Input data

Features: hotel, is_canceled, lead_time, arrival_date_year, arrival_date_month, arrival_date_week_number, arrival_date_day_of_month, stays_in_weekend_nights, stays_in_week_nights, adults, children, babies, meal, country, market_segment, distribution_channel, is_repeated_guest, previous_cancellations, previous_bookings_not_canceled, reserved_room_type, assigned_room_type, booking_changes, deposit_type, company, days_in_waiting_list, customer_type, adr, required_car_parking_spaces, total_of_special_requests, reservation_status, reservation_status_date (total: 31 features)
Meta attributes: agent

Output data

Features: hotel, customer_type, market_segment, distribution_channel, is_repeated_guest, reserved_room_type, adr, lead_time, previous_cancellations, arrival_date_month, booking_changes, stays_in_weekend_nights, stays_in_week_nights, total_of_special_requests, adults, deposit_type (total: 16 features)
Target: is_canceled

Removed: 15 (reservation_status, days_in_waiting_list, meal, country, babies, reservation_status_date, company, arrival_date_week_number, assigned_room_type, previous_bookings_not_canceled, arrival_date_day_of_month, required_car_parking_spaces, arrival_date_year, agent, children)

Name: Tree

Model parameters

Pruning: at least two instances in leaves, at least five instances in internal nodes, maximum depth 100
Splitting: Stop splitting when majority reaches 95% (classification only)
Binary trees: No

Data

Data instances: 2000
Features: hotel, customer_type, market_segment, distribution_channel, is_repeated_guest, reserved_room_type, adr, lead_time, previous_cancellations, arrival_date_month, booking_changes, stays_in_weekend_nights, stays_in_week_nights, total_of_special_requests, adults, deposit_type (total: 16 features)
Target: is_canceled

Name: kNN

Model parameters

Number of neighbours: 5
Metric: Euclidean
Weight: Uniform

Name: Neural Network

Model parameters

Hidden layers: 100
Activation: ReLu
Solver: Adam
Alpha: 0.0001
Max iterations: 200
Replicable training: True

Name: Random Forest

Model parameters

Number of trees: 10
Maximal number of considered features: unlimited
Replicable training: No
Maximal tree depth: unlimited
Stop splitting nodes with maximum instances: 5

Data

Data instances: 2000
Features: hotel, customer_type, market_segment, distribution_channel, is_repeated_guest, reserved_room_type, adr, lead_time, previous_cancellations, arrival_date_month, booking_changes, stays_in_weekend_nights, stays_in_week_nights, total_of_special_requests, adults, deposit_type (total: 16 features)
Target: is_canceled

Name: Gradient Boosting

Model parameters

Method: Gradient Boosting (scikit-learn)
Number of trees: 100
Learning rate: 0.1
Replicable training: Yes
Maximum tree depth: 3
Fraction of training instances: 1
Stop splitting nodes with maximum instances: 2

Name: SVM

Model parameters

SVM type: SVM, C=1.0, ε=0.1
Kernel: RBF, exp(-auto|x-y|²)
Numerical tolerance: 0.001
Iteration limt: 100

Name: AdaBoost

Model parameters

Base estimator: tree
Number of estimators: 50
Algorithm (classification): Samme.r
Loss (regression): Linear

Name: Logistic Regression

Model parameters

Regularization: Ridge (L2), C=1, class weights=False
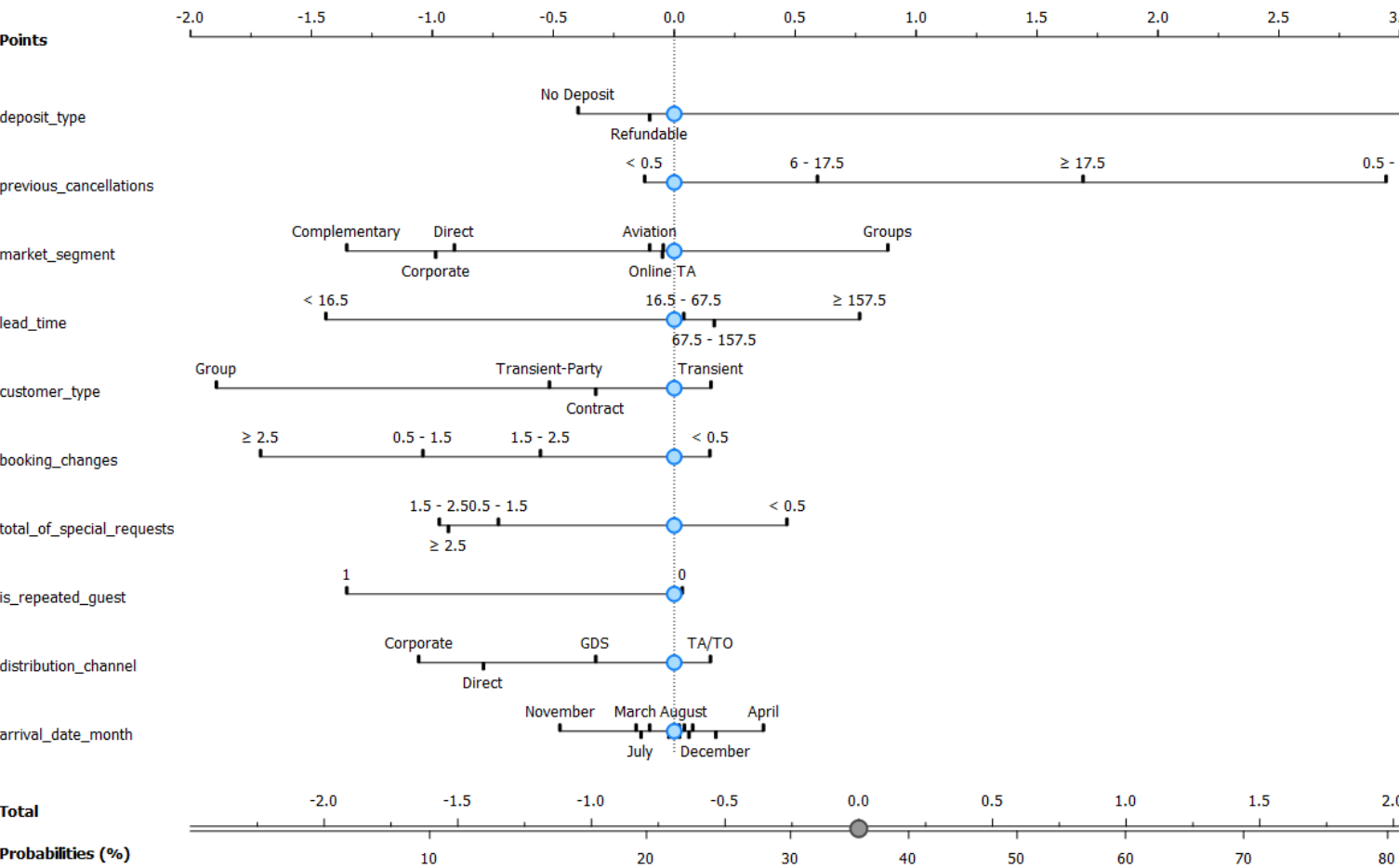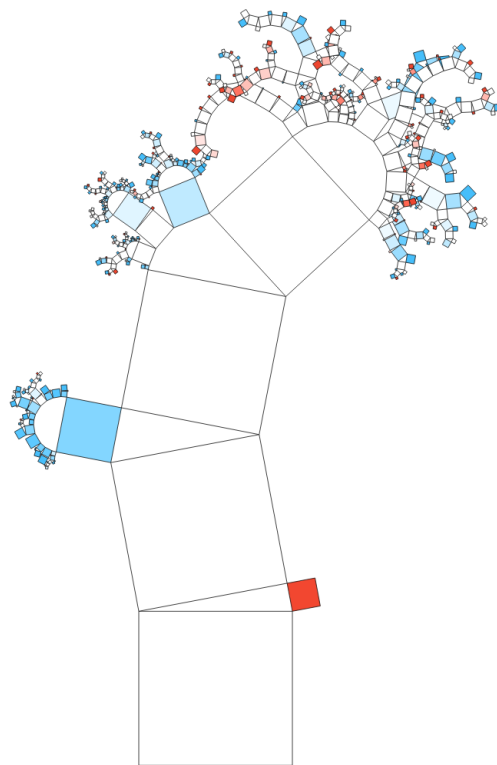
Data

Data instances: 2000
Features: hotel, customer_type, market_segment, distribution_channel, is_repeated_guest, reserved_room_type, adr, lead_time, previous_cancellations, arrival_date_month, booking_changes, stays_in_weekend_nights, stays_in_week_nights, total_of_special_requests, adults, deposit_type (total: 16 features)
Target: is_canceled

Tree size: 628 nodes, 357 leaves
Edge widths: Relative to parent
Target class: 1

≤ 11

9.0%, 38/424

arrival_date_month

April	August	December	February	January	July	Ju
19.0%, 4/21	5.4%, 2/37	6.7%, 2/30	3.7%, 2/54	9.1%, 4/44	4.5%, 1/22	7.1%, 2/28
adr	customer_type	reserved_room_type		adr		hotel
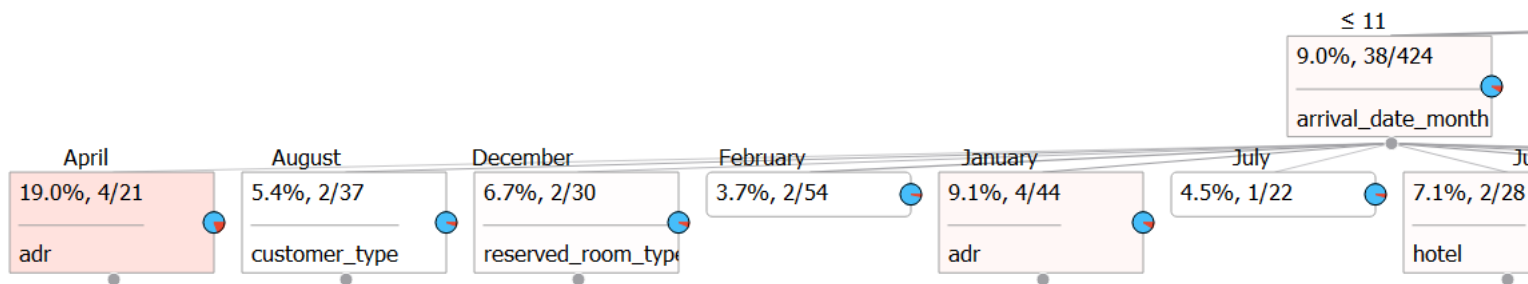
Settings

Sampling type: Stratified 3-fold Cross validation
Target class: None, show average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
kNN	0.6624742654756088	0.6535	0.6421715891355101	0.6388438160574605	0.6535
Tree	0.7124248682741294	0.7415	0.7373121535005172	0.7360901818181818	0.7415
Stack	0.8438688586084164	0.8065	0.7995365386573069	0.8055433525647089	0.8065
SVM	0.6511183613650638	0.6305	0.6298473104813086	0.6292318397011197	0.6305
Random Forest	0.810457594737944	0.7785	0.7681770692148525	0.7765589898263188	0.7785
Neural Network	0.8117323260520622	0.759	0.7559581408798545	0.75488287246684	0.759
Naive Bayes	0.792109205980878	0.75	0.7417590792153749	0.743587955120668	0.75
Logistic Regression	0.8195126578965735	0.795	0.7836570832303595	0.797769395196781	0.795
Gradient Boosting	0.844617994975225	0.8015	0.7933308064973943	0.801168617660585	0.8015
Constant	0.49927321777514133	0.644	0.5045450121654501	0.414736	0.644
AdaBoost	0.7369599021215717	0.7525	0.7511294173599707	0.7501840998306499	0.7525

Target class: 0
Costs: FP = 500, FN = 500
Target probability: 64.0 %

Confusion matrix for Random Forest (showing proportion of predicted)
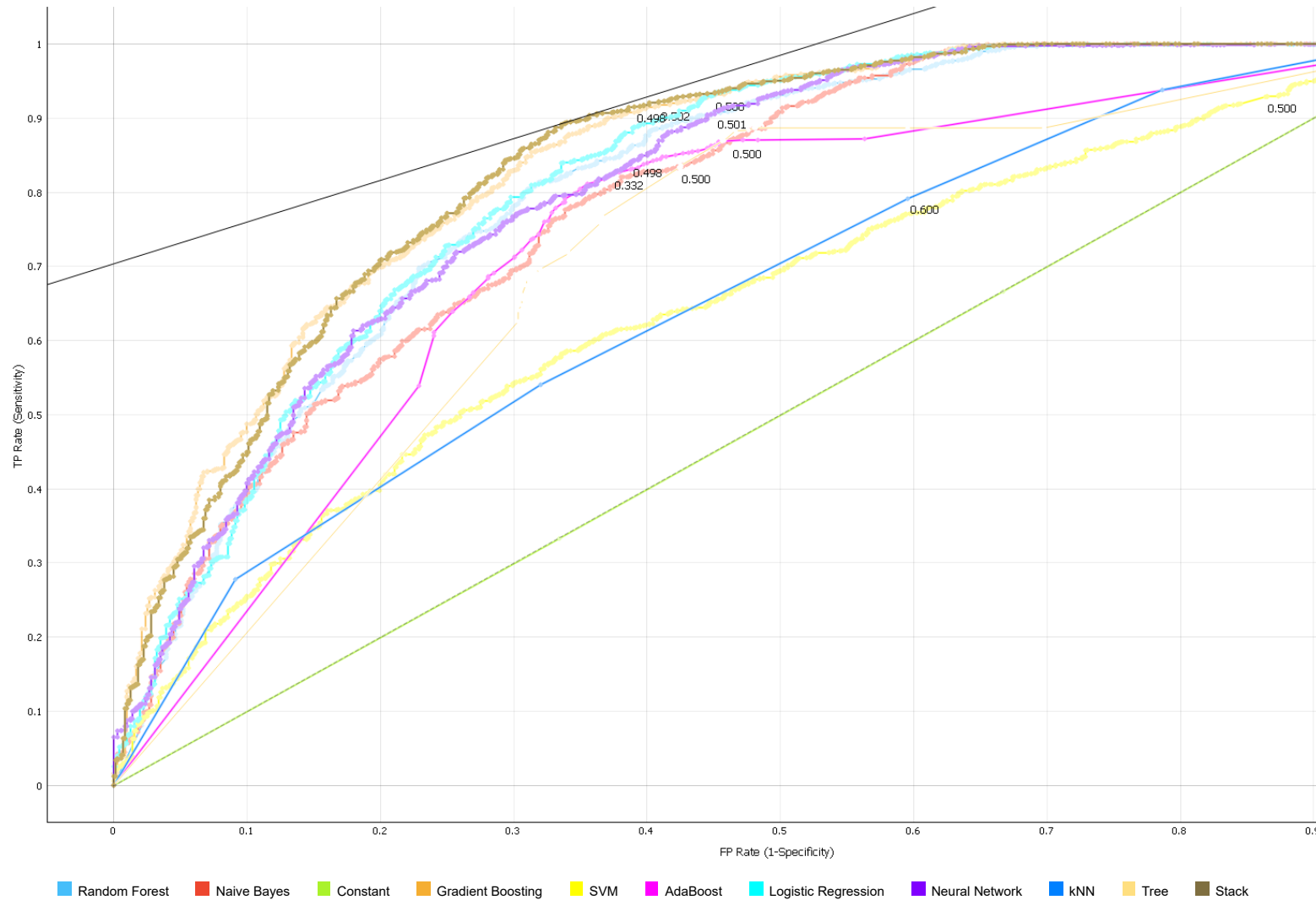
		Predicted		
		0	**1**	**∑**
Actual	**0**	78.3 %	23.6 %	1288
	1	21.7 %	76.4 %	712
	∑	1491	509	2000

Target class: 1
Costs: FP = 500, FN = 500
Target probability: 36.0 %
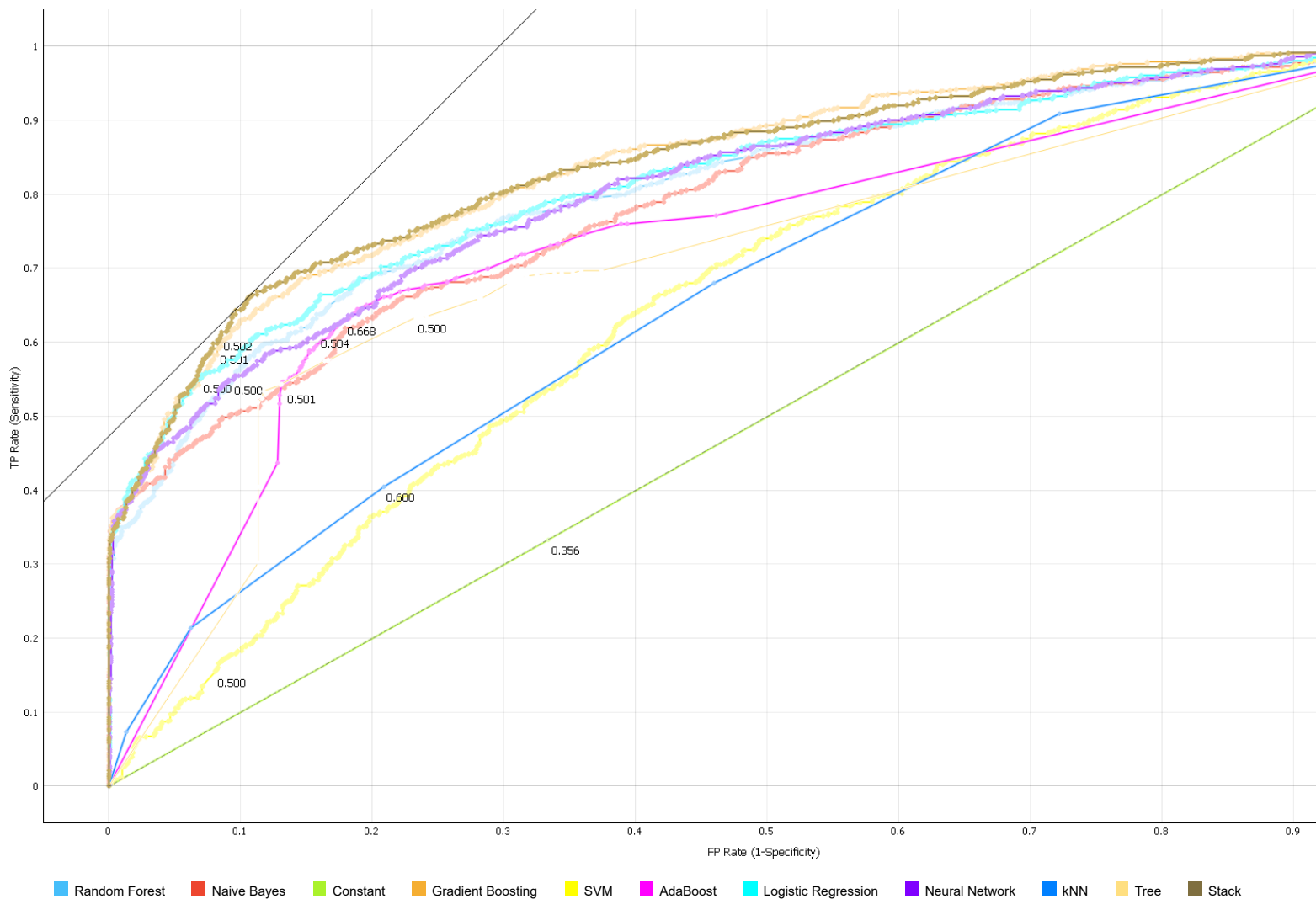
Target class: 1

Pos and neg predictive value

| Random Forest | Naive Bayes | Constant | Gradient Boosting | SVM | AdaBoost | Logistic Regression | Neural Network | kNN | Tree | Stack |
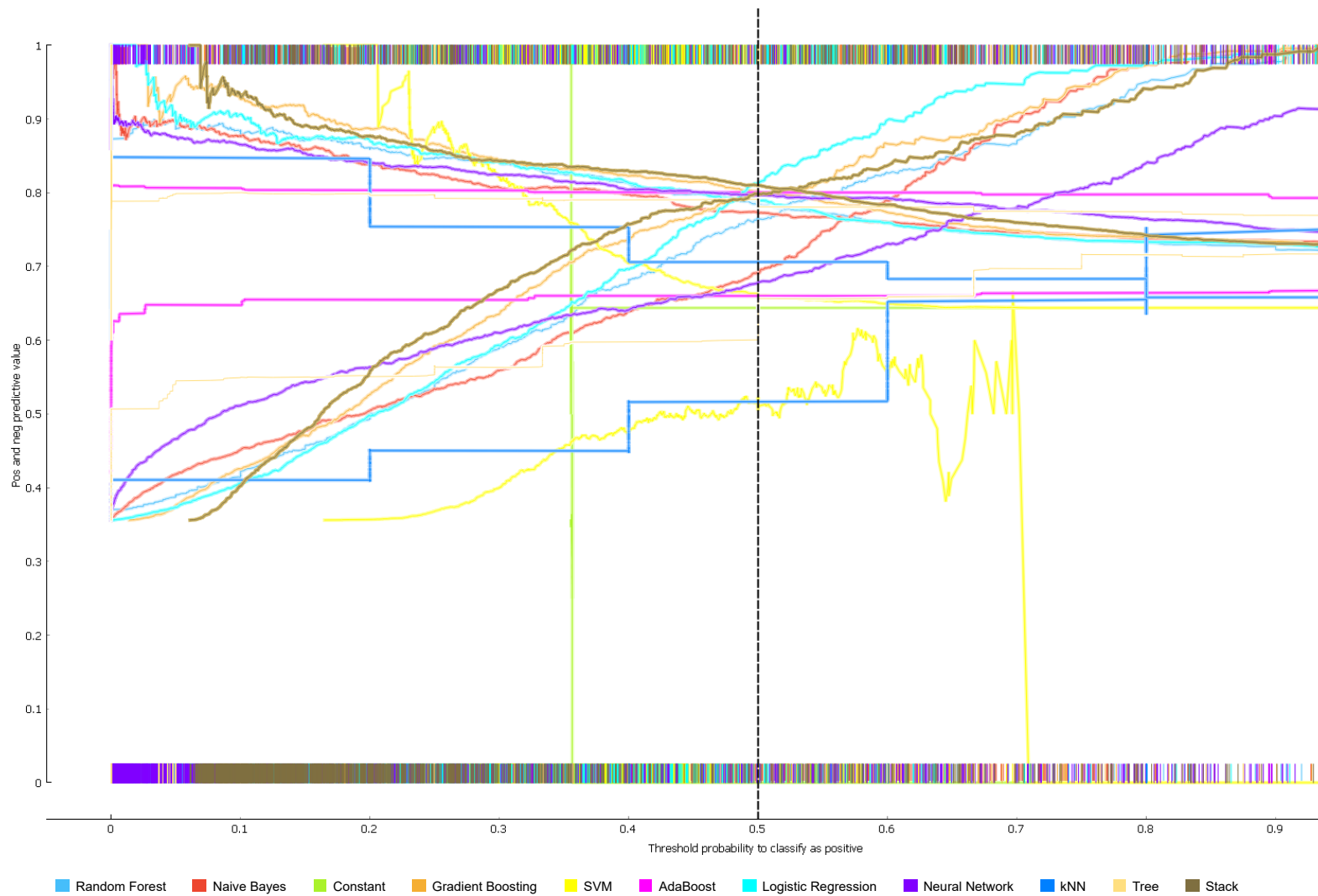
Pos and neg predictive value

Threshold: p = 0.50

	PPV		TPV
Random Forest:	0.765	/	0.784
Naive Bayes:	0.691	/	0.773
Constant:	0.000	/	0.644
Gradient Boosting:	0.799	/	0.802
SVM:	0.514	/	0.663
AdaBoost:	0.660	/	0.800
Logistic Regression:	0.813	/	0.789
Neural Network:	0.679	/	0.797
kNN:	0.517	/	0.706
Tree:	0.601	/	0.791
Stack:	0.798	/	0.810

Target class: 1
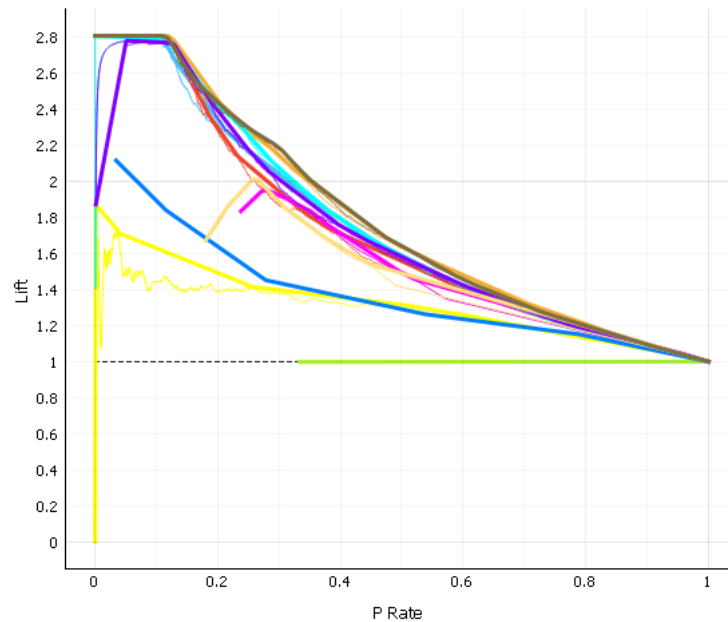
| | Random Forest | | Naive Bayes | | Constant | | Gradient Boosting | | SVM | | AdaBoost | | Logistic Regression | | Neural Network | | kNN | | Tree | | Stack |

Info

Data:
hotel_bookings dataset (sample): 2000 instances, 17 variables
Features: 16 (8 categorical, 8 numeric) (no missing values)
Target: categorical

Model: 1 model

- Gradient Boosting (1)

Showing probabilities for all classes that appear in the data

Data & Predictions

	Gradient Boosting (1)	is_canceled	hotel	customer_type	market_segment	distribution_channel	is_repeated_guest	reserved_room_type	adr	lead_time	previous_cancellations
1	0.86 : 0.14 → 0	0	City Hotel	Transient-Party	Groups	TA/TO	0	A	90	116	0
2	0.72 : 0.28 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	105.3	113	0
3	0.72 : 0.28 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	G	253.33	66	0
4	0.91 : 0.09 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	104	17	0
5	0.92 : 0.08 → 0	0	Resort Hotel	Transient-Party	Groups	TA/TO	0	D	67	203	0
6	0.95 : 0.05 → 0	0	Resort Hotel	Transient-Party	Groups	Corporate	0	A	36	2	0
7	0.84 : 0.16 → 0	1	Resort Hotel	Transient-Party	Groups	Corporate	0	A	55	131	0
8	0.77 : 0.23 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	51.4	148	0
9	0.83 : 0.17 → 0	0	Resort Hotel	Transient-Party	Groups	TA/TO	0	A	54.5	197	0
10	0.79 : 0.21 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	125.1	33	0
11	0.93 : 0.07 → 0	0	City Hotel	Transient-Party	Groups	TA/TO	0	A	90	174	0
12	0.83 : 0.17 → 0	0	Resort Hotel	Transient-Party	Offline TA/TO	TA/TO	0	A	25	46	0
13	0.02 : 0.98 → 1	1	Resort Hotel	Transient	Groups	TA/TO	0	D	89	238	0
14	0.43 : 0.57 → 1	1	Resort Hotel	Transient	Online TA	TA/TO	0	A	48	65	0
15	0.65 : 0.35 → 0	0	City Hotel	Transient-Party	Direct	Direct	0	A	160	8	0

#	Prob	Pred	Hotel	Customer Type	Market Segment	Channel		Room			
16	0.58 : 0.42 → 0	1	City Hotel	Transient	Online TA	TA/TO	0	A	185.67	8	0
17	0.02 : 0.98 → 1	1	Resort Hotel	Transient	Groups	TA/TO	0	A	70	74	0
18	0.31 : 0.69 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	A	162	121	0
19	0.81 : 0.19 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	109.8	55	0
20	0.87 : 0.13 → 0	0	Resort Hotel	Transient-Party	Online TA	TA/TO	0	A	52.71	166	0
21	0.92 : 0.08 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	97.5	3	0
22	0.66 : 0.34 → 0	0	Resort Hotel	Transient	Direct	Direct	0	A	93.4	65	0
23	0.77 : 0.23 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	48	11	0
24	0.67 : 0.33 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	D	195	141	0
25	0.43 : 0.57 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	D	94.4	24	0
26	0.36 : 0.64 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	A	69.6	56	0
27	0.25 : 0.75 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	D	126	59	0
28	0.43 : 0.57 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	A	76.5	103	0
29	0.86 : 0.14 → 0	0	City Hotel	Transient-Party	Groups	TA/TO	0	A	84	271	0
30	0.90 : 0.10 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	54	0	0
31	0.73 : 0.27 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	140	49	0
32	0.74 : 0.26 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	H	309	226	0
33	0.02 : 0.98 → 1	1	City Hotel	Transient	Groups	TA/TO	0	A	62	468	0
34	0.73 : 0.27 → 0	1	City Hotel	Transient	Online TA	TA/TO	0	A	141.3	133	0
35	0.80 : 0.20 → 0	0	Resort Hotel	Transient-Party	Groups	Corporate	0	A	36	19	0
36	0.62 : 0.38 → 0	0	City Hotel	Transient-Party	Groups	TA/TO	0	A	62	59	0
37	0.80 : 0.20 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	37.8	116	0
38	0.65 : 0.35 → 0	1	City Hotel	Transient	Online TA	TA/TO	0	A	115	26	0
39	0.44 : 0.56 → 1	0	City Hotel	Transient	Online TA	TA/TO	0	A	129	28	0
40	0.61 : 0.39 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	138.5	15	0
41	0.94 : 0.06 → 0	0	Resort Hotel	Transient-Party	Groups	Direct	0	A	60	0	0
42	0.97 : 0.03 → 0	0	City Hotel	Transient-Party	Offline TA/TO	TA/TO	0	A	60	5	0
43	0.80 : 0.20 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	D	95.48	38	0
44	0.66 : 0.34 → 0	1	City Hotel	Transient	Online TA	TA/TO	0	D	182.33	0	0
45	0.69 : 0.31 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	112	33	0
46	0.01 : 0.99 → 1	1	City Hotel	Contract	Groups	TA/TO	0	A	62	332	1
47	0.01 : 0.99 → 1	1	City Hotel	Transient	Offline TA/TO	TA/TO	0	A	90	329	1
48	0.81 : 0.19 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	D	152	6	0
49	0.03 : 0.97 → 1	1	City Hotel	Transient	Offline TA/TO	TA/TO	0	A	90	108	0
50	0.37 : 0.63 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	F	168	81	0
51	0.32 : 0.68 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	A	119.85	90	0
52	0.81 : 0.19 → 0	1	City Hotel	Transient	Online TA	TA/TO	0	A	94.5	134	0
53	0.66 : 0.34 → 0	1	Resort Hotel	Transient	Offline TA/TO	TA/TO	0	A	26.1	118	0
54	0.69 : 0.31 → 0	1	City Hotel	Transient-Party	Groups	TA/TO	0	A	120	118	0
55	0.80 : 0.20 → 0	0	City Hotel	Transient	Direct	Direct	0	A	120	9	0
56	0.92 : 0.08 → 0	0	Resort Hotel	Transient-Party	Offline TA/TO	TA/TO	0	A	48.2	128	0
57	0.73 : 0.27 → 0	1	City Hotel	Transient-Party	Offline TA/TO	TA/TO	0	A	160	11	0
58	0.36 : 0.64 → 1	1	Resort Hotel	Transient	Direct	Direct	0	A	185	45	0
59	0.39 : 0.61 → 1	0	City Hotel	Transient	Online TA	TA/TO	0	A	99.45	147	0
60	0.86 : 0.14 → 0	0	Resort Hotel	Transient	Direct	TA/TO	0	A	176.1	12	0
61	0.94 : 0.06 → 0	0	Resort Hotel	Transient	Direct	Direct	1	A	26.5	0	0
62	0.80 :	0	City	Transient	Online TA	TA/TO	0	E	205.2	99	0

	0.20 → 0		Hotel								
63	0.49 : 0.51 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	A	125.1	249	0
64	0.73 : 0.27 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	D	102.55	293	0
65	0.74 : 0.26 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	D	221	112	0
66	0.84 : 0.16 → 0	0	City Hotel	Transient	Offline TA/TO	TA/TO	0	A	75	16	0
67	0.76 : 0.24 → 0	0	City Hotel	Transient-Party	Groups	TA/TO	0	A	62	83	0
68	0.68 : 0.32 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	80	230	0
69	0.82 : 0.18 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	99.9	92	0
70	0.03 : 0.97 → 1	1	Resort Hotel	Transient	Groups	TA/TO	0	A	46	86	0
71	0.95 : 0.05 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	A	89	3	0
72	0.64 : 0.36 → 0	1	Resort Hotel	Transient	Online TA	TA/TO	0	A	48	47	0
73	0.34 : 0.66 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	D	171	107	0
74	0.03 : 0.97 → 1	1	City Hotel	Transient	Offline TA/TO	TA/TO	0	A	75	105	0
75	0.90 : 0.10 → 0	0	City Hotel	Transient	Corporate	Corporate	0	A	120	3	0
76	0.68 : 0.32 → 0	0	Resort Hotel	Transient-Party	Groups	TA/TO	0	A	85	199	0
77	0.61 : 0.39 → 0	1	City Hotel	Transient	Online TA	TA/TO	0	A	125.1	244	0
78	0.24 : 0.76 → 1	1	Resort Hotel	Transient	Online TA	TA/TO	0	A	54	237	0
79	0.09 : 0.91 → 1	1	Resort Hotel	Transient	Online TA	TA/TO	0	A	55	0	0
80	0.39 : 0.61 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	A	105.3	61	0
81	0.38 : 0.62 → 1	1	Resort Hotel	Transient	Online TA	TA/TO	0	E	258.27	28	0
82	0.82 : 0.18 → 0	1	Resort Hotel	Transient	Direct	Direct	0	A	129.33	69	0
83	0.79 : 0.21 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	E	69	19	0
84	0.58 : 0.42 → 0	1	Resort Hotel	Transient-Party	Groups	Corporate	0	C	158	239	0
85	0.84 : 0.16 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	39	11	0
86	0.87 : 0.13 → 0	0	Resort Hotel	Transient	Direct	Direct	0	F	135.9	26	0
87	0.54 : 0.46 → 0	0	City Hotel	Transient	Direct	Direct	0	A	117	90	0
88	0.75 : 0.25 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	103.5	51	0
89	0.78 : 0.22 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	D	145	232	0
90	0.38 : 0.62 → 1	0	City Hotel	Transient	Online TA	TA/TO	0	A	120	109	0
91	0.01 : 0.99 → 1	1	City Hotel	Transient	Offline TA/TO	TA/TO	0	A	90	34	1
92	0.96 : 0.04 → 0	0	Resort Hotel	Transient-Party	Direct	Direct	0	A	70	3	0
93	0.90 : 0.10 → 0	0	Resort Hotel	Transient-Party	Groups	TA/TO	0	A	75.57	131	0
94	0.79 : 0.21 → 0	1	Resort Hotel	Transient	Offline TA/TO	TA/TO	0	A	55	30	0
95	0.20 : 0.80 → 1	1	Resort Hotel	Transient-Party	Online TA	TA/TO	0	G	137.5	91	0
96	0.02 : 0.98 → 1	1	City Hotel	Transient	Groups	TA/TO	0	A	100	128	0
97	0.73 : 0.27 → 0	0	City Hotel	Transient	Online TA	TA/TO	0	E	154.16	11	0
98	0.30 : 0.70 → 1	1	City Hotel	Transient	Online TA	TA/TO	0	A	116.1	130	0
99	0.88 : 0.12 → 0	0	City Hotel	Transient-Party	Groups	TA/TO	0	A	112.67	448	0
100	0.74 : 0.26 → 0	0	Resort Hotel	Transient	Online TA	TA/TO	0	A	46.1	115	0
	+ 1800 more										

Scores

Target class: (Average over classes)

Model	AUC	CA	F1	Precision	Recall
Gradient Boosting (1)	0.919	0.851	0.846	0.853	0.851

Color: adr
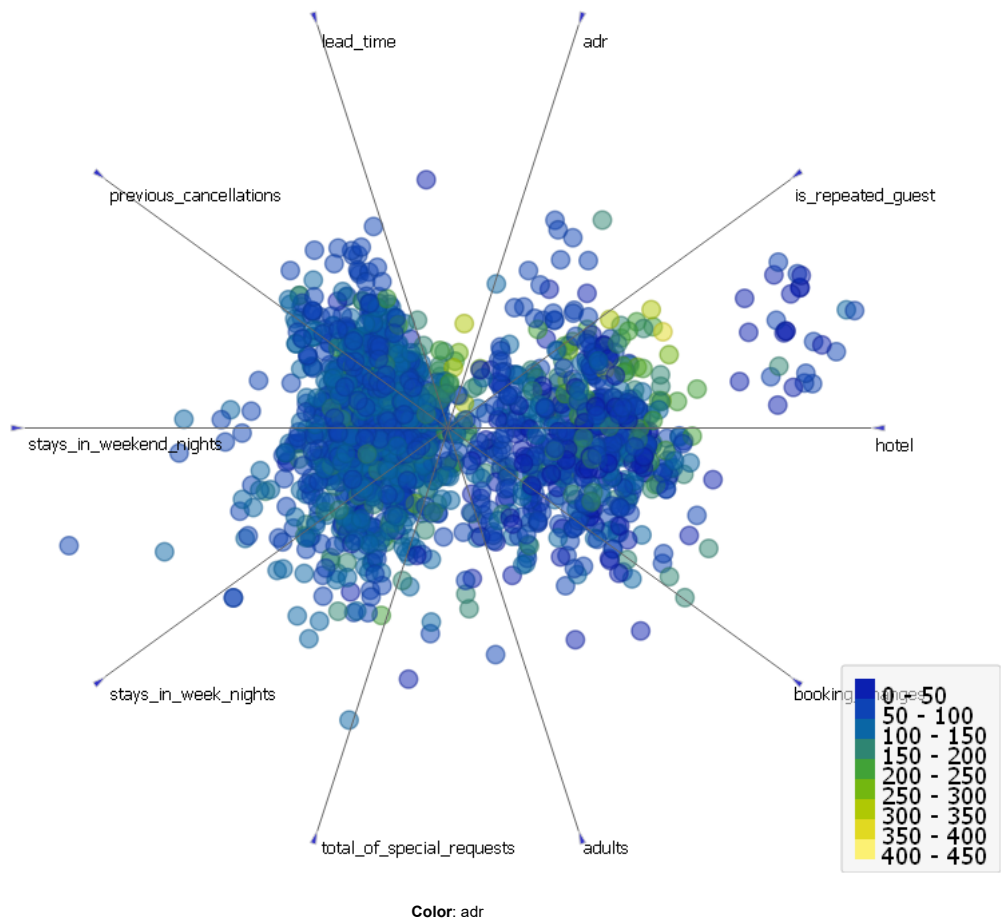
Score: AUC
Permutations: 5
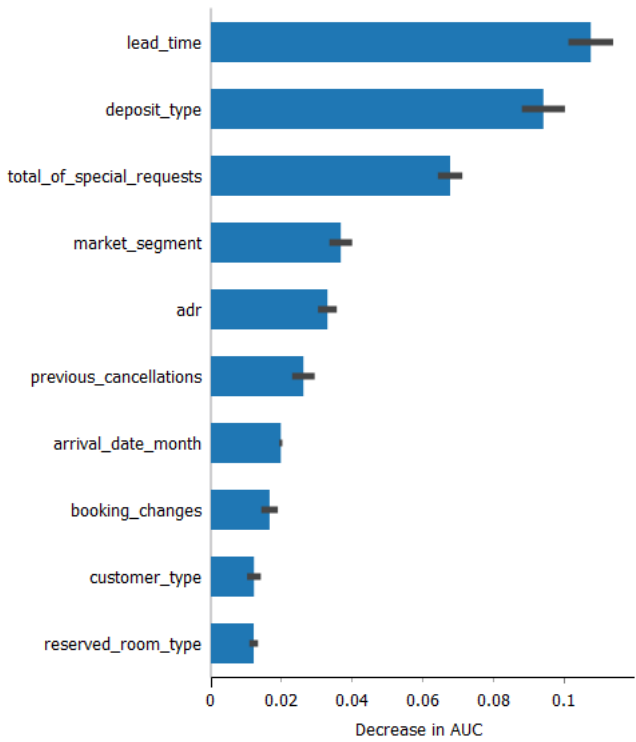
Target class: 0
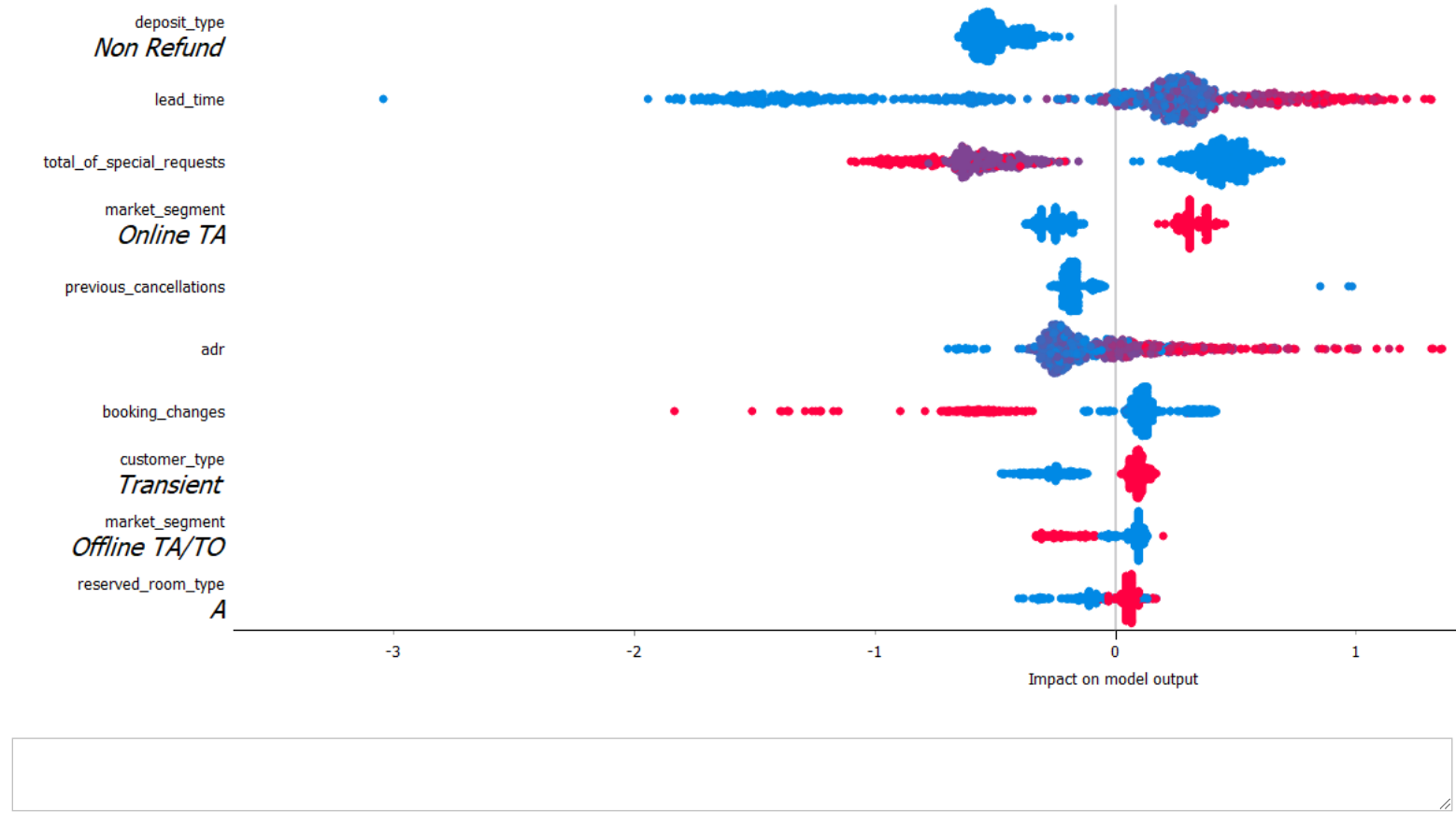
Data instances: 16
Features: 6
Meta attributes: 1

	Feature	Info. gain	Gain ratio	Gini	χ^2	ReliefF	FCBF
12	deposit_type	0.188755	0.355372	0.108676	410.168	0.232	0.345404
8	lead_time	0.0816309	0.0408179	0.0471645	154.277	0.0347273	5.88134e-06
13	total_of_special_requests	0.0597745	0.0409268	0.0362697	142.652	0.0135	5.24272e-06
9	previous_cancellations	0.0501743	0.174552	0.0311737	133.87	0.00130769	8.90871e-06
3	market_segment	0.0427653	0.0220399	0.0262962	1.0675	0.172	3.06107e-06
15	booking_changes	0.0248186	0.0314771	0.0138452	72.0463	0.00514286	2.95786e-06
4	distribution_channel	0.0206647	0.0251402	0.0120147	14.9434	0.024	0.0240292
2	customer_type	0.0152091	0.0153299	0.00856457	0.856252	0.05	1.6001e-06
11	stays_in_week_nights	0.00930196	0.0047828	0.00589798	4.23532	0.0124211	6.49219e-07
1	hotel	0.00920356	0.0099569	0.00573699	16.528	-0.008	9.97543e-07
14	adults	0.0072935	0.00725201	0.00435713	2.31801	-0.007	7.55628e-07
5	is_repeated_guest	0.00709089	0.0355643	0.00378117	15.9814	0.038	1.26115e-06
16	arrival_date_month	0.00568949	0.00160699	0.00358706	7.30608	0.012	2.54654e-07
6	reserved_room_type	0.00519624	0.00378129	0.00322043	34.115	0.012	4.51235e-07
10	stays_in_weekend_nights	0.00167158	0.000990369	0.00106012	2.16946	0.00766667	1.27417e-07
7	adr	0.000226332	0.000113166	0.00014342	0.438114	0.0068054	1.54027e-08

Feature Ranking

Target class: 0

previous_cancellations = 0

customer_type=Transient = 0

market_segment=Online TA = 0

adr = 90

deposit_type=Non Refund = 0

booking_changes = 1

total_of_special_requests = 0

lead_time = 116