

Pyridine Phosphination – 3 categories – Best Model Analysis

Table of Contents

Best Model.....	1
6th Place.....	1
Model Analysis - Variable Impact.....	6
1. Dist_(P,C)_P Removed.....	6
2. NPA_6_SM Removed.....	11
3. para_SM Removed.....	16

Best Model

formula	McFadden R2
class ~ para_P + NPA_6_SM + Avg_NPA_SM	0.703
class ~ NPA_6_SM + para_SM + fr_C_O	0.662
class ~ NPA_6_SM + para_SM + fr_C_O_noCOO	0.662
class ~ Dist_(P,C)_P + para_SM + Iso.Polar_SM	0.651
class ~ para_P + NPA_6_SM + NPA_sum_SM	0.649
class ~ Dist_(P,C)_P + NPA_6_SM + para_SM	0.649
class ~ NPA_6_SM + Avg_NPA_SM + para_SM	0.646
class ~ NPA_1_P + Dist_(P,C)_P + para_SM	0.627
class ~ NPA_1_P + para_SM + fr_pyridine	0.626
class ~ NPA_6_SM + NPA_sum_SM + para_SM	0.623

6th Place

Table:

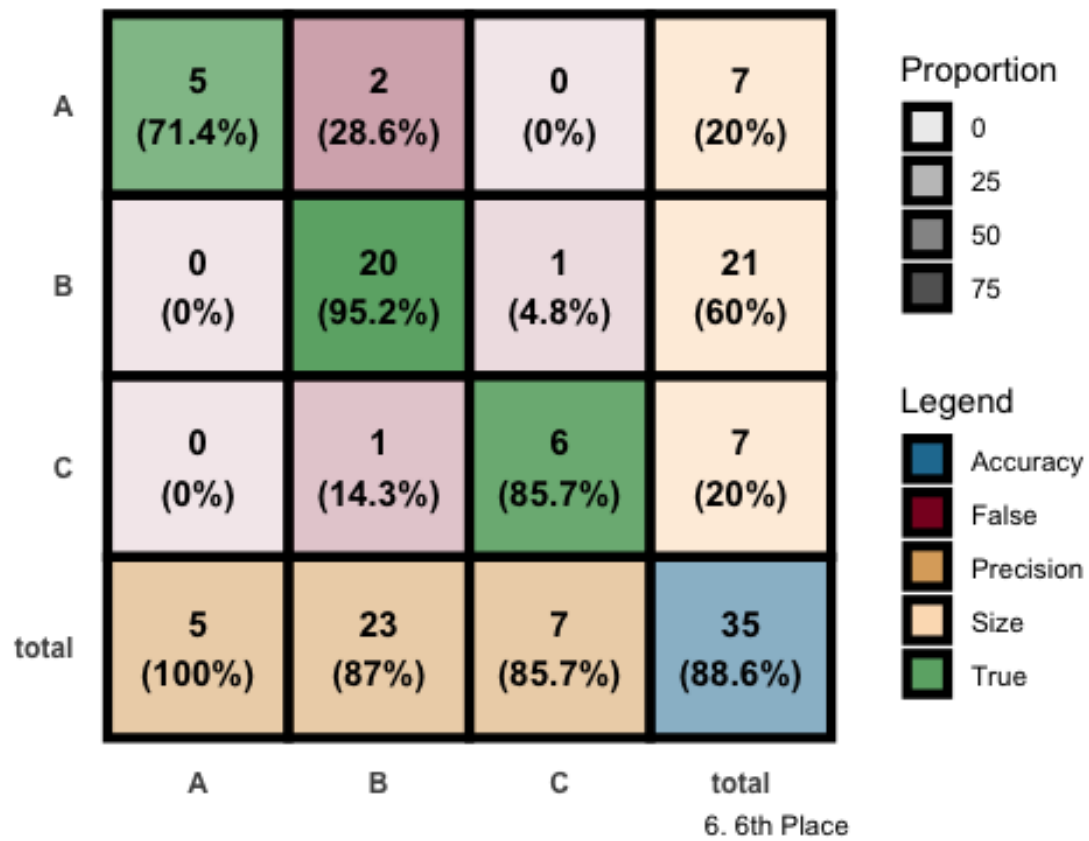
Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
88.57%	0.649

Model Coefficients

	x
Dist_(P,C)_P	-181.4237850
NPA_6_SM	33.1536447
para_SM	0.2071308

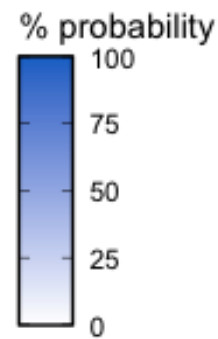
Training Set Confusion Matrix



Training Set

Probability Heatmap

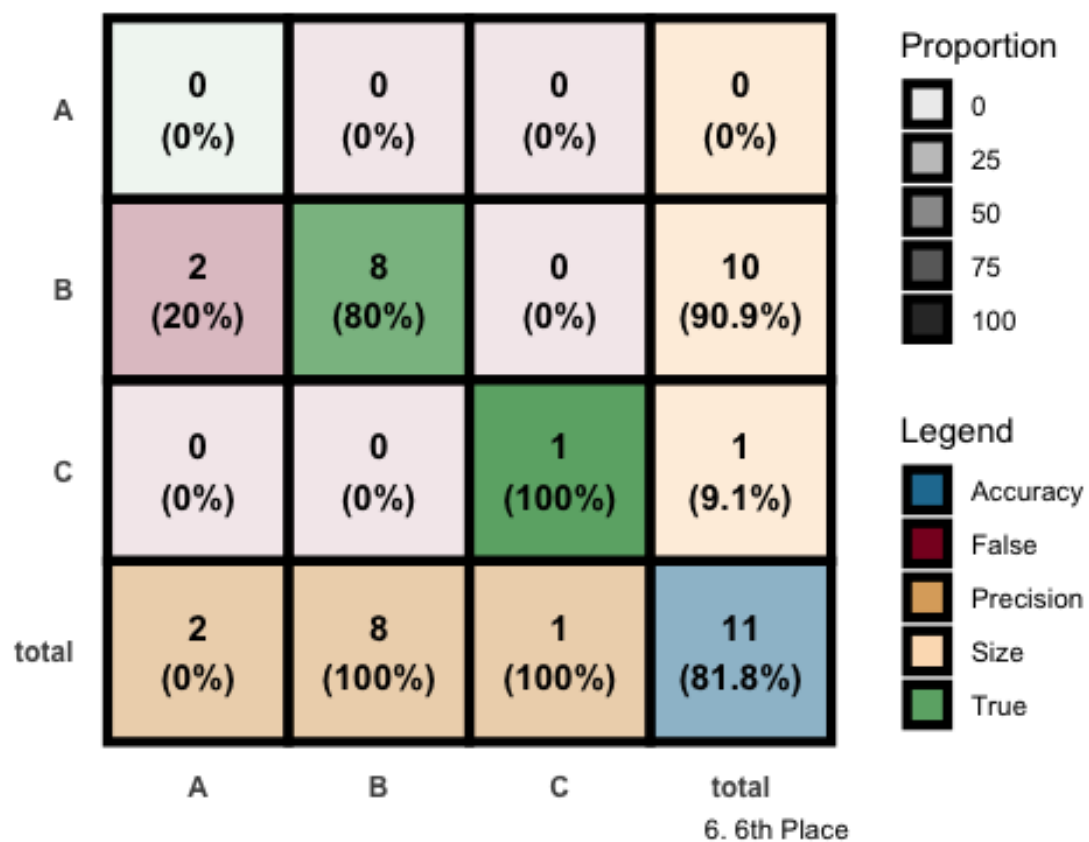
	A	B	C	Exp
A	40	60	0	A
AF	84	16	0	A
B	83	17	0	A
C	61	39	0	A
D	99	1	0	A
E	41	59	0	A
F	100	0	0	A
AA	18	82	0	B
AB	26	74	0	B
AD	26	74	0	B
AH	2	98	1	B
AJ	21	79	0	B
AK	0	14	86	B
AL	1	98	1	B
AM	0	84	16	B
H	20	80	0	B
I	15	85	0	B
J	1	98	1	B
L	0	95	5	B
M	0	90	10	B
O	0	97	2	B
P	3	96	0	B
Q	1	98	2	B
T	0	81	19	B
U	13	86	0	B
V	38	62	0	B
Y	2	97	1	B
Z	1	98	1	B
AN	0	8	92	C
AP	0	2	98	C
AQ	0	97	3	C
AR	0	6	94	C
AS	0	20	80	C
AT	0	10	90	C
AU	0	4	96	C



6. 6th Place

Test Set

Confusion Matrix



Test Set

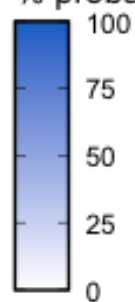
Probability Heatmap

	A	B	C
AC	12	88	0
AE	0	96	4
AI	1	98	2
G	38	62	0
K	0	83	17
N	100	0	0
R	23	77	0
S	80	20	0
W	0	88	12
X	24	76	0
AO	0	6	94

Exp

B
B
B
B
B
B
B
B
B
B
B
C

% probability



6. 6th Place

Model Analysis - Variable Impact

1. Dist_(P,C)_P Removed

formula

class ~ NPA_6_SM + para_SM

Full Model Stats - Overall Accuracy and Pseudo-R2

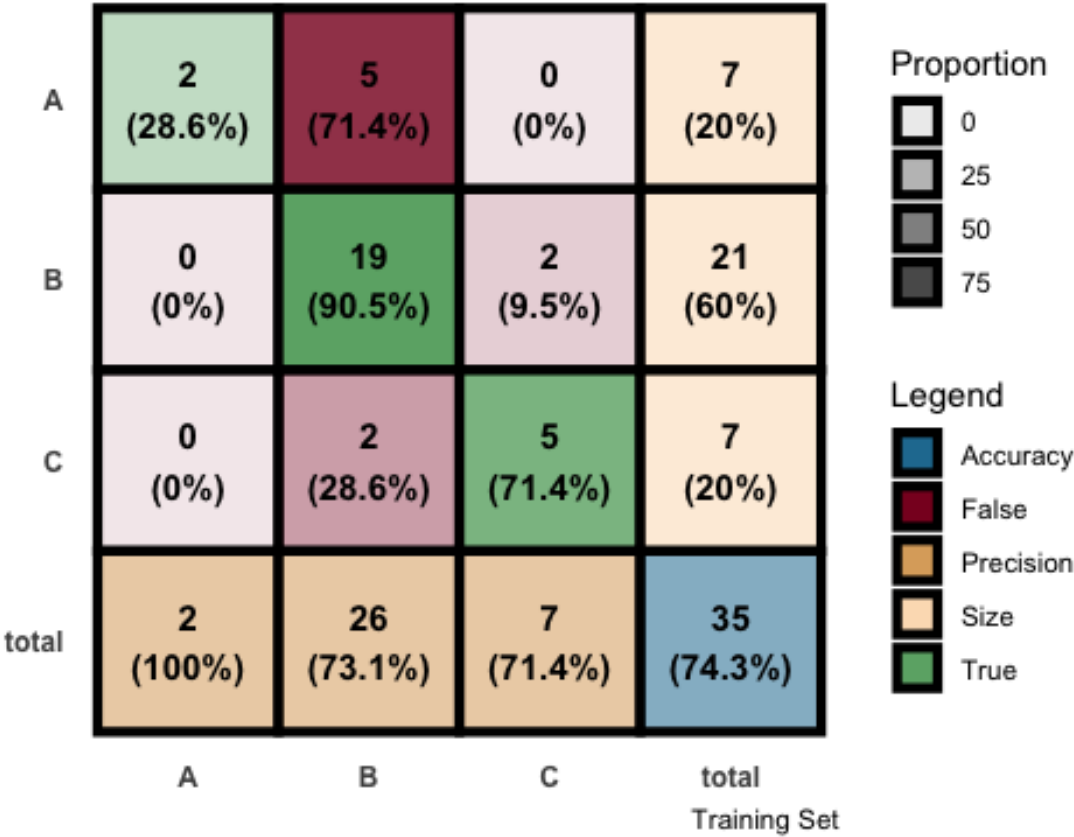
Accuracy	McFadden_R2
74.29%	0.416

Model Coefficients

	x
NPA_6_SM	24.5249253
para_SM	0.0630958

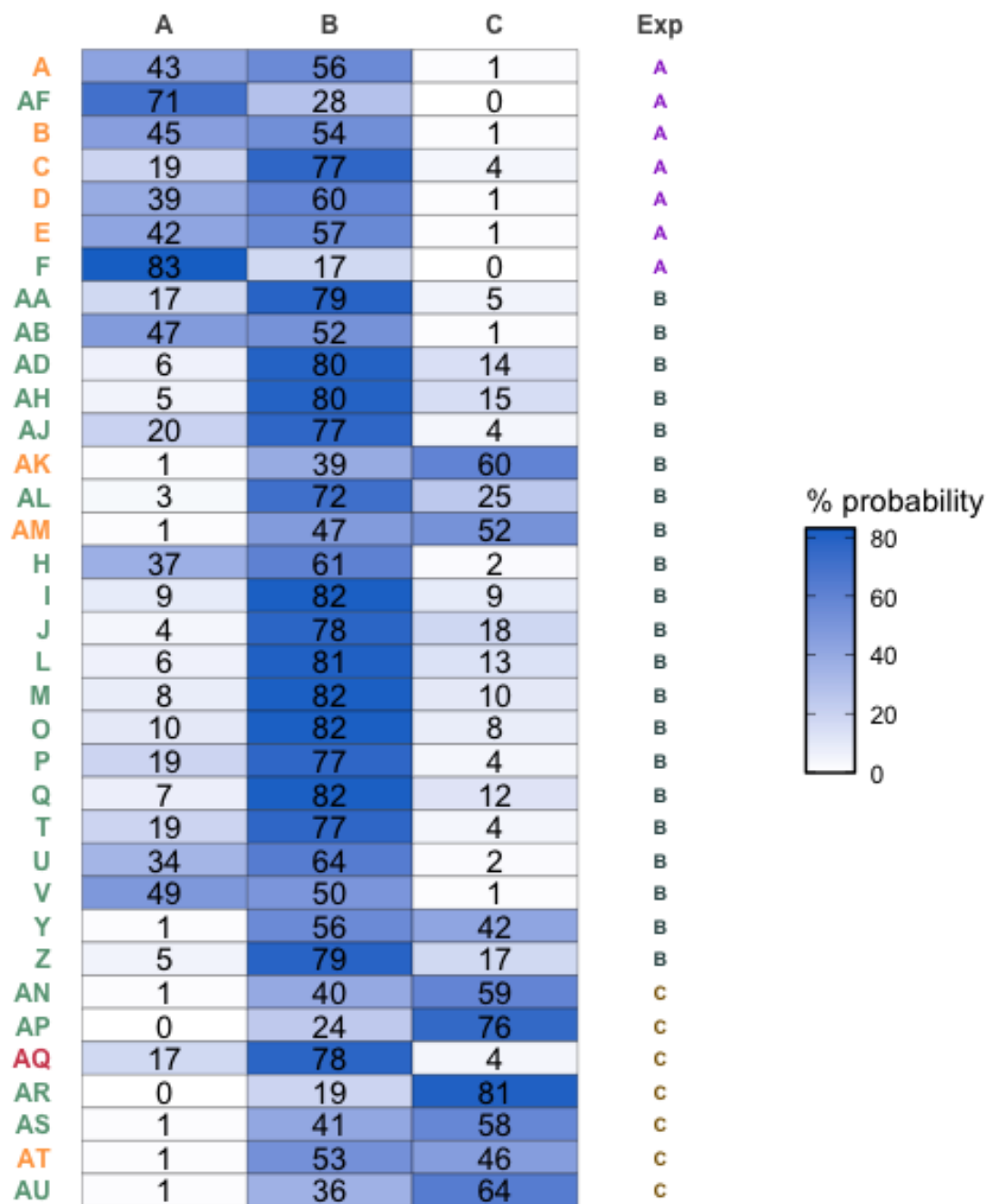
Dist_(P,C)_P Removed

Confusion Matrix



Dist_(P,C)_P Removed

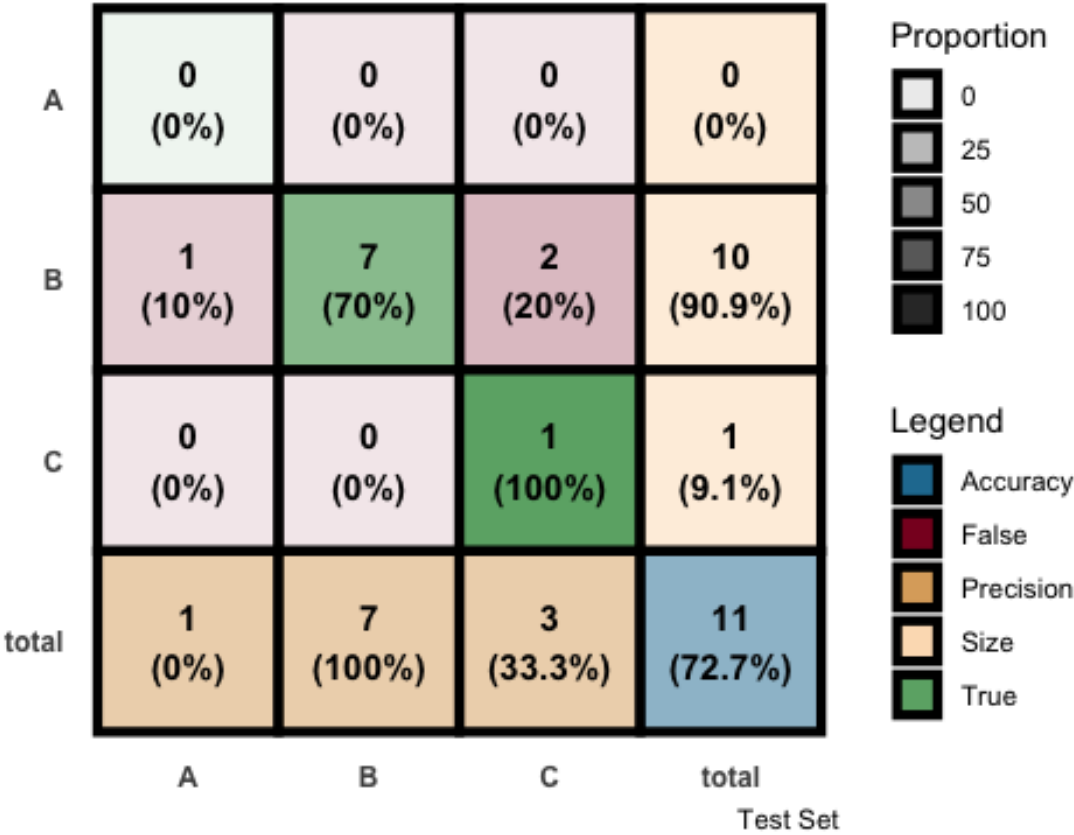
Probability Heatmap



Training Set

Dist_(P,C)_P Removed

Confusion Matrix

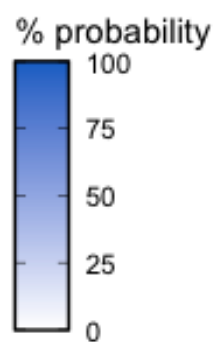


Dist_(P,C)_P Removed Probability Heatmap

	A	B	C
AC	42	57	1
AE	6	81	13
AI	46	53	1
G	39	60	1
K	1	35	65
N	100	0	0
R	41	58	1
S	5	80	15
W	0	24	76
X	49	50	1
AO	1	45	55

Exp

B
B
B
B
B
B
B
B
B
B
B
C



Test Set

2. NPA_6_SM Removed

formula

class ~ Dist_(P,C)_P + para_SM

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy	McFadden_R2
----------	-------------

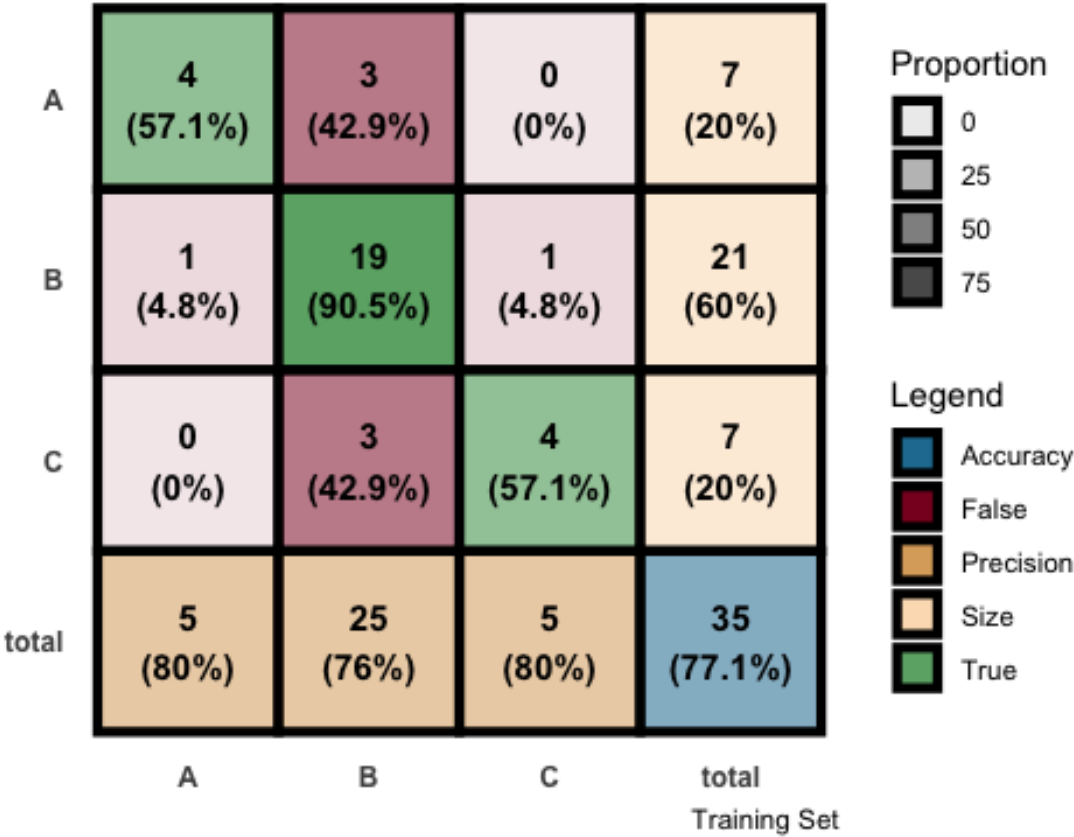
77.14%	0.407
--------	-------

Model Coefficients

	x
Dist_(P,C)_P	-134.4220030
para_SM	0.1140709

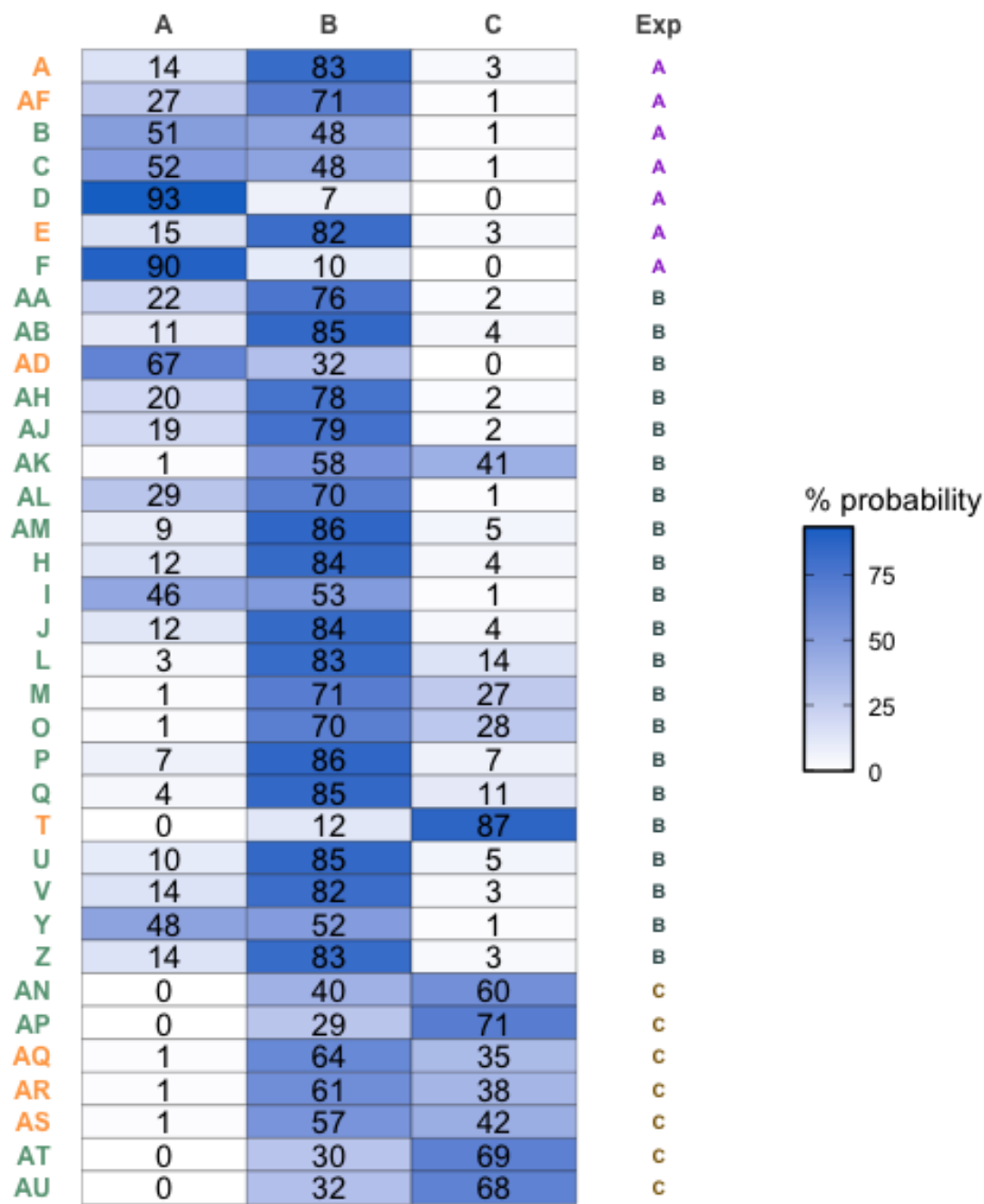
NPA_6_SM Removed

Confusion Matrix



NPA_6_SM Removed

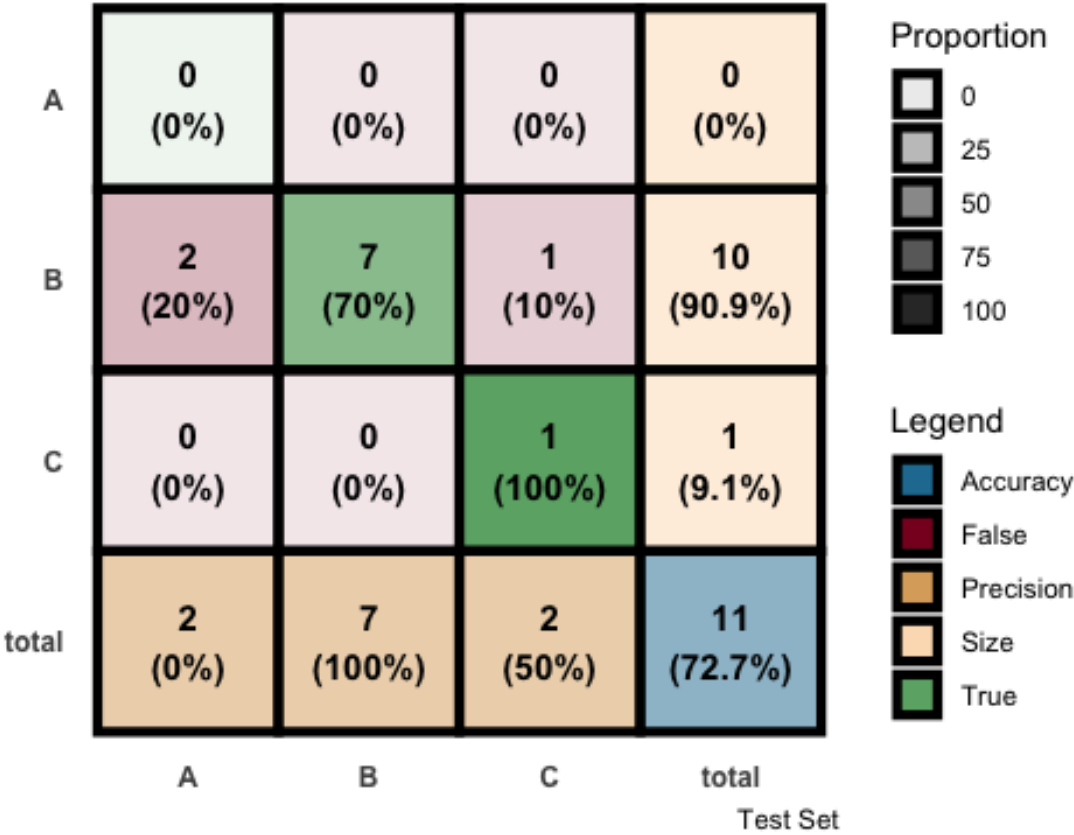
Probability Heatmap



Training Set

NPA_6_SM Removed

Confusion Matrix



NPA_6_SM Removed

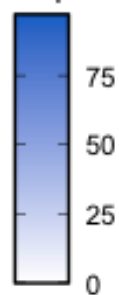
Probability Heatmap

	A	B	C
AC	8	86	6
AE	3	81	16
AI	0	45	55
G	16	81	3
K	12	84	4
N	63	36	0
R	11	85	4
S	97	3	0
W	27	71	1
X	8	86	6
AO	0	26	74

Exp

B
B
B
B
B
B
B
B
B
B
B
C

% probability



Test Set

3. para_SM Removed

formula

class ~ Dist_(P,C)_P + NPA_6_SM

Full Model Stats - Overall Accuracy and Pseudo-R2

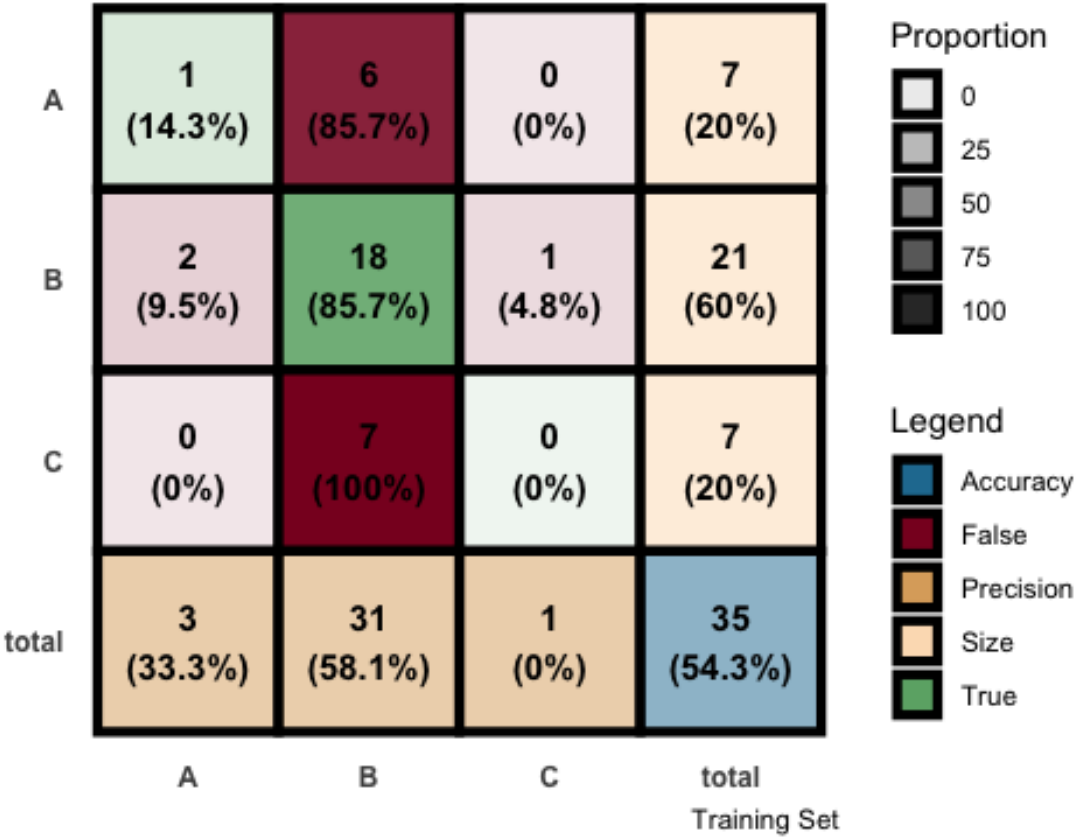
Accuracy	McFadden_R2
----------	-------------

54.29%	0.227
--------	-------

Model Coefficients

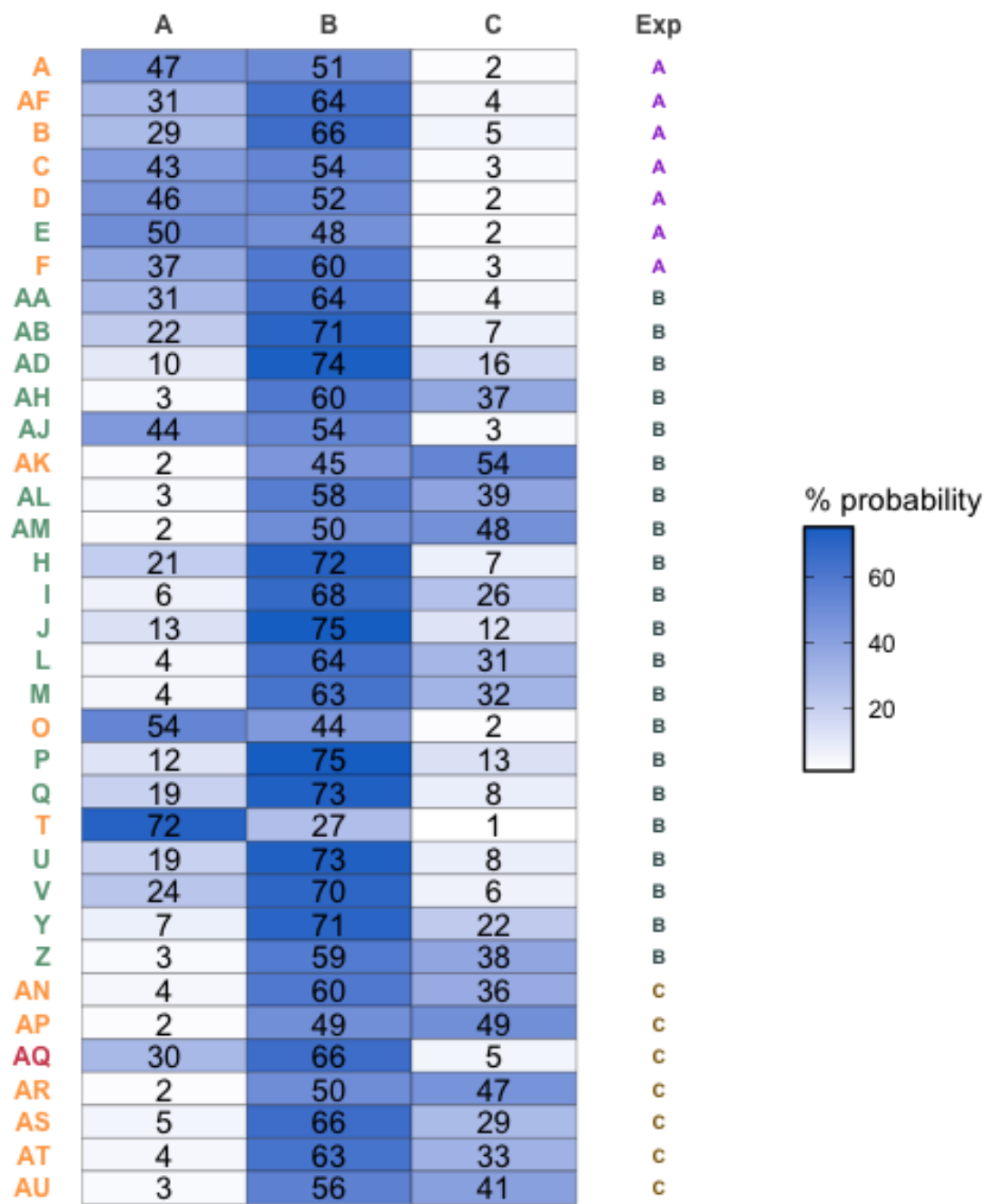
	x
Dist_(P,C)_P	-54.92384
NPA_6_SM	14.72591

para_SM Removed
Confusion Matrix



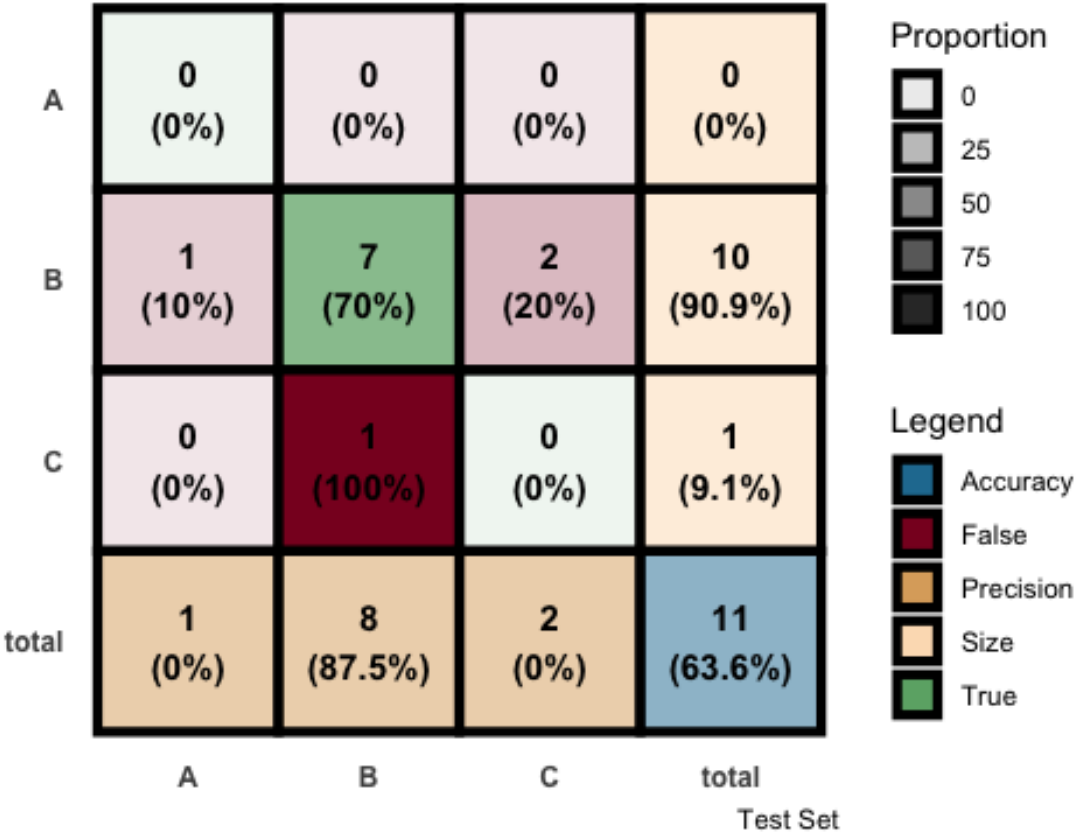
para_SM Removed

Probability Heatmap



Training Set

para_SM Removed
Confusion Matrix



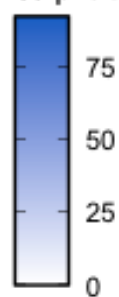
para_SM Removed Probability Heatmap

	A	B	C
AC	12	75	13
AE	13	75	12
AI	21	72	7
G	44	53	3
K	3	56	41
N	92	8	0
R	27	68	5
S	1	31	68
W	2	45	53
X	29	66	5
AO	4	65	31

Exp

B
B
B
B
B
B
B
B
B
B
B
C

% probability



Test Set