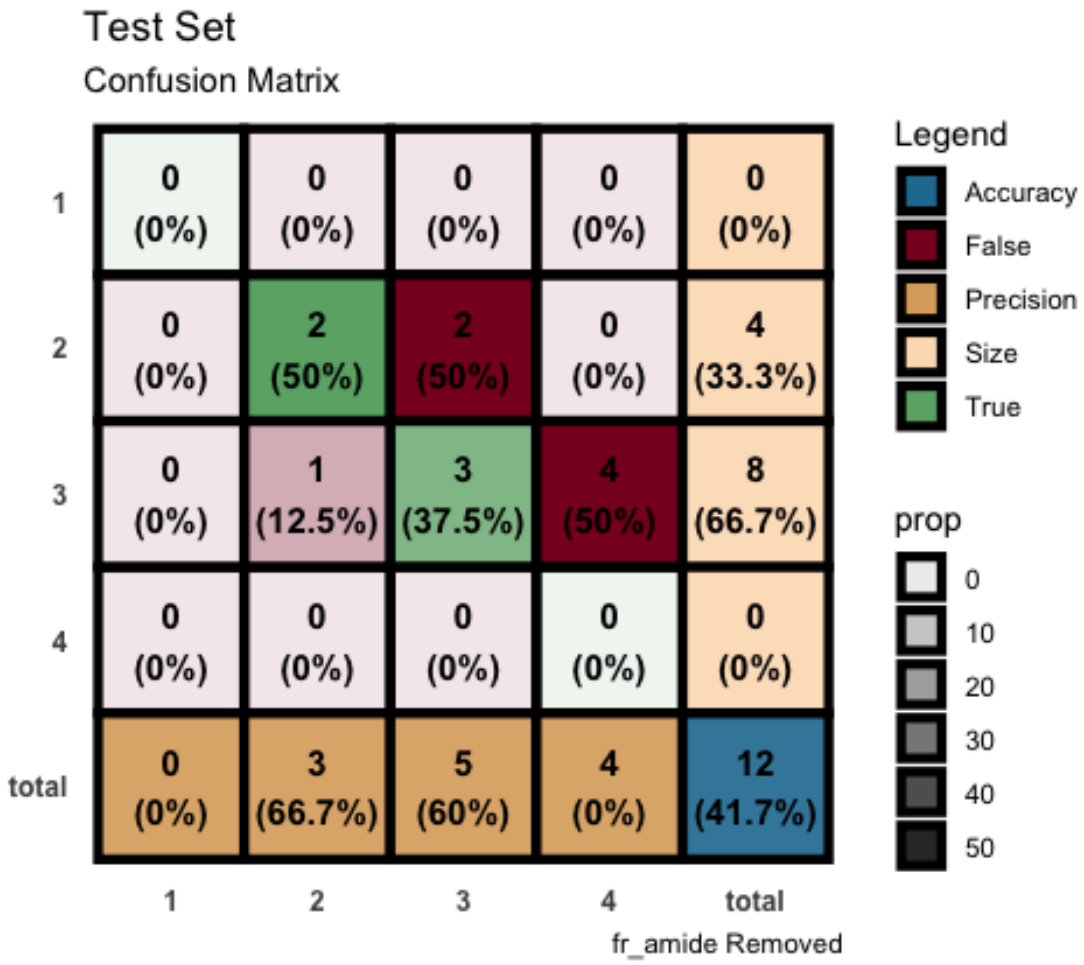
Accuracy	McFadden_R2
68.75%	0.584

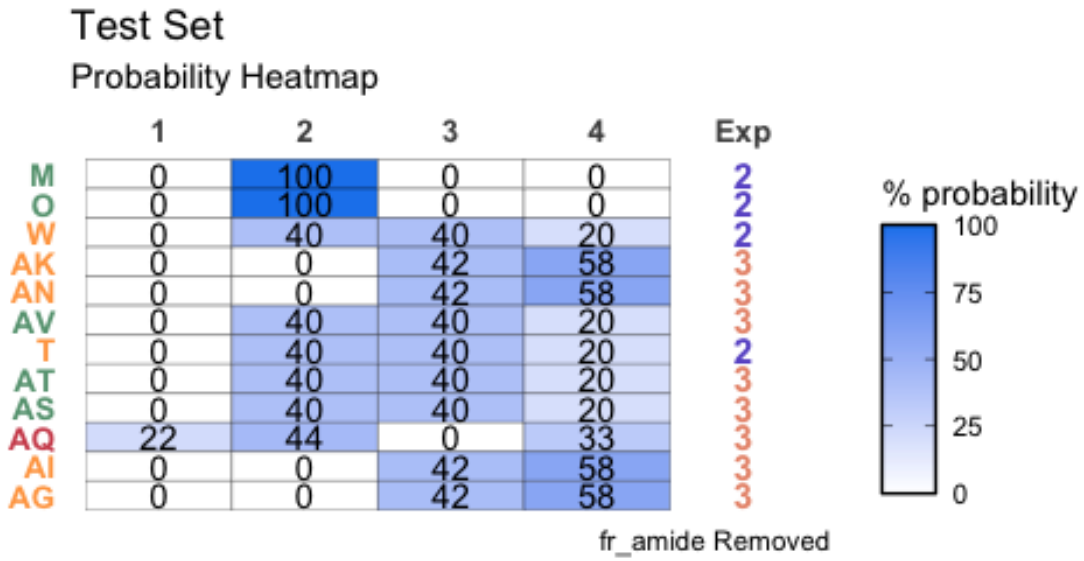
Model Coefficients

	(Intercept)	fr_NH01	fr_NH02	fr_NH05	fr_aldehyde1	fr_pyridine1	fr_pyridine2
2	22.90927	16.194008	-22.21621	-3.56486	0.9163724	-40.71274	29.776318
3	40.56576	-1.462439	-63.22807	-15.86415	-50.3003813	-61.60598	-6.107742
4	40.88420	-2.474048	-40.47883	23.87834	-51.1005312	-58.05919	-15.830804


Training Set
Confusion Matrix
9 (90%)
1 (10%)
0 (0%)
0 (0%)
10 (20.8%)
2 (14.3%)
10 (71.4%)
2 (14.3%)
0 (0%)
14 (29.2%)
0 (0%)
0 (0%)
2 (20%)
8 (80%)
10 (20.8%)
0 (0%)
1 (7.1%)
1 (7.1%)
12 (85.7%)
14 (29.2%)
11 (81.8%)
12 (83.3%)
5 (40%)
20 (60%)
48 (68.8%)
1
2
3
4
total
fr_amide Removed
prop
0
25
50
75
Legend
Accuracy
False
Precision
Size
True


Training Set
Probability Heatmap
1 2 3 4 Exp
AF 0 0 42 58 3
AH 0 0 42 58 3
AL 0 0 42 58 3
AM 0 0 42 58 3
AO 0 0 42 58 3
AP 0 0 42 58 3
AR 0 40 40 20 3
AU 0 40 40 20 3
AW 0 0 42 58 3
AX 0 0 42 58 3
AB 17 83 0 0 2
AC 0 100 0 0 2
K 0 40 40 20 2
L 0 100 0 0 2
N 0 100 0 0 2
P 0 100 0 0 2
U 83 17 0 0 2
V 83 17 0 0 2
X 0 40 40 20 2
Z 17 83 0 0 2
AD 0 100 0 0 2
Q 0 100 0 0 2
S 0 100 0 0 2
Y 22 44 0 33 2
A 83 17 0 0 1
B 83 17 0 0 1
C 83 17 0 0 1
D 83 17 0 0 1
E 67 33 0 0 1
F 83 17 0 0 1
G 83 17 0 0 1
H 83 17 0 0 1
I 83 17 0 0 1
J 22 44 0 33 1
AZ 0 0 42 58 4
BA_rot 0 0 42 58 4
BB 0 0 42 58 4
BC 0 0 42 58 4
BD 0 0 42 58 4
BG 0 0 42 58 4
BI 0 0 42 58 4
BK 0 0 0 100 4
NM1 0 0 42 58 4
NM2 0 40 40 20 4
BE 0 0 42 58 4
BF 0 0 42 58 4
BH 0 0 42 58 4
BJ 22 44 0 33 4
% probability
100
75
50
25
0
fr_amide Removed


Test Set
Confusion Matrix
1
2
3
4
total
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
2 (50%)
2 (50%)
0 (0%)
4 (33.3%)
0 (0%)
1 (12.5%)
3 (37.5%)
4 (50%)
8 (66.7%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
3 (66.7%)
5 (60%)
4 (0%)
12 (41.7%)
1
2
3
4
total
fr_amide Removed
Legend
Accuracy
False
Precision
Size
True
prop
0
10
20
30
40
50


Test Set
Probability Heatmap
1
2
3
4
Exp
M 0 100 0 0 2
O 0 100 0 0 2
W 0 40 40 20 2
AK 0 0 42 58 3
AN 0 0 42 58 3
AV 0 40 40 20 3
T 0 40 40 20 2
AT 0 40 40 20 3
AS 0 40 40 20 3
AQ 22 44 0 33 3
AI 0 0 42 58 3
AG 0 0 42 58 3
% probability
100
75
50
25
0
fr_amide Removed

4.

fr_pyridine Removed

formula

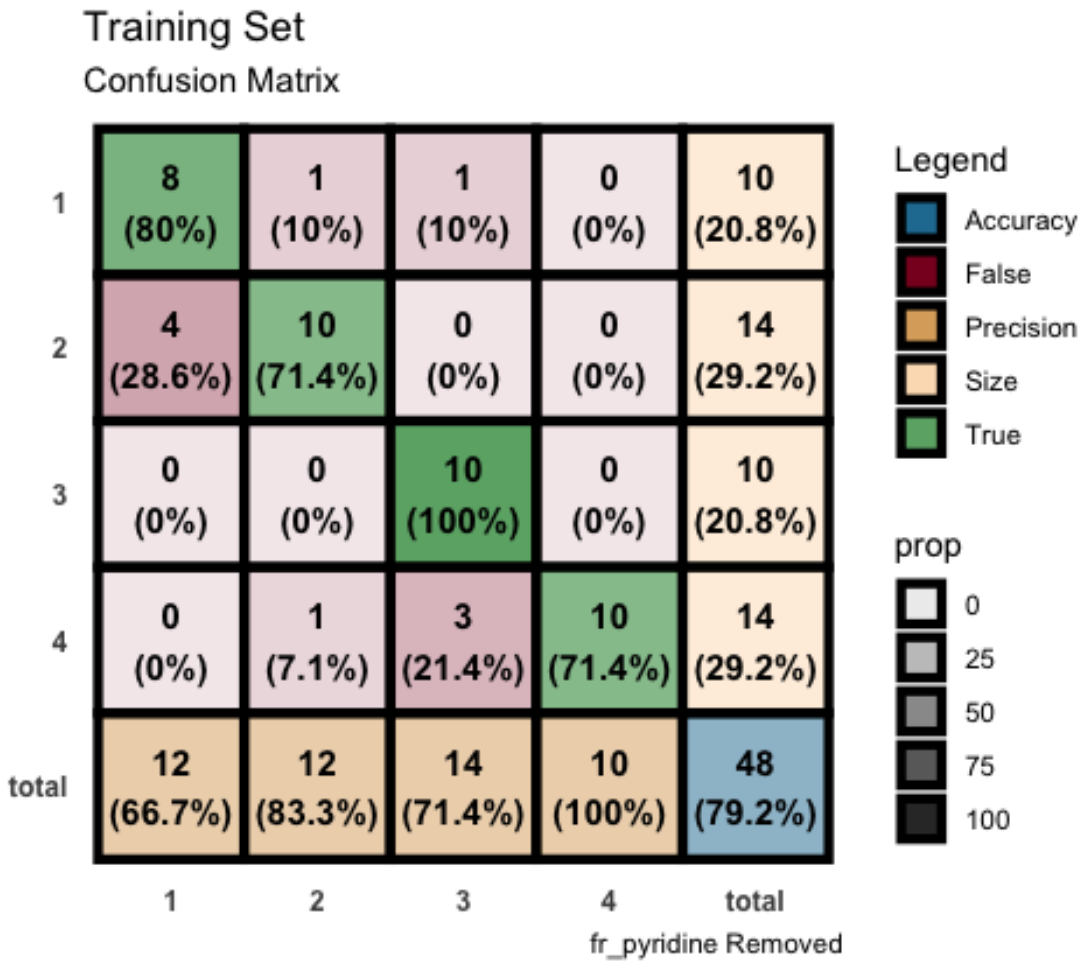
`class ~ fr_NH0 + fr_aldehyde + fr_amide`

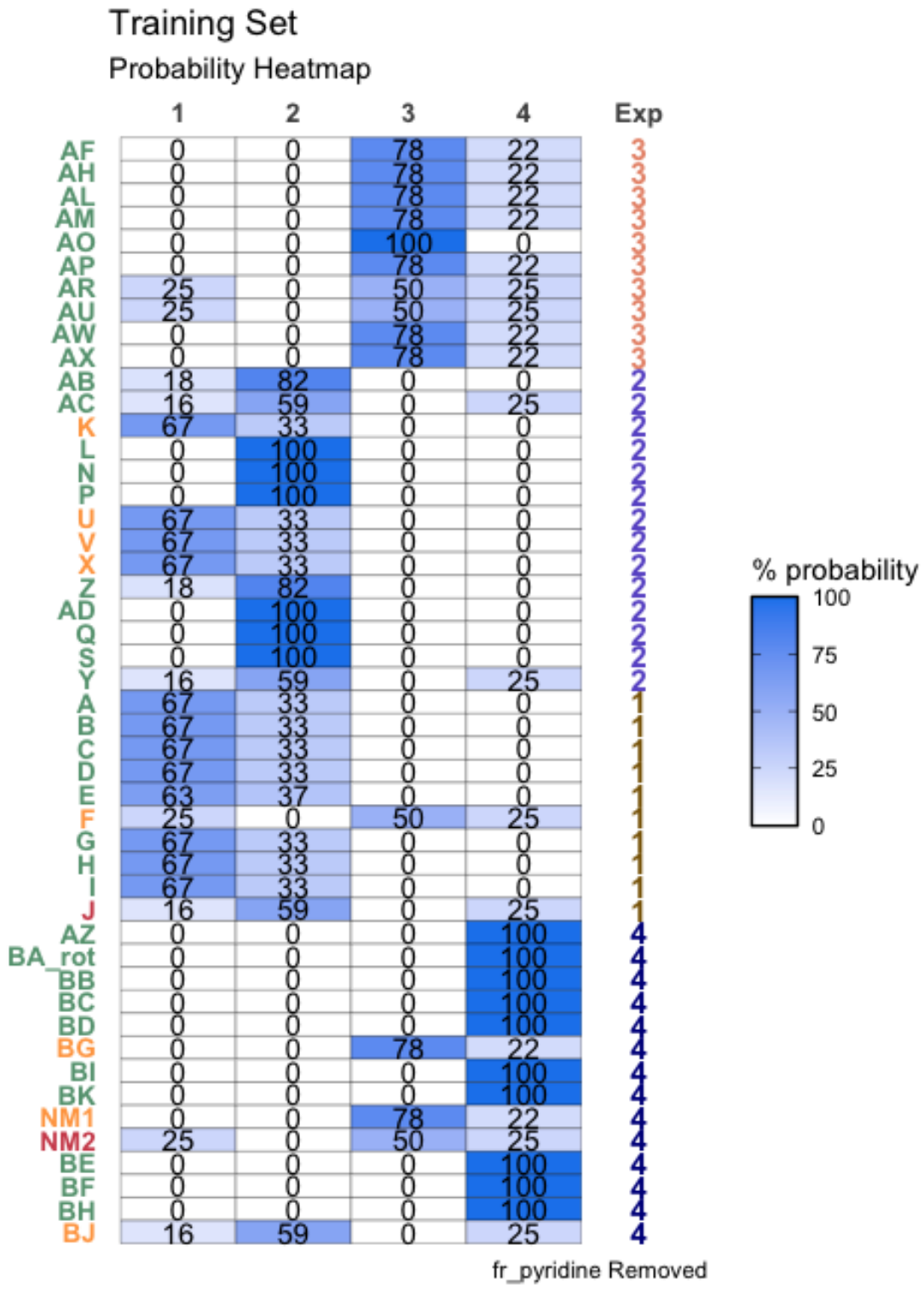
Full Model Stats - Overall Accuracy and Pseudo-R2

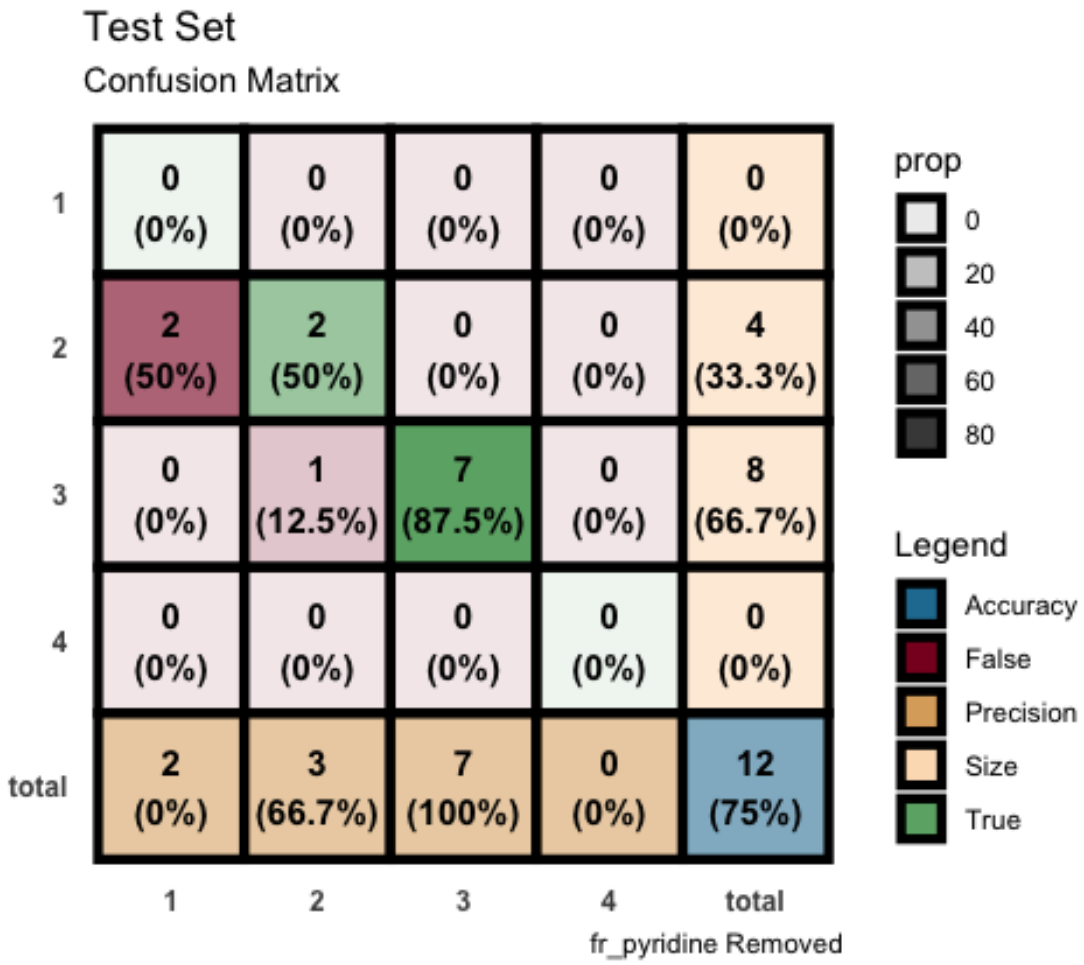
Accuracy	McFadden_R2
79.17%	0.676

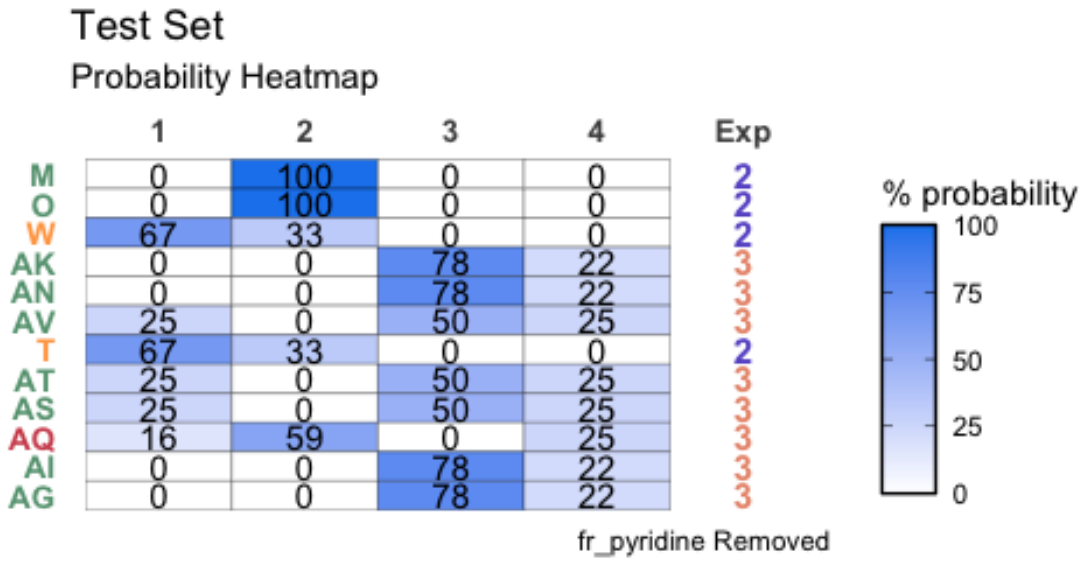
Model Coefficients

	(Intercept)	fr_NH01	fr_NH02	fr_NH05	fr_aldehyde1	fr_amide1	fr_amide2
2	40.004467	-40.71191	-38.68329	-6.192403	0.1676057	-22.71663	-9.730619
3	9.550755	-85.33347	-25.62559	-2.528876	-95.4630660	76.47584	60.011119
4	58.940766	-84.77387	-58.48177	16.210552	-104.9750111	25.83309	-36.687190


Training Set
Confusion Matrix
8 (80%)
1 (10%)
1 (10%)
0 (0%)
10 (20.8%)
4 (28.6%)
10 (71.4%)
0 (0%)
0 (0%)
14 (29.2%)
0 (0%)
0 (0%)
10 (100%)
0 (0%)
10 (20.8%)
0 (0%)
1 (7.1%)
3 (21.4%)
10 (71.4%)
14 (29.2%)
12 (66.7%)
12 (83.3%)
14 (71.4%)
10 (100%)
48 (79.2%)
1
2
3
4
total
fr_pyridine Removed
Legend
Accuracy
False
Precision
Size
True
prop
0
25
50
75
100


Training Set
Probability Heatmap
1
2
3
4
Exp
AF 0 0 78 22 3
AH 0 0 78 22 3
AL 0 0 78 22 3
AM 0 0 78 22 3
AO 0 0 100 0 3
AP 0 0 78 22 3
AR 25 0 50 25 3
AU 25 0 50 25 3
AW 0 0 78 22 3
AX 0 0 78 22 3
AB 18 82 0 0 2
AC 16 59 0 25 2
K 67 33 0 0 2
L 0 100 0 0 2
N 0 100 0 0 2
P 0 100 0 0 2
U 67 33 0 0 2
V 67 33 0 0 2
X 67 33 0 0 2
Z 18 82 0 0 2
AD 0 100 0 0 2
Q 0 100 0 0 2
S 0 100 0 0 2
Y 16 59 0 25 2
A 67 33 0 0 1
B 67 33 0 0 1
C 67 33 0 0 1
D 67 33 0 0 1
E 63 37 0 0 1
F 25 0 50 25 1
G 67 33 0 0 1
H 67 33 0 0 1
I 67 33 0 0 1
J 16 59 0 25 1
AZ 0 0 0 100 4
BA_rot 0 0 0 100 4
BB 0 0 0 100 4
BC 0 0 0 100 4
BD 0 0 0 100 4
BG 0 0 78 22 4
BI 0 0 0 100 4
BK 0 0 0 100 4
NM1 0 0 78 22 4
NM2 25 0 50 25 4
BE 0 0 0 100 4
BF 0 0 0 100 4
BH 0 0 0 100 4
BJ 16 59 0 25 4
% probability
100
75
50
25
0
fr_pyridine Removed


Test Set
Confusion Matrix
1
2
3
4
total
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
2 (50%)
2 (50%)
0 (0%)
0 (0%)
4 (33.3%)
0 (0%)
1 (12.5%)
7 (87.5%)
0 (0%)
8 (66.7%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
0 (0%)
2 (0%)
3 (66.7%)
7 (100%)
0 (0%)
12 (75%)
1
2
3
4
total
fr_pyridine Removed
prop
0
20
40
60
80
Legend
Accuracy
False
Precision
Size
True


Test Set
Probability Heatmap
1
2
3
4
Exp
M 0 100 0 0 2
O 0 100 0 0 2
W 67 33 0 0 2
AK 0 0 78 22 3
AN 0 0 78 22 3
AV 25 0 50 25 3
T 67 33 0 0 2
AT 25 0 50 25 3
AS 25 0 50 25 3
AQ 16 59 0 25 3
AI 0 0 78 22 3
AG 0 0 78 22 3
% probability
100
75
50
25
0
fr_pyridine Removed

Section Results Summary

```
Error: object 'model.sum.1.1' not found
```