# Pyridine Phosphination – 3 categories – Best Model Analysis

### **Table of Contents**

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

### **Best Model**

formula	McFadden R2
class $\sim$ para_P + NPA_6_SM + Avg_NPA_SM	0.703
class $\sim$ NPA_6_SM + para_SM + fr_C_O	0.662
class $\sim$ NPA_6_SM + para_SM + fr_C_0_noCOO	0.662
${\tt class \sim Dist\_(P,C)\_P + para\_SM + Iso.Polar\_SM}$	0.651
class $\sim$ para_P + NPA_6_SM + NPA_sum_SM	0.649
class $\sim$ Dist_(P,C)_P + NPA_6_SM + para_SM	0.649
class $\sim$ NPA_6_SM + Avg_NPA_SM + para_SM	0.646
class $\sim$ NPA_1_P + Dist_(P,C)_P + para_SM	0.627
class $\sim$ NPA_1_P + para_SM + fr_pyridine	0.626
class $\sim$ NPA_6_SM + NPA_sum_SM + para_SM	0.623

### **6th Place**

Table:

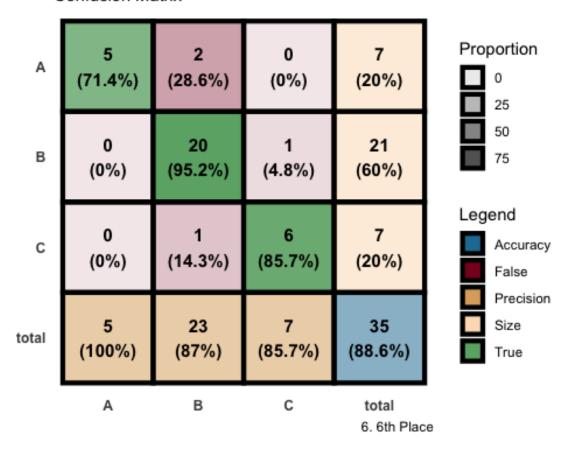
Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy McFadden\_R2 88.57% 0.649

**Model Coefficients** 

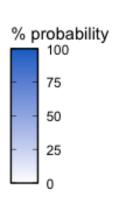
	X
Dist_(P,C)_P	-181.4237850
NPA_6_SM	33.1536447
para_SM	0.2071308

Training Set Confusion Matrix



Training Set Probability Heatmap

	Α	В	С
Α	40	60	0
AF	84	16	0
В	83	17	0
С	61	39	0
D	99	1	0
E	41	59	0
F	100	0 82	0
AA	18	82	0
AB	26	74	0
AD	26	74	0
AH	2	98	1
AJ	21	79	0
AK	0	14	86
AL	1	98	1
AM	0	84	16
Н	20	80	0
- 1	15	85	0
J	1	98	1
L	0	95	5
M	0	90	10
0	0	97	2
Р	3	96	0
Q	1	98	2
Т	0	81	19
U	13	86	0
V	38	62	0
Υ	2	97	1
Z	1	98	1
AN	0	8	92
AP	0	2	98
AQ	0	97	3
AR	0	6	94
AS	0	20	80
AT	0	10	90
AU	0	4	96



6. 6th Place

Ехр

B B

В

В

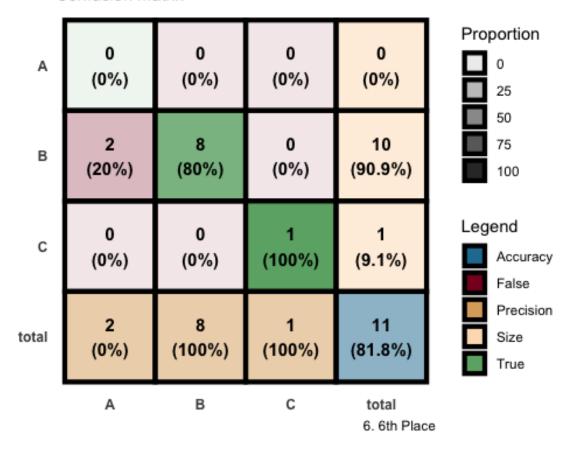
B B

В

B B

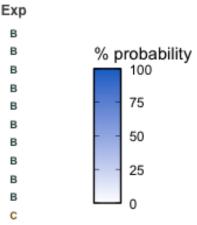
c c

Test Set Confusion Matrix



Test Set Probability Heatmap

	Α	В	С	E
AC	12	88	0	
AE	0	96	4	
ΑI	1	98	2	
G	38	62	0	
K	0	83	17	
N	100 23	0	0	
R S	23	77	0	
S	80	20 88 76	0	
W	0	88	12	
Х	0 24	76	0	
AO	0	6	94	



6. 6th Place

### **Model Analysis - Variable Impact**

1. Dist\_(P,C)\_P Removed

formula

 ${\tt class} \sim {\tt NPA\_6\_SM} + {\tt para\_SM}$ 

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy McFadden\_R2

74.29% 0.416

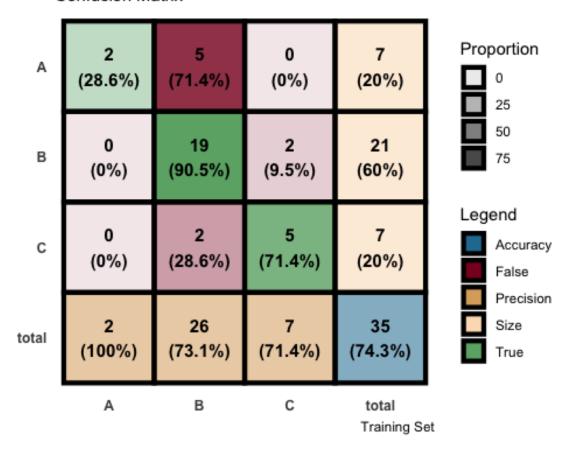
**Model Coefficients** 

X

NPA\_6\_SM 24.5249253

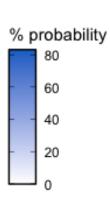
para\_SM 0.0630958

Dist\_(P,C)\_P Removed Confusion Matrix



# Dist\_(P,C)\_P Removed Probability Heatmap

	Α	В	С
Α	43	56	1
AF	71	28	0
В	45	54	1
С	19	77	4
D	39	60	1
E	42	57	1
F	83 17	17 79 52	0
AA	17	79	5
AB	47	52	1
AD	6	80	14
AH	5	80	15
AJ	20	77	4
AK	1	39	60
AL	3	72	25
AM	1	47	52
Н	37	61	2
- 1	9	82	9
J	4	78	18
L	6	81	13
M	8	82	10
0	10	82	8
P	19	77	4
Q	7	82 77	12
Т	19	77	4
U	34	64	2
V	49	50	1
Υ	1	56	42
Z	5	79	17
AN	1	40	17 59
AP	0	24	76 4
AQ	17	78	4
AR	0	19	81
AS	1	41	58
AT	1	53	46
AU	1	36	64



Training Set

Exp

В

B B

В

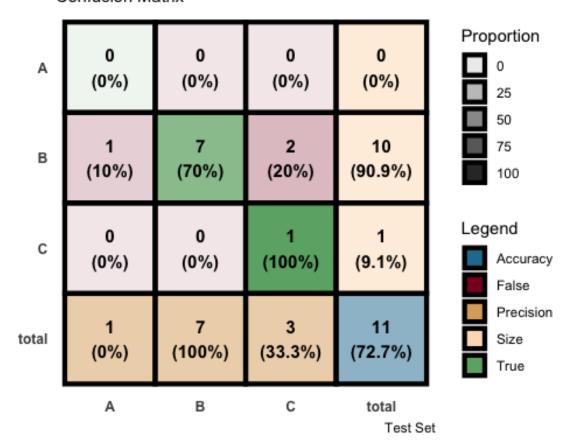
В

B B

B B B

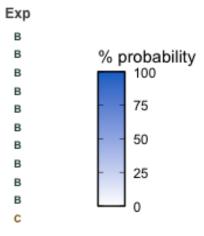
c c c

Dist\_(P,C)\_P Removed Confusion Matrix



Dist\_(P,C)\_P Removed Probability Heatmap

	Α	В	С
AC	42	57	1
AE	42 6	81	13
ΑI	46	53	1
G K N	39	60	1
K	1	35	65
N	100	0 58	0
R S	41		1
	5	80	15 76
W X	0 49	24	76
Х	49	24 50	1
AO	1	45	55



Test Set

### 2. NPA\_6\_SM Removed

formula

 ${\tt class} \sim {\tt Dist\_(P,C)\_P + para\_SM}$ 

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy McFadden\_R2

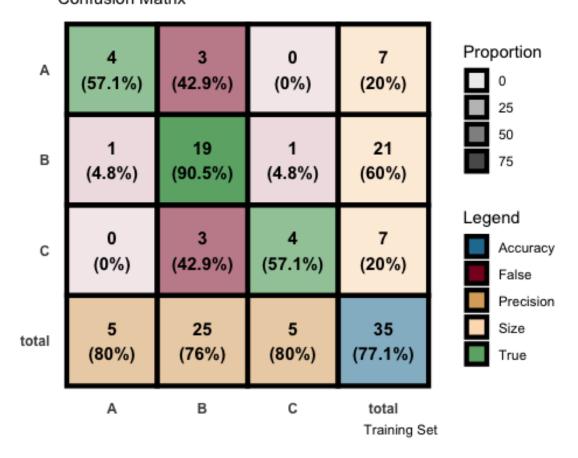
77.14% 0.407

**Model Coefficients** 

Dist\_(P,C)\_P -134.4220030

para\_SM 0.1140709

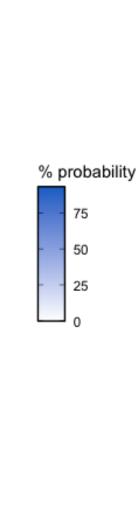
NPA\_6\_SM Removed Confusion Matrix



# NPA\_6\_SM Removed

## Probability Heatmap

	Α	В	С
Α	14	83	3
AF	27	71	1
В	51	48	1
С	52	48	1
D	93 15	7 82	0
E	15	82	3
F	90 22	10 76	0
AA	22	76	2
AB	11	85	4
AD	11 67	85 32	0
AH	20	78	2 2
AJ	19	79	2
AK	1	58	41
AL	29	70	1
AM	9	86	5
Н	12	84	4
- 1	46	53	1
J	12	84	4
L	3	83	14
M	1	71	27 28
0	1	70	28
P	7	86	7
Q	4	85 12 85 82	11 87
T	0	12	87
U	10	85	5 3
V	14	82	3
Υ	48	52	1
Z	14	83	3
AN	0	40	60
AP	0	29	71
AQ	1	64	35
AR	1	61	38
AS	1	57	42
AT	0	30	69
AU	0	32	68



Training Set

Exp

В

B B B

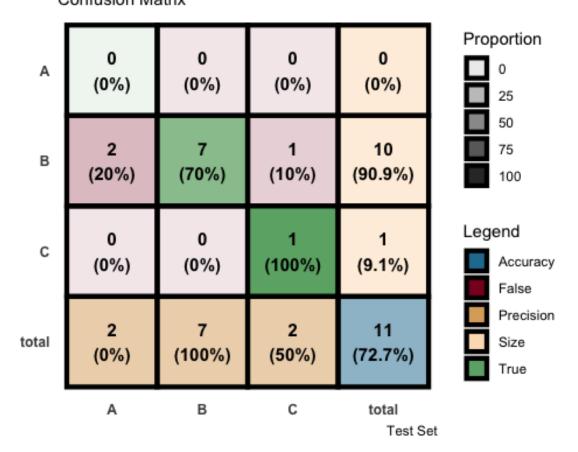
В

B B

B B B

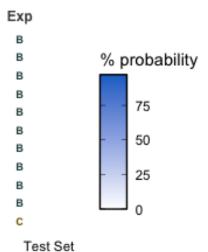
c c c

NPA\_6\_SM Removed Confusion Matrix



NPA\_6\_SM Removed Probability Heatmap

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



### 3. para\_SM Removed

### formula

 ${\tt class \sim Dist\_(P,C)\_P + NPA\_6\_SM}$ 

Full Model Stats - Overall Accuracy and Pseudo-R2

Accuracy McFadden\_R2

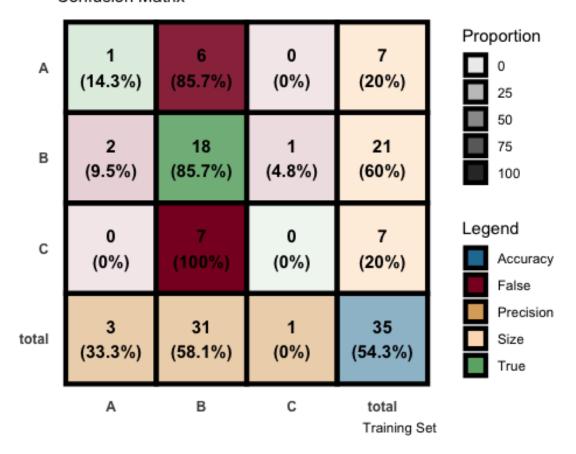
54.29% 0.227

**Model Coefficients** 

Dist\_(P,C)\_P -54.92384

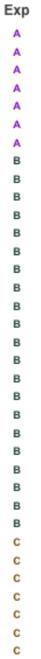
NPA\_6\_SM 14.72591

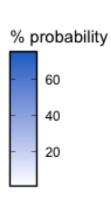
para\_SM Removed Confusion Matrix



## para\_SM Removed Probability Heatmap

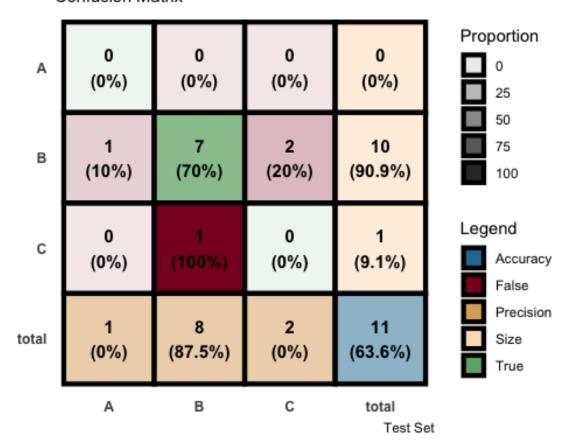
	,		
	Α	В	С
Α	47	51	2
AF	31	64	
В	31 29	66	4 5 3 2 2 2 3 4 7
B C D	43	54	3
D	46	52	2
Е	50	48	2
F	50 37 31 22 10 3	60	3
AA	31	64	4
AB	22	71	7
AD	10	74	16
AH	3	60	37
AJ	44	54	3
AK	2	45	16 37 3 54
AL	3	58	39
AM	2	50	48
Н	21	72	7
- 1	2 3 2 21 6 13 4	68	7 26 12 31 32 2 13
J	13	75	12
L	4	64	31
M	4	63	32
0	54	44	2
Р	12 19 <b>72</b> 19 24 7	75 73 27 73 70	13
Q T U	19	73	8 1 8
T	72	27	1
U	19	73	8
V Y Z	24	70	6
Υ	7	71	22
	3	59	38
AN	4	60	36
AP	2	49	49
AQ	4 2 30	66	5 47 29 33
AR	2 5 4 3	50	47
AS	5	66	29
AT	4	63	33
AU	3	56	41





Training Set

para\_SM Removed
Confusion Matrix



# para\_SM Removed Probability Heatmap

	Α	В	С
AC	12 13	75 75	13 12
AE	13	75	12
ΑI	21	72	7
AI G K N R S W	44	53	3
K	44 3 92 27	56 8 68	41
N	92	8	0
R	27	68	0 5
S	1	31	68
W	2	45	<b>53</b> 5
X	2 29	45 66	5
AO	4	65	31

