UAS IS655 BDA-II Theory WillibrordusBayu-00000034000

Creation Date: Friday, June 3, 2022 05:09:45 PM

Author: willi.brordus@student.umn.ac.id

UAS IS655 BDA-II Theory WillibrordusBayu-00000034000

Contents

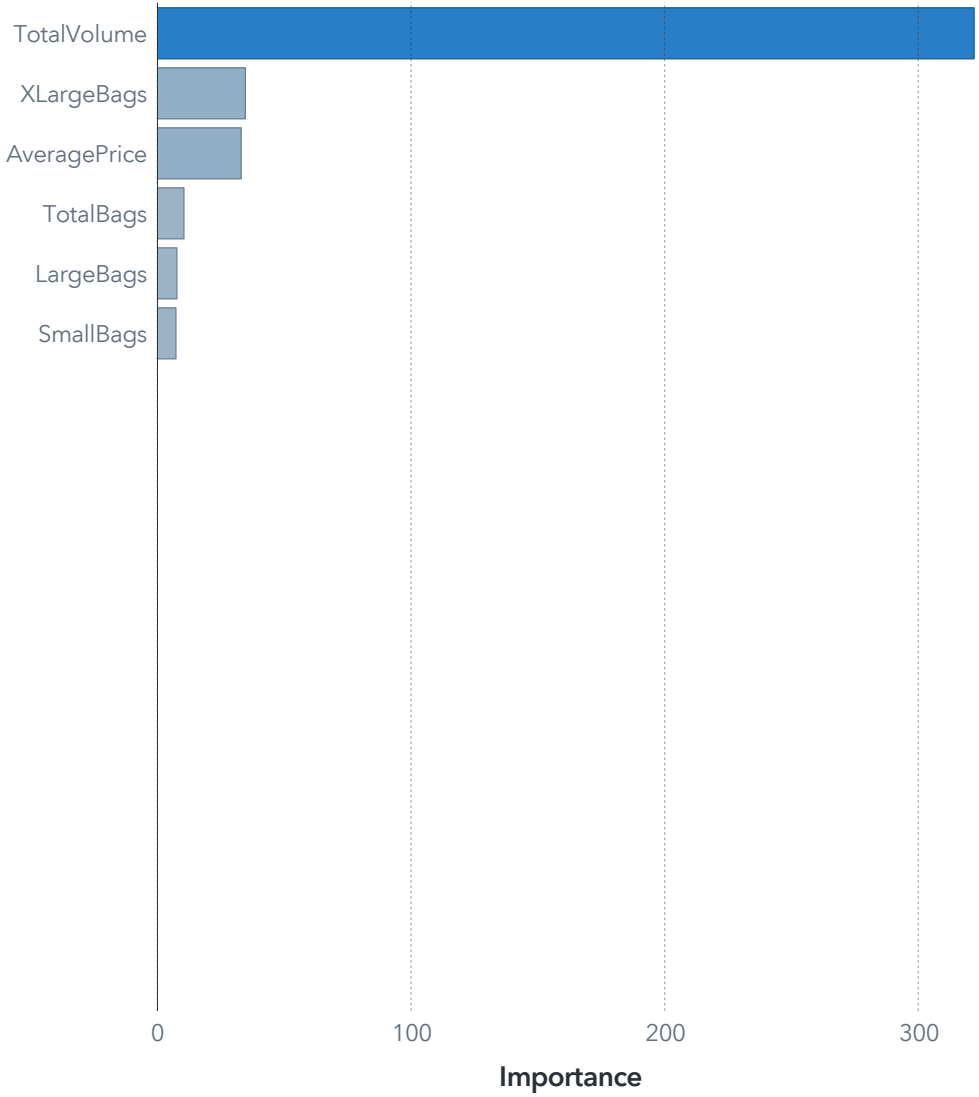
UAS IS655 BDA-II Theory WillibrordusBayu-00000034000

Contents

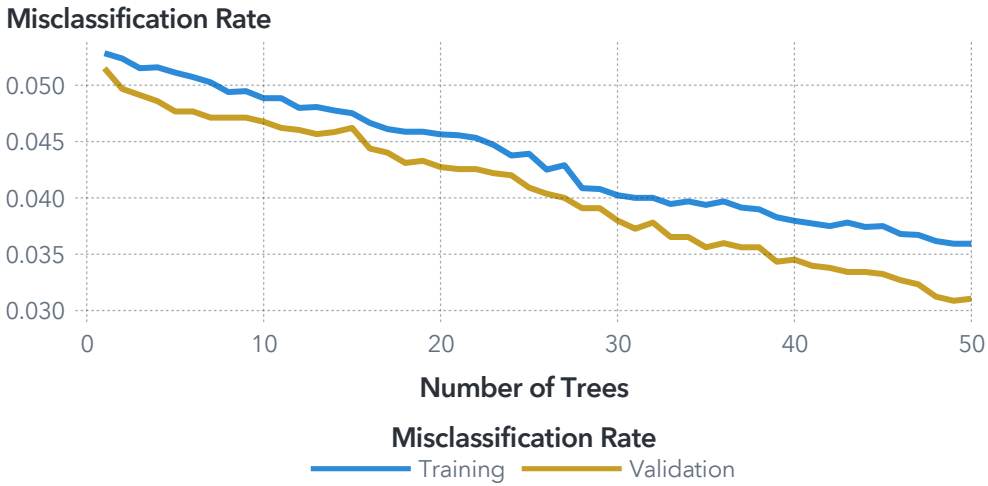
Gradient Boosting

Gradient Boosting **type** (event=organic) Validation F1 Score **0.969** Observations Used **18,249**

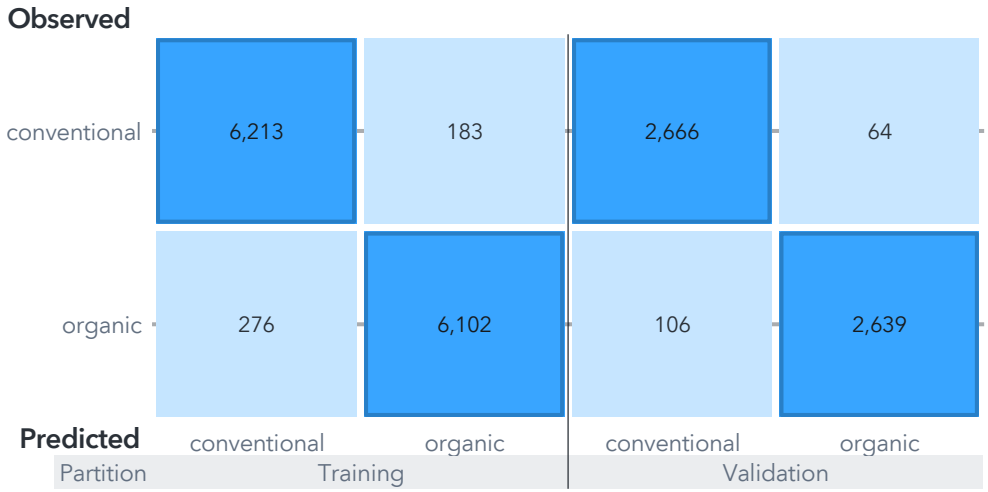
Variable Importance



Iteration Plot

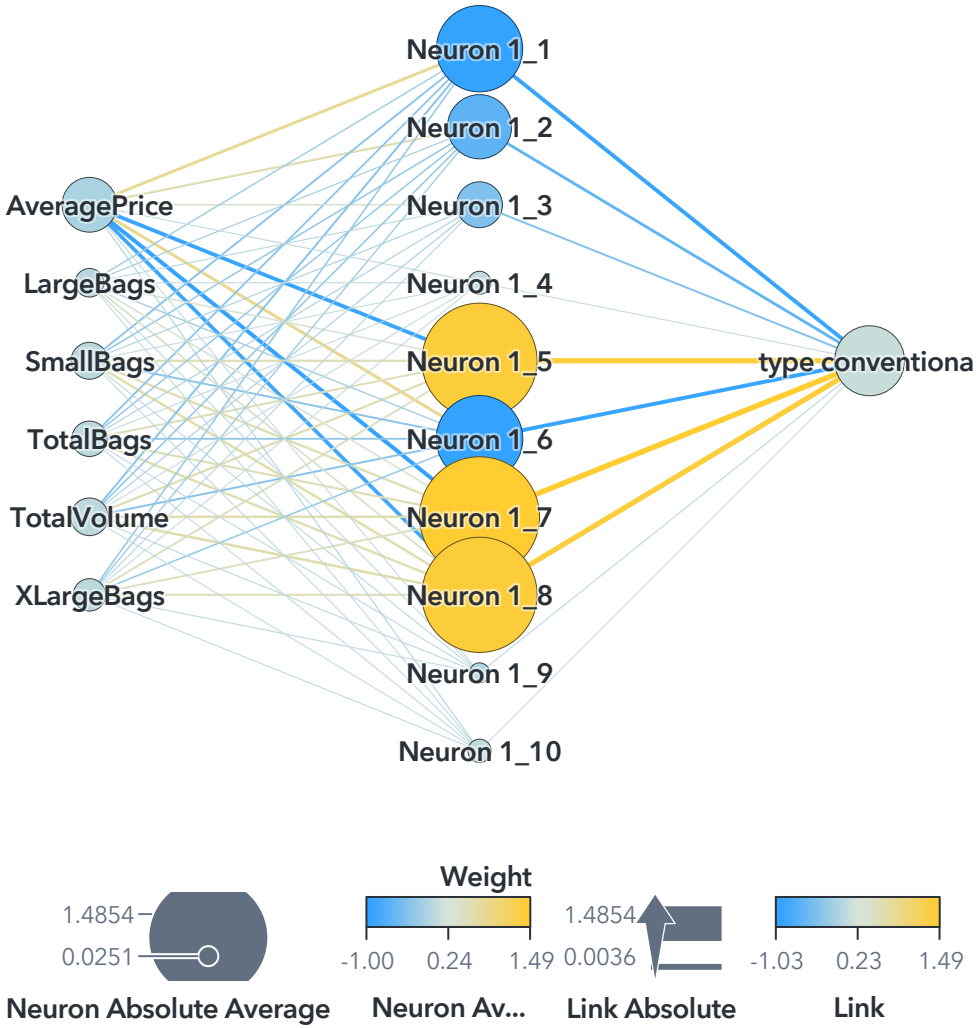


Confusion Matrix

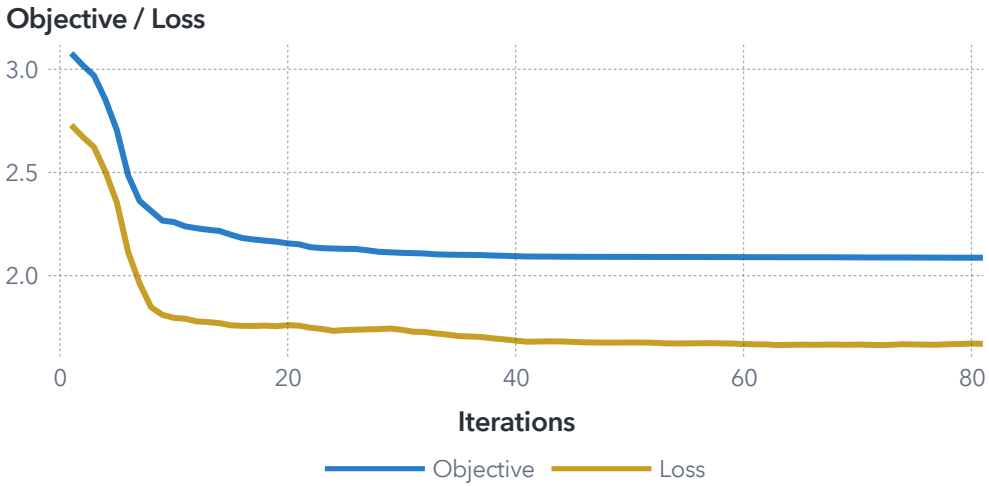


Neural Network **type** (event=organic) Validation F1 Score **0.820** Observations Used **18,249**

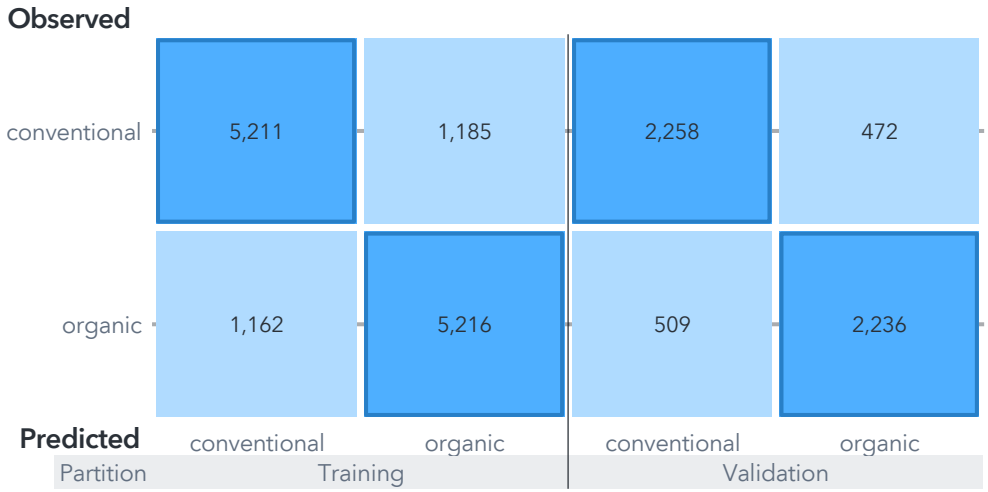
Network



Iteration Plot



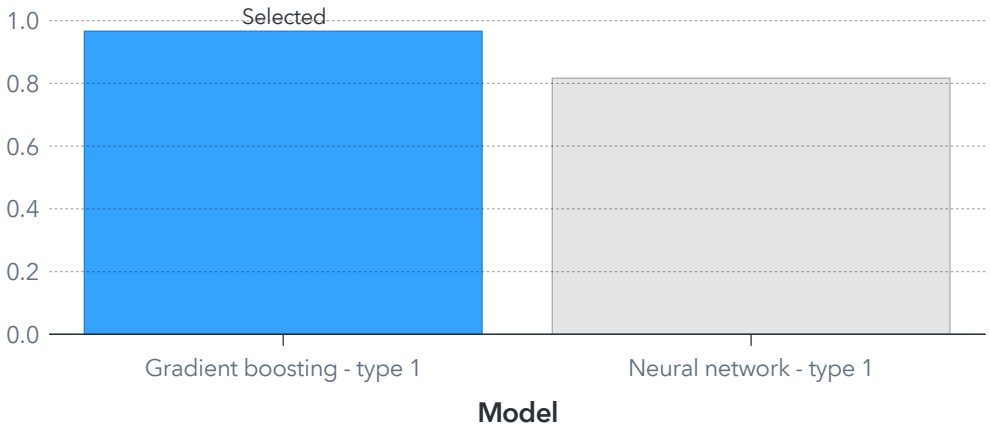
Confusion Matrix



Model Comparison **type** (event=organic)

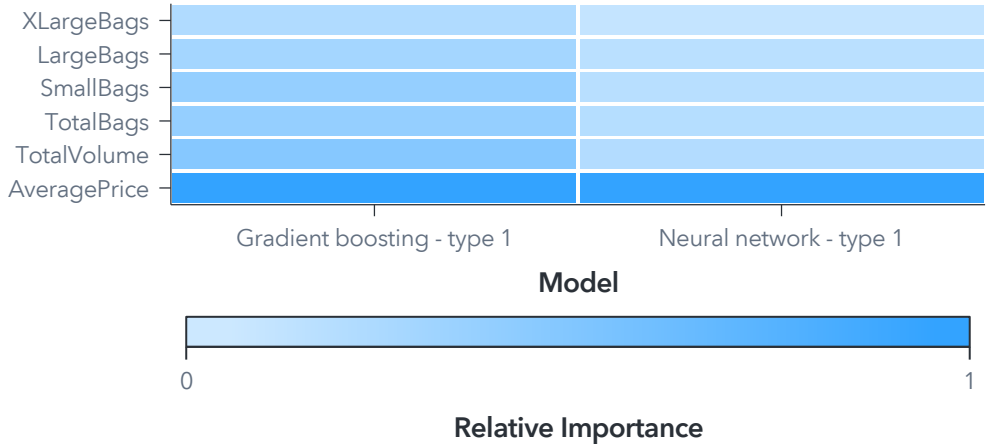
Fit Statistic

F1 Score



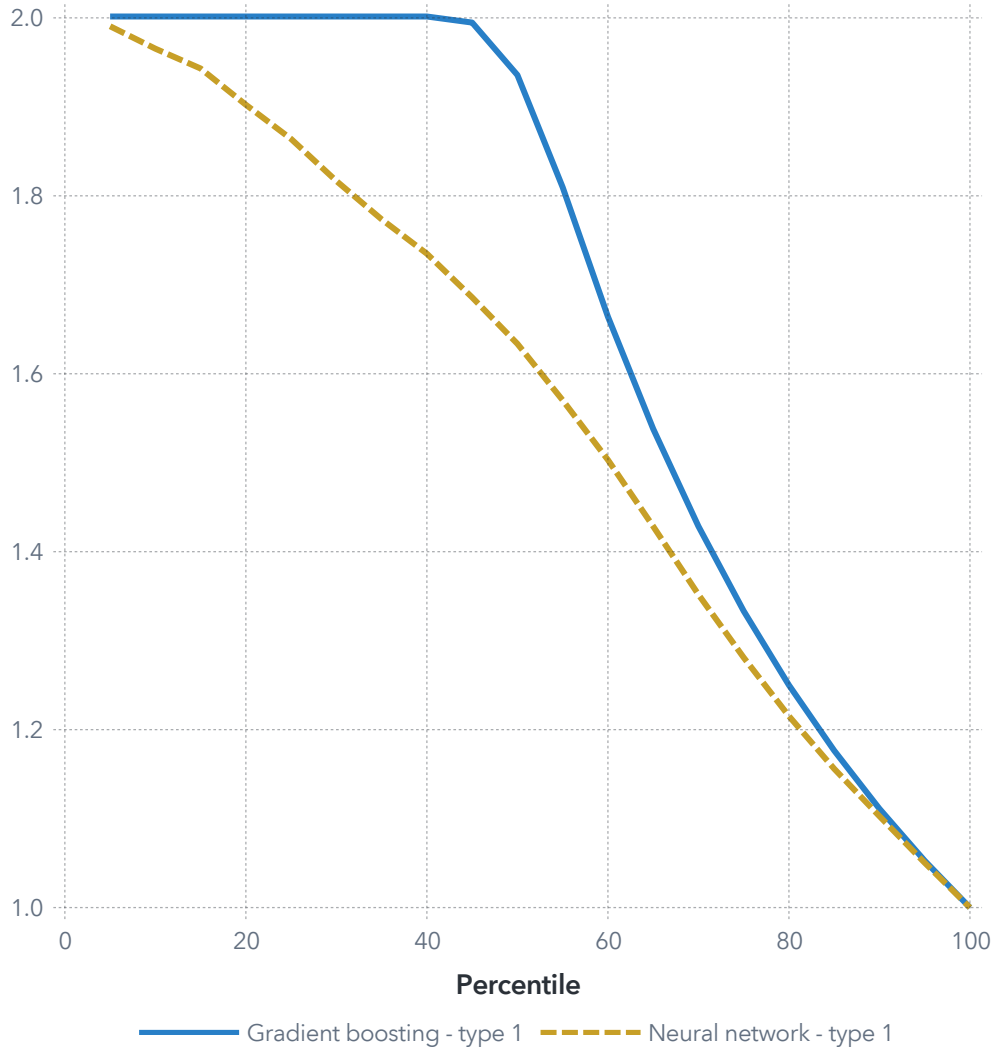
Relative Importance

Variable



Lift

Cumulative Lift



Gradient boosting - type 1 Supplement 1

Variable	Importance	Standard Deviation
TotalVolume	322.0358	309.3621
XLargeBags	34.6993	28.4736
AveragePrice	33.0451	26.5493
TotalBags	10.4722	7.7386
LargeBags	7.6942	5.2090
SmallBags	7.3199	6.7932

Number of Trees	Misclassification Rate	Validation Misclassification Rate
1	0.0528	0.0515
2	0.0524	0.0497
3	0.0515	0.0491
4	0.0516	0.0486
5	0.0511	0.0477
6	0.0507	0.0477
7	0.0503	0.0471
8	0.0494	0.0471
9	0.0495	0.0471
10	0.0488	0.0468
11	0.0488	0.0462
12	0.0480	0.0460
13	0.0481	0.0457
14	0.0478	0.0458
15	0.0475	0.0462
16	0.0467	0.0444
17	0.0461	0.0440
18	0.0459	0.0431
19	0.0459	0.0433
20	0.0456	0.0427
21	0.0456	0.0426
22	0.0453	0.0426
23	0.0447	0.0422
24	0.0438	0.0420
25	0.0439	0.0409
26	0.0425	0.0404
27	0.0429	0.0400
28	0.0409	0.0391
29	0.0408	0.0391
30	0.0402	0.0380
31	0.0400	0.0373
32	0.0400	0.0378
33	0.0395	0.0365
34	0.0397	0.0365
35	0.0394	0.0356
36	0.0397	0.0360
37	0.0391	0.0356
38	0.0390	0.0356
39	0.0383	0.0343
40	0.0380	0.0345
41	0.0377	0.0340
42	0.0375	0.0338
43	0.0378	0.0334
44	0.0374	0.0334
45	0.0375	0.0332
46	0.0368	0.0327
47	0.0367	0.0323
48	0.0362	0.0312
49	0.0359	0.0309
50	0.0359	0.0311

Gradient boosting - type 1 Supplement 3

Predicted	Observed	Training Frequency	Training Percentage	Validation Frequency	Validation Percentage
conventional	conventional	6,213	97.14%	2,666	97.66%
organic	conventional	183	2.86%	64	2.34%
conventional	organic	276	4.33%	106	3.86%
organic	organic	6,102	95.67%	2,639	96.14%

Gradient boosting - type 1 Supplement 4

Percentile	Training Observations	Training Events	Training Lift	Training Lift Best	Training Cumulative Lift	Training Cumulative Lift Best	Validation Observations	Validation Events	Validation Lift	Validation Lift Best	Validation Cumulative Lift	Validation Cumulative Lift Best
5.00	639	639	2.0038	2.0038	2.0038	2.0038	274	274	1.9964	1.9964	1.9964	1.9964
10.00	639	639	2.0038	2.0038	2.0038	2.0038	274	274	1.9964	1.9964	1.9964	1.9964
15.00	639	639	2.0038	2.0038	2.0038	2.0038	274	274	1.9964	1.9964	1.9964	1.9964
20.00	639	639	2.0038	2.0038	2.0038	2.0038	274	274	1.9964	1.9964	1.9964	1.9964
25.00	639	639	2.0038	2.0038	2.0038	2.0038	274	274	1.9964	1.9964	1.9964	1.9964
30.00	639	639	2.0038	2.0038	2.0038	2.0038	274	274	1.9964	1.9964	1.9964	1.9964
35.00	639	639	2.0038	2.0038	2.0038	2.0038	274	274	1.9964	1.9964	1.9964	1.9964
40.00	639	638	2.0006	2.0038	2.0034	2.0038	274	274	1.9964	1.9964	1.9964	1.9964
45.00	639	618	1.9379	2.0038	1.9961	2.0038	274	265	1.9308	1.9964	1.9891	1.9964
50.00	639	438	1.3735	1.9661	1.9338	2.0000	274	208	1.5155	1.9964	1.9417	1.9964
55.00	639	181	0.5676	0.0000	1.8096	1.8182	274	70	0.5100	0.0364	1.8116	1.8182
60.00	639	22	0.0690	0.0000	1.6646	1.6667	274	7	0.0510	0.0000	1.6648	1.6667
65.00	639	7	0.0220	0.0000	1.5382	1.5385	274	3	0.0219	0.0000	1.5385	1.5385
70.00	639	1	0.0031	0.0000	1.4286	1.4286	274	0	0.0000	0.0000	1.4286	1.4286
75.00	639	0	0.0000	0.0000	1.3333	1.3333	274	0	0.0000	0.0000	1.3333	1.3333
80.00	639	0	0.0000	0.0000	1.2500	1.2500	274	0	0.0000	0.0000	1.2500	1.2500
85.00	639	0	0.0000	0.0000	1.1765	1.1765	274	0	0.0000	0.0000	1.1765	1.1765
90.00	639	0	0.0000	0.0000	1.1111	1.1111	274	0	0.0000	0.0000	1.1111	1.1111
95.00	639	0	0.0000	0.0000	1.0526	1.0526	274	0	0.0000	0.0000	1.0526	1.0526
100.00	633	0	0.0000	0.0000	1.0000	1.0000	269	0	0.0000	0.0000	1.0000	1.0000

Cutoff	Training Sensitivity	Training 1 - Specificity	Training KS (Youden)	Validation Sensitivity	Validation 1 - Specificity	Validation KS (Youden)
0.00	1.0000	1.000		1.0000	1.000	
0.01	1.0000	0.579		1.0000	0.584	
0.02	1.0000	0.389		1.0000	0.400	
0.03	0.9998	0.303		1.0000	0.306	
0.04	0.9997	0.263		0.9993	0.266	
0.05	0.9994	0.234		0.9989	0.229	
0.06	0.9989	0.211		0.9989	0.203	
0.07	0.9984	0.192		0.9989	0.178	
0.08	0.9984	0.174		0.9982	0.166	
0.09	0.9984	0.164		0.9982	0.157	
0.10	0.9980	0.156		0.9978	0.149	
0.11	0.9973	0.148		0.9974	0.141	
0.12	0.9973	0.144		0.9971	0.134	
0.13	0.9967	0.138		0.9964	0.127	
0.14	0.9958	0.132		0.9964	0.119	
0.15	0.9956	0.123		0.9964	0.109	
0.16	0.9955	0.113		0.9964	0.100	
0.17	0.9950	0.105		0.9956	0.092	
0.18	0.9948	0.101		0.9949	0.089	
0.19	0.9944	0.096		0.9945	0.085	
0.20	0.9937	0.091		0.9942	0.079	
0.21	0.9936	0.087		0.9920	0.076	
0.22	0.9929	0.082		0.9920	0.074	
0.23	0.9918	0.078		0.9920	0.071	
0.24	0.9915	0.076		0.9916	0.068	
0.25	0.9915	0.073		0.9909	0.067	
0.26	0.9909	0.072		0.9898	0.064	
0.27	0.9903	0.069		0.9894	0.061	
0.28	0.9881	0.066		0.9872	0.058	
0.29	0.9878	0.063		0.9869	0.056	
0.30	0.9864	0.061		0.9869	0.052	
0.31	0.9857	0.059		0.9858	0.050	
0.32	0.9849	0.056		0.9851	0.049	
0.33	0.9824	0.054		0.9818	0.046	
0.34	0.9810	0.051		0.9811	0.044	
0.35	0.9795	0.050		0.9796	0.043	
0.36	0.9779	0.047		0.9781	0.041	
0.37	0.9771	0.045		0.9778	0.040	
0.38	0.9759	0.044		0.9770	0.038	
0.39	0.9744	0.042		0.9760	0.036	
0.40	0.9735	0.040	Yes	0.9760	0.033	
0.41	0.9716	0.038		0.9745	0.032	
0.42	0.9705	0.037		0.9745	0.031	Yes
0.43	0.9697	0.036		0.9730	0.030	
0.44	0.9683	0.035		0.9716	0.029	
0.45	0.9664	0.034		0.9709	0.028	
0.46	0.9647	0.033		0.9690	0.027	
0.47	0.9633	0.033		0.9672	0.027	
0.48	0.9613	0.030		0.9665	0.026	
0.49	0.9599	0.030		0.9628	0.025	
0.50	0.9567	0.029		0.9614	0.023	
0.51	0.9555	0.028		0.9581	0.023	
0.52	0.9542	0.026		0.9566	0.022	
0.53	0.9525	0.025		0.9545	0.021	
0.54	0.9508	0.023		0.9526	0.020	
0.55	0.9495	0.021		0.9505	0.019	
0.56	0.9479	0.020		0.9490	0.018	
0.57	0.9476	0.019		0.9464	0.017	
0.58	0.9454	0.018		0.9435	0.017	
0.59	0.9436	0.017		0.9413	0.015	
0.60	0.9425	0.016		0.9395	0.014	
0.61	0.9407	0.016		0.9384	0.014	
0.62	0.9392	0.015		0.9344	0.013	
0.63	0.9381	0.014		0.9330	0.012	
0.64	0.9343	0.013		0.9311	0.012	
0.65	0.9338	0.013		0.9304	0.012	
0.66	0.9318	0.012		0.9271	0.011	
0.67	0.9299	0.012		0.9260	0.011	
0.68	0.9271	0.010		0.9202	0.010	
0.69	0.9249	0.009		0.9191	0.010	
0.70	0.9225	0.008		0.9162	0.010	
0.71	0.9210	0.007		0.9148	0.008	
0.72	0.9169	0.006		0.9104	0.007	
0.73	0.9153	0.006		0.9067	0.006	
0.74	0.9127	0.005		0.9046	0.006	
0.75	0.9077	0.005		0.9024	0.006	
0.76	0.9045	0.005		0.9005	0.005	
0.77	0.9011	0.004		0.8984	0.004	
0.78	0.8989	0.003		0.8958	0.003	
0.79	0.8956	0.003		0.8922	0.003	
0.80	0.8937	0.003		0.8874	0.003	
0.81	0.8882	0.003		0.8838	0.003	
0.82	0.8863	0.003		0.8816	0.003	
0.83	0.8832	0.002		0.8783	0.003	
0.84	0.8788	0.002		0.8732	0.003	
0.85	0.8752	0.002		0.8718	0.003	
0.86	0.8700	0.002		0.8667	0.002	
0.87	0.8634	0.001		0.8619	0.001	
0.88	0.8584	0.001		0.8521	0.001	
0.89	0.8534	0.001		0.8459	0.001	
0.90	0.8463	0.001		0.8383	0.001	
0.91	0.8391	0.001		0.8299	0.000	
0.92	0.8300	0.000		0.8226	0.000	
0.93	0.8236	0.000		0.8164	0.000	
0.94	0.8114	0.000		0.8058	0.000	
0.95	0.8032	0.000		0.7964	0.000	
0.96	0.7923	0.000		0.7869	0.000	
0.97	0.7836	0.000		0.7763	0.000	
0.98	0.7714	0.000		0.7628	0.000	
0.99	0.7526	0.000		0.7435	0.000	

Gradient boosting - type 1 Supplement 6

Response	Event	Value	Training Frequency	Validation Frequency
Correct	organic	True Positive	6,102	2,639
Incorrect	organic	False Negative	276	106
Correct	conventional	True Negative	6,213	2,666
Incorrect	conventional	False Positive	183	64

Gradient boosting - type 1 Supplement 7

Partition	KS (Youden)	Misclassification Rate	Misclassification Rat...	C Statistic	FPR	FDR	F1 Score	Lift	Cumulative Lift	Cumulative % Events	Cumulative % Captu...	Gain	Gini	Gamma	Tau	Observations Used	Unused
Training	0.9336	0.0359	0.0359	0.996	0.029	0.029	0.964	2.004	2.004	100.000	10.019	1.004	0.993	0.993	0.497	12,774	0
Validation	0.9437	0.0311	0.0311	0.997	0.023	0.024	0.969	1.996	1.996	100.000	9.982	0.996	0.994	0.994	0.497	5,475	0

Plot	Summary
Confusion Matrix	The confusion matrix plot displays the number of observations predicting each response level. A greater number of observations where the observed level and predicted level are the same indicates a better model. For this data, the percentages of each observed value that are correctly predicted in the validation partition are as follows: conventional - 97.66%, and organic - 96.14%.
Lift	The lift plot measures the ratio of percent captured response to the baseline percent response. The validation partition has a lift of 2.0 at the 5% quantile meaning there are about 1.996 times more events in that quantile than expected by random (5% of the total number of events).
Cumulative Lift	Cumulative lift measures the ratio of percent captured response to the baseline percent response, up to and including the specified quantile. The validation partition has a cumulative lift of 2.0 in the 10% quantile meaning there are about 1.996 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the validation partition.
ROC	The receiver operator characteristic (ROC) is a plot of sensitivity (the true positive rate) against 1-specificity (the false positive rate), which are both measures of classification based on the confusion matrix. These measures are calculated at various cutoff values. To help identify the best cutoff to use when scoring your data, the KS cutoff reference line is drawn at the value of 1-specificity where the greatest difference between sensitivity and 1-specificity is observed for the validation partition. The KS cutoff line is drawn at the cutoff value 0.42 where the 1-specificity value is 0.031 and the sensitivity value is 0.974. Cutoff values range from 0 to 0.99, inclusive, in increments of 0.01. At each cutoff value, the predicted response classification is determined by whether the predicted probability of the response type being organic is greater than or equal to the cutoff value. When the predicted probability of the event is greater than or equal to the cutoff value, then the predicted classification is organic, otherwise it is NOT organic.
Misclassification	The misclassification plot is a visual representation of the accuracy of the prediction at the specified cutoff value, 0.50. The plot displays the number of true positives for events that are correctly classified, false positives for NOT events that are classified as events, false negatives for events that are classified as NOT events, and true negatives for NOT events that are classified as NOT events. True negatives include NOT event classifications that predict a different level from observed, as long as both are NOT events. The predicted response classification is determined by whether the predicted probability of the level organic for the response type is greater than or equal to the cutoff value. When it is greater than or equal to the cutoff value, the predicted classification is an event, otherwise it is a NOT event. For this data, for the bar corresponding to the event level of type, organic, the segment of the bar colored as "Correct" corresponds to true positives.

Description	Value
Model	Neural Net
Number of Observations Used for Training	12,774
Number of Observations Read for Training	12,774
Target/Response Variable	type
Number of Neurons	18
Number of Input Neurons	6
Number of Output Neurons	2
Number of Hidden Neurons	10
Number of Hidden Layers	1
Number of Weight Parameters	70
Number of Bias Parameters	12
Architecture	MLP
Number of Neural Nets	1
Objective Value	2.086877
Misclassification Error for Validation (%)	17.9178

Iterations	Objective	Loss	Validation Error
1	3.0781	2.7299	0.4837
2	3.0202	2.6730	0.3458
3	2.9692	2.6228	0.4449
4	2.8525	2.5030	0.4462
5	2.7055	2.3547	0.2533
6	2.4849	2.1127	0.2216
7	2.3618	1.9617	0.1954
8	2.3139	1.8470	0.1978
9	2.2666	1.8102	0.1953
10	2.2598	1.7951	0.1954
11	2.2390	1.7909	0.1929
12	2.2296	1.7784	0.1956
13	2.2224	1.7748	0.1892
14	2.2167	1.7697	0.1903
15	2.1984	1.7591	0.1918
16	2.1824	1.7566	0.1916
17	2.1753	1.7559	0.1927
18	2.1696	1.7576	0.1914
19	2.1646	1.7551	0.1903
20	2.1563	1.7597	0.1938
21	2.1518	1.7574	0.1934
22	2.1378	1.7473	0.1929
23	2.1332	1.7416	0.1912
24	2.1313	1.7324	0.1879
25	2.1297	1.7359	0.1898
26	2.1291	1.7378	0.1905
27	2.1226	1.7396	0.1901
28	2.1156	1.7406	0.1907
29	2.1129	1.7436	0.1887
30	2.1105	1.7376	0.1856
31	2.1091	1.7283	0.1847
32	2.1072	1.7268	0.1845
33	2.1032	1.7196	0.1872
34	2.1019	1.7145	0.1837
35	2.1009	1.7069	0.1836
36	2.1005	1.7051	0.1837
37	2.0997	1.7023	0.1841
38	2.0975	1.6960	0.1837
39	2.0961	1.6911	0.1817
40	2.0942	1.6852	0.1828
41	2.0928	1.6794	0.1801
42	2.0925	1.6806	0.1792
43	2.0922	1.6815	0.1794
44	2.0916	1.6809	0.1803
45	2.0914	1.6789	0.1795
46	2.0912	1.6767	0.1795
47	2.0911	1.6758	0.1795
48	2.0910	1.6753	0.1792
49	2.0908	1.6755	0.1797
50	2.0908	1.6756	0.1801
51	2.0907	1.6756	0.1810
52	2.0905	1.6750	0.1805
53	2.0904	1.6721	0.1803
54	2.0903	1.6711	0.1794
55	2.0902	1.6717	0.1792
56	2.0902	1.6721	0.1792
57	2.0901	1.6728	0.1792
58	2.0899	1.6716	0.1790
59	2.0898	1.6707	0.1795
60	2.0896	1.6684	0.1792
61	2.0895	1.6671	0.1779
62	2.0894	1.6666	0.1777
63	2.0893	1.6631	0.1784
64	2.0892	1.6644	0.1786
65	2.0891	1.6656	0.1790
66	2.0890	1.6647	0.1786
67	2.0890	1.6655	0.1788
68	2.0889	1.6658	0.1790
69	2.0887	1.6648	0.1792
70	2.0885	1.6658	0.1797
71	2.0881	1.6645	0.1777
72	2.0880	1.6635	0.1781
73	2.0880	1.6650	0.1784
74	2.0879	1.6675	0.1794
75	2.0878	1.6663	0.1788
76	2.0876	1.6653	0.1784
77	2.0873	1.6653	0.1781
78	2.0871	1.6673	0.1792
79	2.0870	1.6685	0.1792
80	2.0869	1.6705	0.1794
81	2.0869	1.6695	0.1792

Reason

The optimization achieved the desired objective value.

Neural network - type 1 Supplement 4

Predicted	Observed	Training Frequency	Training Percentage	Validation Frequency	Validation Percentage
conventional	conventional	5,211	81.47%	2,258	82.71%
organic	conventional	1,185	18.53%	472	17.29%
conventional	organic	1,162	18.22%	509	18.54%
organic	organic	5,216	81.78%	2,236	81.46%

Neural network - type 1 Supplement 5

Percentile	Training Observations	Training Events	Training Lift	Training Lift Best	Training Cumulative Lift	Training Cumulative Lift Best	Validation Observations	Validation Events	Validation Lift	Validation Lift Best	Validation Cumulative Lift	Validation Cumulative Lift Best
5.00	639	636	1.9944	2.0038	1.9944	2.0038	274	273	1.9891	1.9964	1.9891	1.9964
10.00	639	621	1.9473	2.0038	1.9708	2.0038	274	263	1.9162	1.9964	1.9526	1.9964
15.00	639	613	1.9222	2.0038	1.9546	2.0038	274	252	1.8361	1.9964	1.9138	1.9964
20.00	639	570	1.7874	2.0038	1.9128	2.0038	274	248	1.8069	1.9964	1.8871	1.9964
25.00	639	541	1.6965	2.0038	1.8696	2.0038	274	239	1.7413	1.9964	1.8579	1.9964
30.00	639	500	1.5679	2.0038	1.8193	2.0038	274	230	1.6758	1.9964	1.8276	1.9964
35.00	639	488	1.5303	2.0038	1.7780	2.0038	274	209	1.5228	1.9964	1.7840	1.9964
40.00	639	462	1.4487	2.0038	1.7368	2.0038	274	206	1.5009	1.9964	1.7486	1.9964
45.00	639	415	1.3013	2.0038	1.6884	2.0038	274	174	1.2678	1.9964	1.6952	1.9964
50.00	639	364	1.1414	1.9661	1.6337	2.0000	274	160	1.1658	1.9964	1.6423	1.9964
55.00	639	317	0.9940	0.0000	1.5756	1.8182	274	121	0.8816	0.0364	1.5731	1.8182
60.00	639	232	0.7275	0.0000	1.5049	1.6667	274	105	0.7650	0.0000	1.5058	1.6667
65.00	639	168	0.5268	0.0000	1.4297	1.5385	274	74	0.5392	0.0000	1.4314	1.5385
70.00	639	112	0.3512	0.0000	1.3526	1.4286	274	51	0.3716	0.0000	1.3557	1.4286
75.00	639	94	0.2948	0.0000	1.2821	1.3333	274	45	0.3279	0.0000	1.2872	1.3333
80.00	639	70	0.2195	0.0000	1.2157	1.2500	274	34	0.2477	0.0000	1.2222	1.2500
85.00	639	61	0.1913	0.0000	1.1554	1.1765	274	26	0.1894	0.0000	1.1615	1.1765
90.00	639	72	0.2258	0.0000	1.1038	1.1111	274	19	0.1384	0.0000	1.1046	1.1111
95.00	639	31	0.0972	0.0000	1.0508	1.0526	274	13	0.0947	0.0000	1.0515	1.0526
100.00	633	11	0.0345	0.0000	1.0000	1.0000	269	3	0.0219	0.0000	1.0000	1.0000

Cutoff	Training Sensitivity	Training 1 - Specificity	Training KS (Youden)	Validation Sensitivity	Validation 1 - Specificity	Validation KS (Youden)
0.00	1.0000	1.000		1.0000	1.000	
0.01	1.0000	0.976		1.0000	0.975	
0.02	1.0000	0.959		1.0000	0.956	
0.03	1.0000	0.943		0.9996	0.942	
0.04	0.9994	0.921		0.9993	0.920	
0.05	0.9983	0.898		0.9989	0.898	
0.06	0.9961	0.875		0.9985	0.877	
0.07	0.9950	0.849		0.9971	0.853	
0.08	0.9948	0.828		0.9949	0.830	
0.09	0.9929	0.802		0.9942	0.806	
0.10	0.9911	0.785		0.9927	0.783	
0.11	0.9882	0.763		0.9913	0.764	
0.12	0.9860	0.741		0.9891	0.744	
0.13	0.9821	0.718		0.9876	0.718	
0.14	0.9796	0.696		0.9847	0.693	
0.15	0.9770	0.672		0.9811	0.670	
0.16	0.9740	0.647		0.9796	0.640	
0.17	0.9715	0.622		0.9763	0.612	
0.18	0.9694	0.597		0.9727	0.587	
0.19	0.9668	0.574		0.9672	0.558	
0.20	0.9638	0.553		0.9650	0.534	
0.21	0.9595	0.532		0.9603	0.511	
0.22	0.9561	0.511		0.9559	0.488	
0.23	0.9528	0.494		0.9530	0.469	
0.24	0.9490	0.475		0.9490	0.451	
0.25	0.9472	0.456		0.9468	0.433	
0.26	0.9429	0.440		0.9446	0.416	
0.27	0.9403	0.422		0.9388	0.401	
0.28	0.9367	0.404		0.9359	0.388	
0.29	0.9327	0.388		0.9308	0.373	
0.30	0.9296	0.375		0.9264	0.358	
0.31	0.9255	0.361		0.9231	0.344	
0.32	0.9213	0.349		0.9202	0.333	
0.33	0.9177	0.337		0.9162	0.320	
0.34	0.9128	0.324		0.9075	0.307	
0.35	0.9084	0.312		0.9035	0.298	
0.36	0.9036	0.301		0.9002	0.288	
0.37	0.8976	0.290		0.8951	0.278	
0.38	0.8920	0.279		0.8889	0.267	
0.39	0.8859	0.267		0.8845	0.259	
0.40	0.8796	0.258		0.8776	0.246	
0.41	0.8733	0.249		0.8710	0.238	
0.42	0.8677	0.241		0.8648	0.233	
0.43	0.8619	0.231		0.8594	0.224	
0.44	0.8559	0.222		0.8539	0.216	
0.45	0.8506	0.216		0.8481	0.209	
0.46	0.8470	0.211		0.8441	0.203	
0.47	0.8404	0.204	Yes	0.8397	0.194	Yes
0.48	0.8341	0.197		0.8306	0.188	
0.49	0.8264	0.191		0.8240	0.180	
0.50	0.8178	0.185		0.8146	0.173	
0.51	0.8075	0.177		0.8029	0.165	
0.52	0.7985	0.170		0.7913	0.157	
0.53	0.7916	0.163		0.7789	0.147	
0.54	0.7817	0.157		0.7723	0.141	
0.55	0.7717	0.151		0.7639	0.138	
0.56	0.7612	0.144		0.7563	0.134	
0.57	0.7535	0.137		0.7486	0.128	
0.58	0.7468	0.130		0.7392	0.122	
0.59	0.7372	0.125		0.7315	0.117	
0.60	0.7256	0.119		0.7188	0.108	
0.61	0.7139	0.113		0.7064	0.103	
0.62	0.7016	0.108		0.6987	0.099	
0.63	0.6919	0.103		0.6903	0.093	
0.64	0.6794	0.099		0.6801	0.090	
0.65	0.6665	0.094		0.6681	0.085	
0.66	0.6522	0.089		0.6565	0.082	
0.67	0.6397	0.085		0.6441	0.078	
0.68	0.6287	0.080		0.6295	0.075	
0.69	0.6149	0.076		0.6189	0.072	
0.70	0.6016	0.070		0.6058	0.068	
0.71	0.5848	0.066		0.5913	0.064	
0.72	0.5663	0.062		0.5727	0.060	
0.73	0.5553	0.058		0.5592	0.054	
0.74	0.5414	0.054		0.5461	0.050	
0.75	0.5265	0.049		0.5268	0.046	
0.76	0.5119	0.044		0.5097	0.043	
0.77	0.4945	0.040		0.4914	0.040	
0.78	0.4828	0.037		0.4791	0.038	
0.79	0.4672	0.033		0.4638	0.035	
0.80	0.4478	0.031		0.4448	0.033	
0.81	0.4354	0.027		0.4284	0.030	
0.82	0.4156	0.023		0.4084	0.026	
0.83	0.4001	0.020		0.3913	0.025	
0.84	0.3871	0.018		0.3756	0.022	
0.85	0.3725	0.016		0.3548	0.020	
0.86	0.3547	0.013		0.3359	0.018	
0.87	0.3373	0.010		0.3177	0.015	
0.88	0.3112	0.008		0.2944	0.013	
0.89	0.2816	0.006		0.2623	0.010	
0.90	0.2579	0.005		0.2434	0.007	
0.91	0.2327	0.004		0.2153	0.005	
0.92	0.2013	0.003		0.1880	0.004	
0.93	0.1789	0.002		0.1654	0.004	
0.94	0.1524	0.001		0.1403	0.003	
0.95	0.1270	0.001		0.1162	0.001	
0.96	0.0978	0.000		0.0922	0.000	
0.97	0.0696	0.000		0.0703	0.000	
0.98	0.0419	0.000		0.0419	0.000	
0.99	0.0174	0.000		0.0138	0.000	

Neural network - type 1 Supplement 7

Response	Event	Value	Training Frequency	Validation Frequency
Correct	organic	True Positive	5,216	2,236
Incorrect	organic	False Negative	1,162	509
Correct	conventional	True Negative	5,211	2,258
Incorrect	conventional	False Positive	1,185	472

Neural network - type 1 Supplement 8

Partition	KS (Youden)	Misclassification Rate	Misclassification Rat...	C Statistic	FPR	FDR	F1 Score	Lift	Cumulative Lift	Cumulative % Events	Cumulative % Captu...	Gain	Gini	Gamma	Tau	Observations Used	Unused
Training	0.6368	0.1837	0.1837	0.893	0.185	0.185	0.816	1.994	1.994	99.531	9.972	0.994	0.787	0.791	0.393	12,774	0
Validation	0.6459	0.1792	0.1792	0.896	0.173	0.174	0.820	1.989	1.989	99.635	9.945	0.989	0.793	0.797	0.397	5,475	0

Plot	Summary
Confusion Matrix	The confusion matrix plot displays the number of observations predicting each response level. A greater number of observations where the observed level and predicted level are the same indicates a better model. For this data, the percentages of each observed value that are correctly predicted in the validation partition are as follows: conventional - 82.71%, and organic - 81.46%.
Lift	The lift plot measures the ratio of percent captured response to the baseline percent response. The validation partition has a lift of 1.99 at the 5% quantile meaning there are about 1.989 times more events in that quantile than expected by random (5% of the total number of events).
Cumulative Lift	Cumulative lift measures the ratio of percent captured response to the baseline percent response, up to and including the specified quantile. The validation partition has a cumulative lift of 1.95 in the 10% quantile meaning there are about 1.953 times more events in the first two quantiles than expected by random (10% of the total number of events). Because this value is greater than 1, it is better to use your model to identify responders than no model, based on the validation partition.
ROC	The receiver operator characteristic (ROC) is a plot of sensitivity (the true positive rate) against 1-specificity (the false positive rate), which are both measures of classification based on the confusion matrix. These measures are calculated at various cutoff values. To help identify the best cutoff to use when scoring your data, the KS cutoff reference line is drawn at the value of 1-specificity where the greatest difference between sensitivity and 1-specificity is observed for the validation partition. The KS cutoff line is drawn at the cutoff value 0.47 where the 1-specificity value is 0.194 and the sensitivity value is 0.84. Cutoff values range from 0 to 0.99, inclusive, in increments of 0.01. At each cutoff value, the predicted response classification is determined by whether the predicted probability of the response type being organic is greater than or equal to the cutoff value. When the predicted probability of the event is greater than or equal to the cutoff value, then the predicted classification is organic, otherwise it is NOT organic.
Misclassification	The misclassification plot is a visual representation of the accuracy of the prediction at the specified cutoff value, 0.50. The plot displays the number of true positives for events that are correctly classified, false positives for NOT events that are classified as events, false negatives for events that are classified as NOT events, and true negatives for NOT events that are classified as NOT events. True negatives include NOT event classifications that predict a different level from observed, as long as both are NOT events. The predicted response classification is determined by whether the predicted probability of the level organic for the response type is greater than or equal to the cutoff value. When it is greater than or equal to the cutoff value, the predicted classification is an event, otherwise it is a NOT event. For this data, for the bar corresponding to the event level of type, organic, the segment of the bar colored as "Correct" corresponds to true positives.

Model comparison 1 Supplement 1

Selected	Model	Visualization Type	Number Of Observa...	Percentile	Prediction cutoff	C Statistic	Cumulative % Captu...	Cumulative % Events	Cumulative Lift	F1 Score	FDR	FPR	Gain	Gamma	Gini	KS (Youden)	Lift	Misclassification Rate	Misclassification Rat...	Tau
Yes	Gradient boosting - type 1	Gradient Boosting	18,249	5	0.5	0.997	10.008	100.000	2.0015	0.966	0.027	0.027	1.002	0.993	0.993	0.937	2.0015	0.033	0.033	0.497
No	Neural network - type 1	Neural Network	18,249	5	0.5	0.892	9.953	99.452	1.9906	0.817	0.184	0.184	0.991	0.788	0.784	0.634	1.9906	0.183	0.183	0.392

Model comparison 1 Supplement 2

Variable	Gradient boosting - type 1 Rank	Neural network - type 1 Rank	Gradient boosting - type 1 Importance	Neural network - type 1 Importance
TotalVolume	1		460.171712	
XLargeBags	2		49.397103	
AveragePrice	3		46.067248	
TotalBags	4		16.837841	
LargeBags	5		10.026202	
SmallBags	6		9.831606	

Model comparison 1 Supplement 3

Variable	Importance	Model
AveragePrice	1.0000	Gradient boosting - type 1
TotalVolume	0.4536	Gradient boosting - type 1
TotalBags	0.3566	Gradient boosting - type 1
SmallBags	0.3530	Gradient boosting - type 1
LargeBags	0.2582	Gradient boosting - type 1
XLargeBags	0.1804	Gradient boosting - type 1
AveragePrice	1.0000	Neural network - type 1
TotalVolume	0.1632	Neural network - type 1
TotalBags	0.1440	Neural network - type 1
SmallBags	0.1302	Neural network - type 1
LargeBags	0.1145	Neural network - type 1
XLargeBags	0.0521	Neural network - type 1

Champion Model Summary

The champion model is Gradient boosting - type 1. The model was chosen based on F1 Score cutoff 0.5 (0.97). 96.65% of the data was correctly classified using the Gradient boosting - type 1 model with a prediction cutoff value of 0.5. The six most important factors with greater than 10% relative importance are AveragePrice, TotalVolume, TotalBags, SmallBags, LargeBags, and XLargeBags.