

**Tipul studiului: clasificare**  
**Scafa Ioan**

**Scop:** estimarea acordarii titlului de superhost pentru cazari folosind metode de clasificare

**Sursa datelor:** <https://www.kaggle.com/datasets/dipeshkhemani/airbnb-cleaned-europe-dataset>

Variabile predictor:

Numele variabilei	Descriere
Price	Pretul pe o noapte
Day	Weekday sau Weekend
Shared Room	Camera se imparte cu alte persoane (0 - Nu, 1 - Da)
Person Capacity	Numărul de persoane care pot fi cazate
Multiple Rooms	Exista mai multe camere (0 - Nu, 1 - Da)
Business	Daca se incadreaza la clasa business (0 - Nu, 1 - Da)
Cleanliness Rating	De la 1 la 10, unde 10 este cel mai curat
Guest Satisfaction	De la 1 la 100 unde 100 este cel mai satisfacut
Bedrooms	Numarul de dormitoare
City Center (km)	Distanța fata de centrul orasului exprimata in Km
Metro Distance (km)	Distanța fata de metrou exprimata in Km

**Variabila țintă:** **airbnb\_superhost** – arata daca cazarea este sau nu **superhost**

**Set de antrenare:** airbnb\_dataset.csv

**Set de aplicare:** airbnb\_dataset\_apply.csv

### Weights of Evidence and Information Value

	Variable	Cutoff	N	Events	% of Events	Non-Events	% of Non-Events	WoE	IV
0	Price	(34.778000000000006, 109.224]	4172	1149	0.0985504760271035	3023	0.100582265845949	-0.0204070943907313	4.14629266153083E-05
1	Price	(109.224, 133.254]	4171	1241	0.106441375761215	2930	0.0974879387789053	0.0878657057334653	0.000786700059190703
2	Price	(133.254, 155.567]	4172	1393	0.119478514452354	2779	0.092463816336716	0.256319166577273	0.00692438490633706
3	Price	(155.567, 178.555]	4174	1267	0.108671412642594	2907	0.096722675095658	0.116480902302234	0.00139179973083966
4	Price	(178.555, 203.819]	4170	1244	0.106698687709066	2926	0.0973548494426884	0.0916463142590263	0.000856328338145952
5	Price	(203.819, 233.792]	4174	1245	0.10678445835835	2929	0.0974546664448511	0.0914250846110454	0.000852977015095063
6	Price	(233.792, 272.538]	4177	1162	0.0996654944677931	3015	0.100316087173515	-0.0065065492094587	4.2331134550958E-06
7	Price	(272.538, 330.7]	4163	1080	0.0926323012265203	3083	0.102578605889203	-0.101991485227914	0.00101443838507635
8	Price	(330.7, 452.046]	4172	975	0.0836263830517197	3197	0.106371651971386	-0.240580055522023	0.00547205805955662
9	Price	(452.046, 18545.45]	4169	903	0.077450896303285	3266	0.108667443021128	-0.338648097573454	0.0105714241588104
0	Day	Weekday	20886	5914	0.507247619864482	14972	0.49815338545999	0.0180912539280741	0.0001645261038931
1	Day	Weekend	20828	5745	0.492752380135518	15083	0.50184661454001	-0.0182877479454252	0.00016631306654597

0	Shared Room	False	41398	11624	0.996998027275067	29774	0.99065047413076	0.00638701935308737	4.05419447774345E-05
1	Shared Room	True	316	35	0.00300197272493353	281	0.00934952586923973	-1.13605598530436	0.00721117574162658
0	Person Capacity	2	18126	4684	0.40174972124539	13442	0.447246714357012	-0.107281065272332	0.0048809658877028
1	Person Capacity	3	5292	1513	0.129770992366412	3779	0.12573615039095	0.031585632978838	0.000127443037764561
2	Person Capacity	4	12162	3630	0.311347456900249	8532	0.283879554150724	0.0923594702076263	0.00253692094566073
3	Person Capacity	5	2542	785	0.0673299596877948	1757	0.058459490933289	0.141271252135282	0.00125314222797594
4	Person Capacity	6	3592	1047	0.0898018698001544	2545	0.0846780901680253	0.0587489044258849	0.000301016439907249
0	Multiple Rooms	0	29397	7473	0.64096406209795	21924	0.729462651805024	-0.129334779695746	0.0114459456031486
1	Multiple Rooms	1	12317	4186	0.35903593790205	8131	0.270537348194976	0.283012329338078	0.0250461920161339
0	Business	0	27482	8497	0.728793206964577	18985	0.631675262019631	0.14301458887761	0.0138892829689398
1	Business	1	14232	3162	0.271206793035423	11070	0.368324737980369	-0.306083385393294	0.0297261893711886
0	Cleanliness Rating	2	88	1	8.57706492838151E-05	87	0.00289469306271835	-3.51895749611461	0.0098844785827598
1	Cleanliness Rating	3	6	0	4.28853246419075E-05	6	0.000199634004325403	-1.53795602724803	0.000241072576682403
2	Cleanliness Rating	4	83	4	0.00034308259713526	79	0.00262851439028448	-2.03620286880716	0.00465360277367353
3	Cleanliness Rating	5	54	0	4.28853246419075E-05	54	0.00179670603892863	-3.73518060458425	0.00655083711592186
4	Cleanliness Rating	6	347	3	0.000257311947851445	344	0.0114456829146565	-3.79507874616532	0.0424607488603348
5	Cleanliness Rating	7	619	9	0.000771935843554336	610	0.0202961237730827	-3.26928375729117	0.0638301104723073
6	Cleanliness Rating	8	3031	80	0.00686165194270521	2951	0.0981866577940443	-2.66092211772909	0.243008727971566
7	Cleanliness Rating	9	12444	1618	0.138776910541213	10826	0.360206288471136	-0.953809207198515	0.211201379413801
8	Cleanliness Rating	10	25042	9944	0.852903336478257	15098	0.502345699550824	0.529357692030948	0.185570381607729
0	Guest Satisfaction	(19.999, 84.0]	4371	92	0.00789089973411099	4279	0.142372317418067	-2.89273541682781	0.389019159839592
1	Guest Satisfaction	(84.0, 89.0]	4331	139	0.0119221202504503	4192	0.139477624355349	-2.45950867033022	0.313723868294338
2	Guest Satisfaction	(89.0, 92.0]	5333	391	0.0335363238699717	4942	0.164431874896024	-1.58986720402322	0.208106543728868
3	Guest Satisfaction	(92.0, 93.0]	2785	371	0.0318209108842954	2414	0.0803194144069206	-0.925887716510071	0.0449041686807191
4	Guest Satisfaction	(93.0, 95.0]	5105	1514	0.129856763015696	3591	0.119480951588754	0.0832750623112823	0.000864046343108718
5	Guest Satisfaction	(95.0, 97.0]	6208	3356	0.287846298996483	2852	0.0948926967226751	1.10967990859288	0.214116735733867
6	Guest Satisfaction	(97.0, 98.0]	3305	2243	0.192383566343597	1062	0.0353352187655964	1.6946109552583	0.266135850310893
7	Guest Satisfaction	(98.0, 99.0]	1976	1554	0.133287588987049	422	0.0140409249708867	2.25053283943155	0.268368533361033
8	Guest Satisfaction	(99.0, 100.0]	8300	1999	0.171455527918346	6301	0.209648976875728	-0.201110672901685	0.00768111022025513
0	Bedrooms	0	3745	837	0.0717900334505532	2908	0.0967559474297122	-0.298446145623963	0.00745098079905942
1	Bedrooms	1	28800	8102	0.69491380049747	20698	0.688670770254533	0.00902449007448852	5.6340164462111E-05
2	Bedrooms	2	7830	2393	0.205249163736169	5437	0.18090168025287	0.126270993302424	0.00307438092385061
3	Bedrooms	3	1253	319	0.027360837121537	934	0.0310763600066545	-0.127334712904027	0.000473115039864774
4	Bedrooms	4	65	8	0.000686165194270521	57	0.00189652304109133	-1.01665910361474	0.00123052132360192
5	Bedrooms	5	7	0	4.28853246419075E-05	7	0.000232906338379637	-1.69210670707529	0.000321535831830858
6	Bedrooms	6	2	0	4.28853246419075E-05	2	6.65446681084678E-05	-0.43934373857992	1.0394584410945E-05
7	Bedrooms	9	10	0	4.28853246419075E-05	10	0.000332723340542339	-2.04878165101402	0.000593814808743114
8	Bedrooms	10	2	0	4.28853246419075E-05	2	6.65446681084678E-05	-0.43934373857992	1.0394584410945E-05
0	City Center (km)	(0.013999999999999999, 9, 0.733]	4172	1225	0.105069045372673	2947	0.0980535684578273	0.0691037627811038	0.000484795852519839
1	City Center (km)	(0.733, 1.107]	4171	1199	0.102839008491294	2972	0.0988853768091832	0.0392033717298355	0.000154995692516657
2	City Center (km)	(1.107, 1.457]	4171	1283	0.110043743031135	2888	0.0960905007486275	0.135587487142891	0.00189188505858109
3	City Center (km)	(1.457, 1.85]	4172	1179	0.101123595505618	2993	0.0995840958243221	0.0153410152246427	2.36174880490929E-05
4	City Center (km)	(1.85, 2.253]	4171	1199	0.102839008491294	2972	0.0988853768091832	0.0392033717298355	0.000154995692516657
5	City Center (km)	(2.253, 2.68]	4171	1162	0.0996654944677931	3009	0.10011645316919	-0.0045145166782179	2.03586057864291E-06
6	City Center (km)	(2.68, 3.241]	4172	1243	0.106612917059782	2929	0.0974546664448511	0.0898173672229485	0.000822569958601054

7	City Center (km)	(3.241, 3.954]	4171	1000	0.0857706492838151	3171	0.105506571285976	-0.207096373016239	0.00408723786477886
8	City Center (km)	(3.954, 5.006]	4171	1102	0.0945192555107642	3069	0.102112793212444	-0.077274442366905	0.000586786391489362
9	City Center (km)	(5.006, 25.285]	4172	1067	0.0915172827858307	3105	0.103310597238396	-0.121212120525855	0.00142949265282369
0	Metro Distance (km)	(0.0013, 0.148]	4172	940	0.0806244103267862	3232	0.107536183663284	-0.288025921836966	0.00775128832351225
1	Metro Distance (km)	(0.148, 0.208]	4171	1027	0.0880864568144781	3144	0.104608218266511	-0.171903321080568	0.00284014566370543
2	Metro Distance (km)	(0.208, 0.265]	4171	1113	0.0954627326528862	3058	0.101746797537847	-0.0637514126734917	0.000400618013748153
3	Metro Distance (km)	(0.265, 0.324]	4172	1181	0.101295136804186	2991	0.0995175511562136	0.0177043801073853	3.1471051985129E-05
4	Metro Distance (km)	(0.324, 0.391]	4171	1154	0.0989793292735226	3017	0.100382631841624	-0.0140781695510452	1.97559314851443E-05
5	Metro Distance (km)	(0.391, 0.471]	4171	1216	0.104297109529119	2955	0.0983197471302612	0.059018755225885	0.000352776488314606
6	Metro Distance (km)	(0.471, 0.593]	4172	1338	0.114761128741745	2834	0.0942937947096989	0.196437442977519	0.00402055076182181
7	Metro Distance (km)	(0.593, 0.815]	4171	1320	0.113217257054636	2851	0.0948594243886209	0.176912549207716	0.0032477309748734
8	Metro Distance (km)	(0.815, 1.291]	4171	1224	0.10498327472339	2947	0.0980535684578273	0.0682871028745478	0.000473209564646854
9	Metro Distance (km)	(1.291, 14.274]	4172	1146	0.0982931640792521	3026	0.100682082848112	-0.0240133745336451	5.73660011270826E-05

Woe ne arata relatia dintre variabilele descrise mai sus (coloana 1 din tabel) si variabila **superhost**.  
Cu cat woe este mai mare cu atat creste sansa ca variabila aferenta (Y - superhost) sa fie true.

Pe de alta parte IV ne arata puterea predictiva.

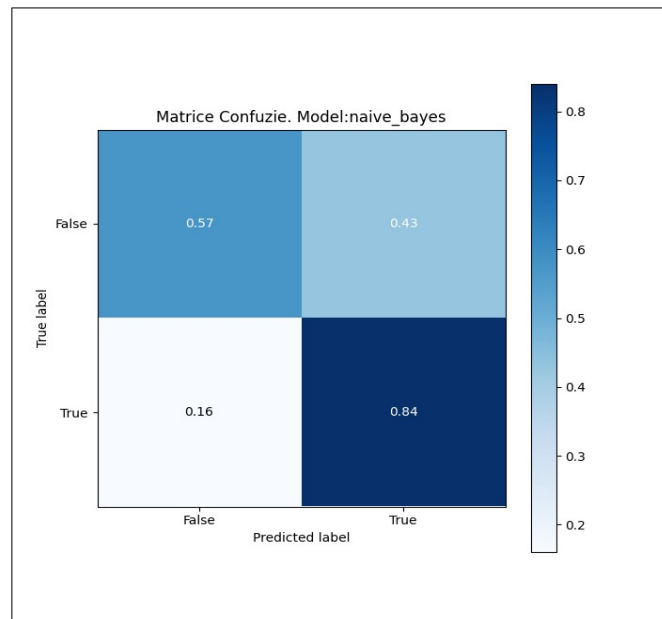
Variabile	IV
Price	0.0279158066931222
Day	0.00033083917043907
Shared Room	0.00725171768640401
Person Capacity	0.00909948853901128
Multiple Rooms	0.0364921376192825
Business	0.0436154723401284
Cleanliness Rating	0.767401339374776
Guest Satisfaction	1.71292001651267
Bedrooms	0.0132214780602347
City Center (km)	0.00963841251245494
Metro Distance (km)	0.0191949127752199

$IV < 0.02 \Rightarrow$  variabila are o putere predictivă foarte mică

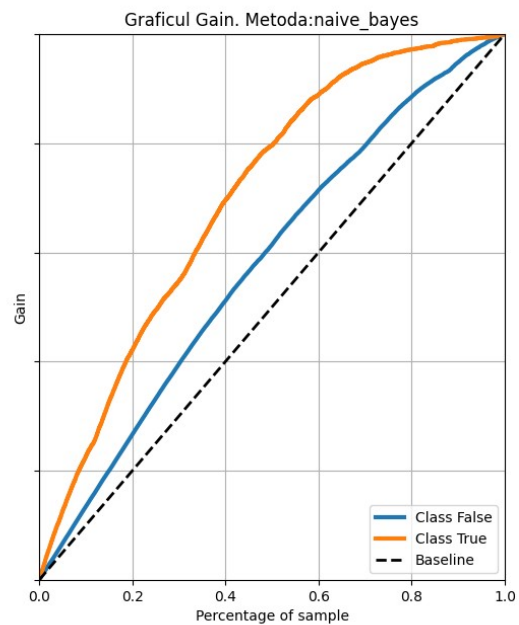
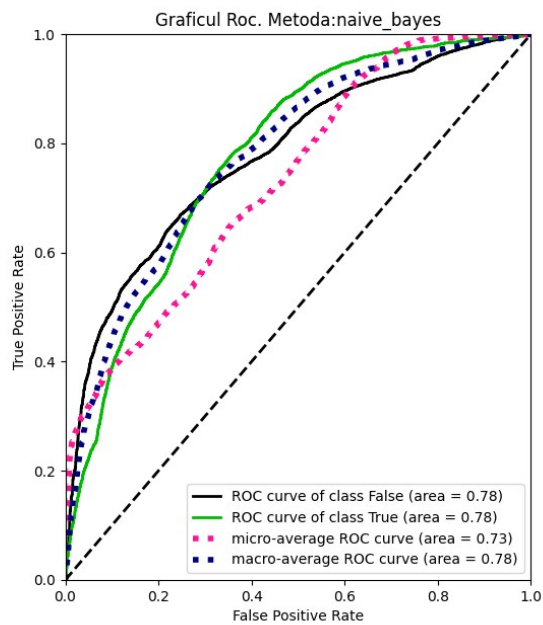
$IV$  intre 0.02 - 0.1  $\Rightarrow$  un predictor slab

$IV > 0.5 \Rightarrow$  un predictor foarte bun (se remarca Cleanliness Rating si Guest Satisfaction)

## Model Bayes

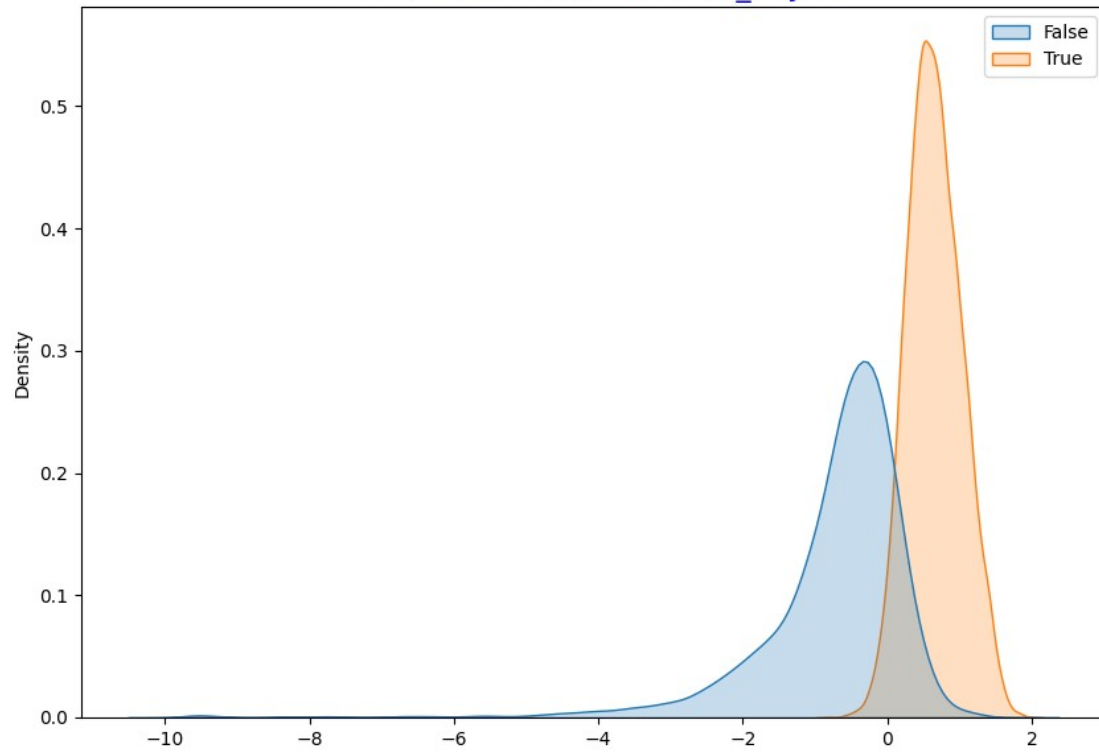


	False	True	Acuratete
False	6849	5100	57.318604067286
True	721	4016	84.7793962423475

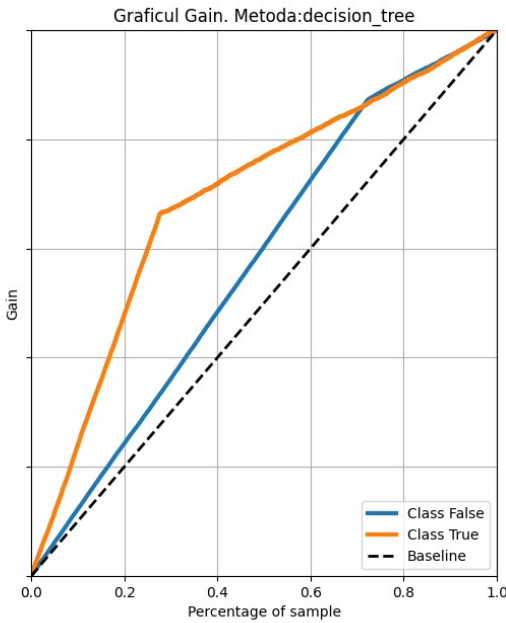
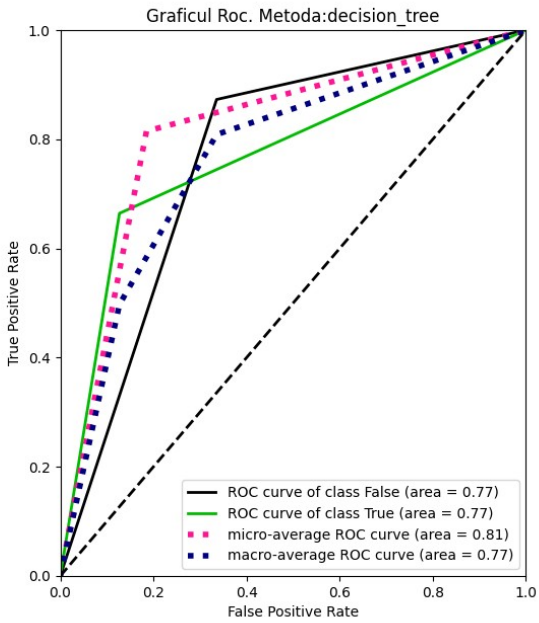
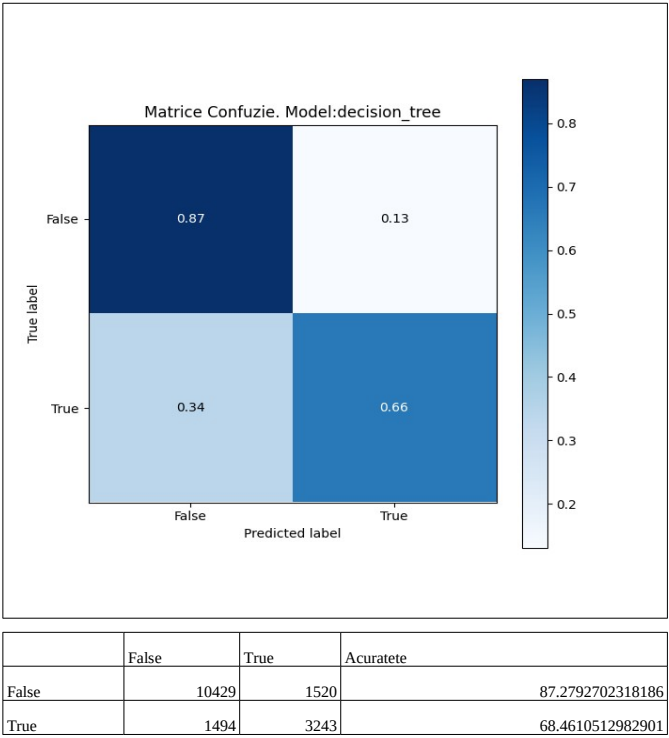


	Acuratete naive_bayes
Acuratete globala	65.1144672180271
Acuratete medie	71.0490001548167
Index Cohen	0.32915413594671

Functii Fisher. Model:naive\_bayes

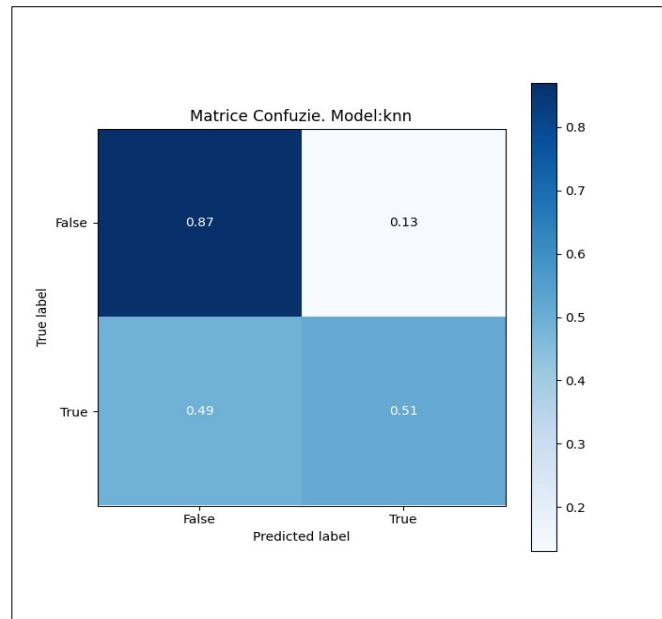


Model Decision Tree

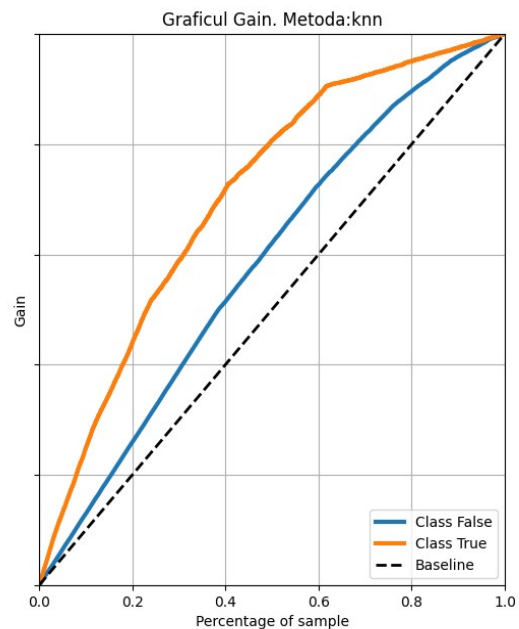
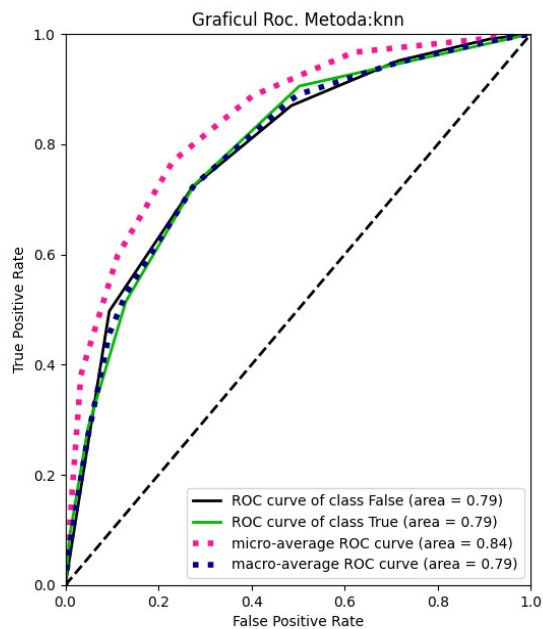


	Acuratete decision_tree
Acuratete globala	81.9369531343641
Acuratete medie	77.8701607650543
Index Cohen	0.556481461736859

## Model KNN

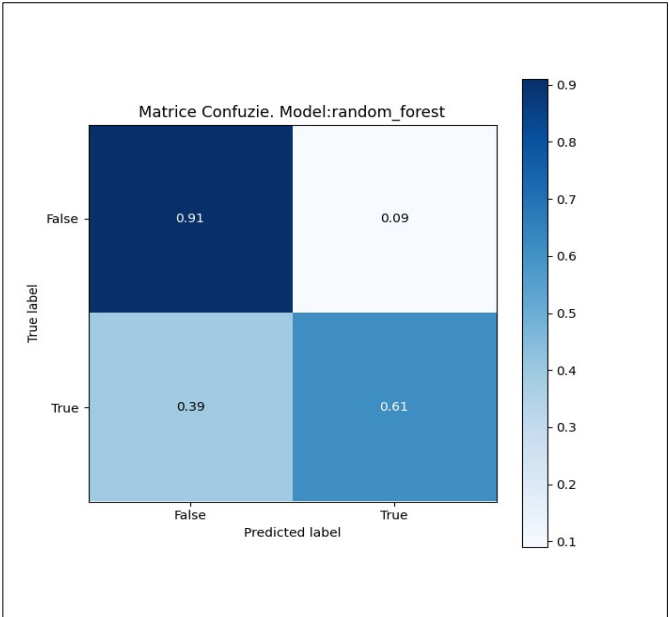


	False	True	Acuratete
False	10394	1555	86.9863586911039
True	2299	2438	51.467173316445

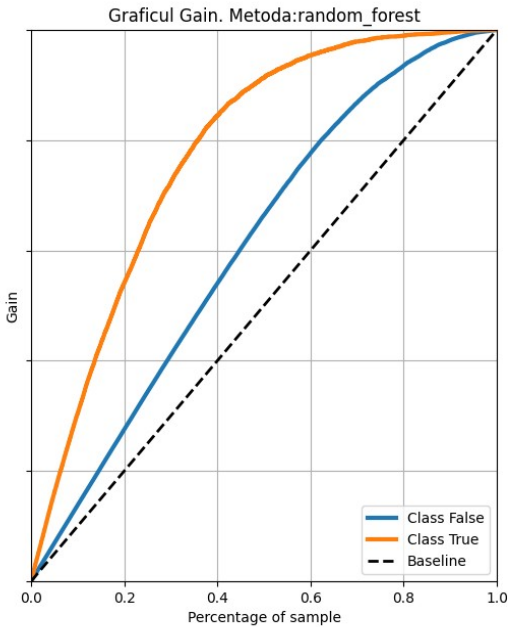
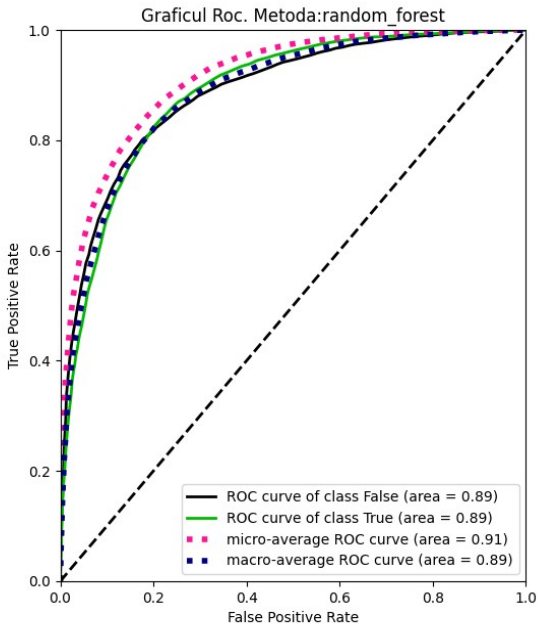


	Acuratete knn
Acuratete globala	76.9027927603979
Acuratete medie	69.2267660037744
Index Cohen	0.403668565551244

Model Random Forest



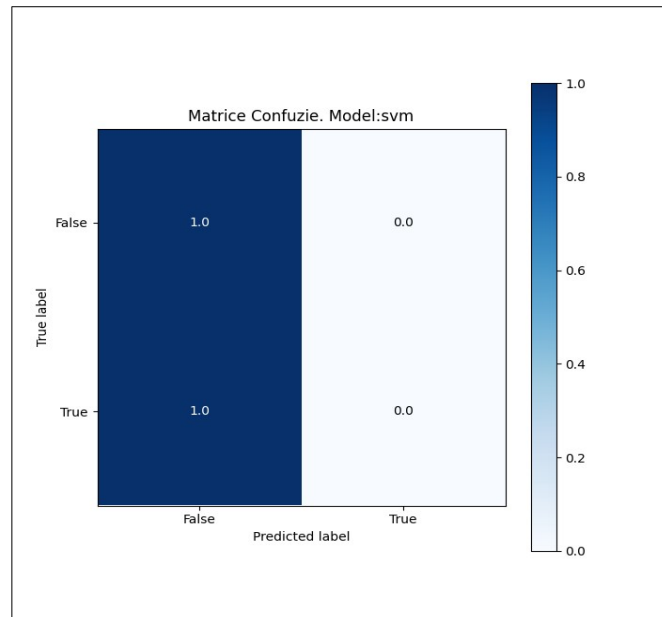
	False	True	Acuratete
False	10915	1034	91.3465561971713
True	1843	2894	61.0935191049187



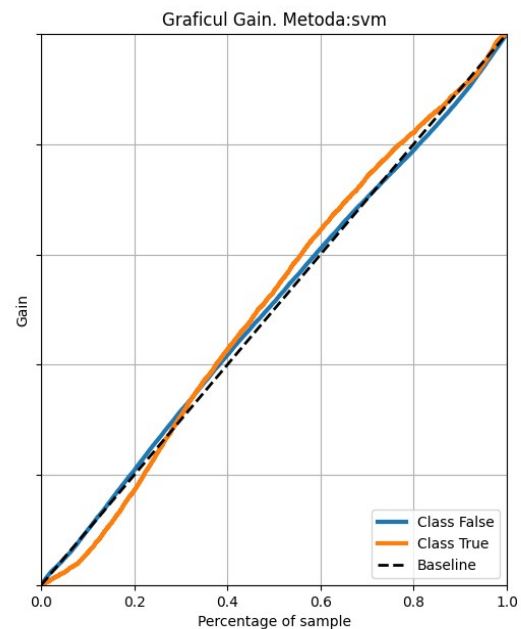
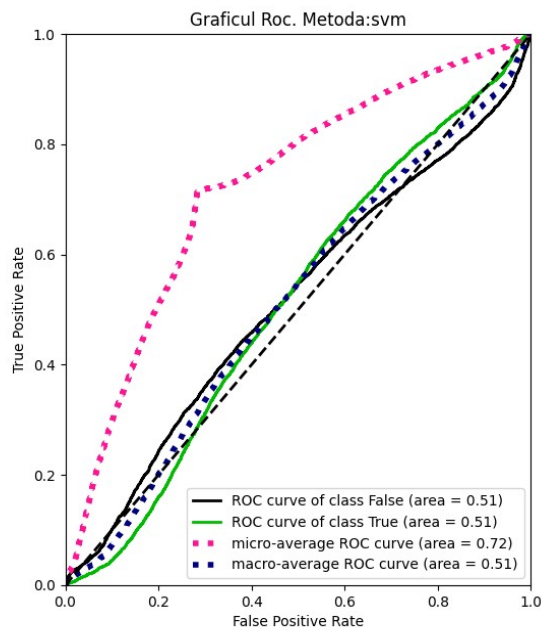
Acuratete random_forest		
Acuratete globala		82.7580007191658
Acuratete medie		76.220037651045
Index Cohen		0.552896713968939



## Model SVM



	False	True	Acuratete
False	11949	0	100
True	4737	0	0



	Acuratete svm
Acuratete globala	71.6109313196692
Acuratete medie	50
Index Cohen	0

Tabel clasificare in setul de testare (primele 120 randuri)

City	Superhost	naive_bayes	decision_tree	knn	random_forest	svm
Athens	True	True	True	False	True	False
Rome	False	False	False	False	False	False
Paris	False	True	False	False	False	False
Budapest	False	True	False	True	False	False
Rome	False	True	True	False	False	False
Paris	True	True	False	False	False	False
Rome	False	False	False	False	False	False
Paris	False	False	False	False	False	False
Paris	False	True	False	True	False	False
Paris	False	True	True	True	False	False
Vienna	True	True	True	True	True	False
Barcelona	False	False	False	False	False	False
Lisbon	False	False	False	False	False	False
Budapest	True	True	True	False	True	False
Paris	False	True	False	True	False	False
Vienna	True	True	True	True	True	False
Paris	False	False	False	False	False	False
Rome	False	True	False	False	False	False
Amsterdam	False	False	False	False	False	False
Lisbon	False	False	True	False	False	False
Barcelona	False	False	False	False	False	False
Athens	False	True	False	True	True	False
Vienna	False	False	False	False	False	False
Athens	False	True	False	False	False	False
Rome	False	False	False	False	False	False
Athens	False	True	False	False	False	False
Paris	False	False	False	False	False	False
Lisbon	False	False	False	False	False	False
Lisbon	False	True	False	False	False	False
Budapest	False	False	False	False	False	False
Athens	True	True	True	True	True	False
Rome	False	True	False	False	False	False
Rome	False	True	False	False	False	False
Athens	True	True	True	True	True	False
Paris	False	False	False	False	False	False
Athens	True	True	False	True	True	False
Paris	False	True	True	False	False	False
Vienna	False	False	False	False	False	False
Athens	False	True	False	True	True	False
Lisbon	False	True	False	True	False	False
Paris	False	True	False	False	False	False
Vienna	False	False	False	False	False	False
Paris	False	True	False	False	False	False
Lisbon	False	True	False	False	False	False
Paris	False	False	False	False	False	False
Rome	True	True	True	False	True	False
Rome	False	False	False	False	False	False
Budapest	True	False	False	False	False	False

Paris	False	False	False	False	False	False
Vienna	False	False	False	False	False	False
Barcelona	False	True	True	True	False	False
Rome	True	True	True	True	True	False
Rome	False	False	False	False	False	False
Rome	True	True	False	True	False	False
Athens	False	True	False	False	False	False
Paris	False	True	False	False	False	False
Rome	False	True	False	True	False	False
Berlin	False	True	False	False	False	False
Barcelona	False	True	False	False	False	False
Rome	False	False	False	False	False	False
Rome	True	True	True	True	True	False
Athens	False	True	False	False	False	False
Lisbon	False	False	False	False	False	False
Paris	False	True	False	False	False	False
Rome	False	True	True	False	True	False
Berlin	True	True	False	False	False	False
Rome	False	True	False	False	False	False
Budapest	False	False	False	False	False	False
Rome	False	False	False	False	False	False
Paris	False	True	False	False	False	False
Lisbon	False	False	False	False	False	False
Athens	False	False	False	False	False	False
Vienna	True	True	False	False	False	False
Athens	True	False	True	False	False	False
Paris	False	False	False	False	False	False
Barcelona	False	True	True	True	False	False
Athens	False	True	False	False	False	False
Budapest	False	True	False	True	False	False
Athens	False	True	False	False	False	False
Budapest	True	True	True	True	False	False
Barcelona	False	False	False	False	False	False
Rome	False	True	False	False	False	False
Rome	True	False	True	False	False	False
Paris	False	False	False	False	False	False
Lisbon	False	True	False	False	False	False
Athens	False	True	False	False	False	False
Vienna	True	True	True	True	True	False
Athens	True	True	True	True	True	False
Amsterdam	False	False	False	False	False	False
Barcelona	False	False	False	False	False	False
Lisbon	False	True	False	False	False	False
Athens	False	False	False	False	False	False
Rome	True	True	True	False	True	False
Vienna	True	True	False	False	True	False
Rome	False	False	False	True	False	False
Athens	False	True	False	False	False	False
Athens	False	True	False	False	False	False
Paris	True	True	True	True	False	False
Paris	False	False	False	False	False	False
Vienna	False	True	False	False	False	False
Budapest	True	True	True	True	True	False

Rome	False	False	False	False	False	False
Vienna	False	False	False	False	False	False
Lisbon	False	False	False	False	False	False
Barcelona	False	False	False	False	False	False
Lisbon	True	True	True	True	True	False
Budapest	False	True	False	False	False	False
Budapest	True	False	True	True	False	False
Amsterdam	False	False	False	False	False	False
Rome	True	False	False	False	False	False
Paris	False	True	False	False	False	False
Rome	False	False	False	False	False	False
Vienna	True	True	True	False	True	False
Lisbon	False	False	False	False	False	False
Rome	True	True	False	False	False	False
Paris	False	False	False	False	False	False
Rome	True	False	True	False	False	False
Athens	False	False	False	False	False	False
Barcelona	False	False	False	False	False	False

Tabel clasificare in setul de aplicare (primele 120 de randuri)

	naive_bayes	decision_tree	knn	random_forest	svm
0	False	True	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	True	False	True	False
5	False	False	False	False	False
6	False	True	False	True	False
7	False	True	False	True	False
8	False	False	False	False	False
9	False	True	True	True	False
10	False	False	False	False	False
11	False	False	False	False	False
12	False	True	True	True	False
13	False	False	False	False	False
14	False	False	True	True	False
15	False	False	False	False	False
16	False	False	False	True	False
17	False	False	False	False	False
18	False	True	False	True	False
19	False	False	False	False	False
20	False	True	False	False	False
21	False	False	False	False	False
22	False	False	False	False	False
23	False	False	False	False	False
24	False	True	True	False	False
25	False	False	False	False	False
26	False	False	False	False	False
27	False	True	False	True	False
28	False	False	False	False	False
29	False	False	False	False	False
30	False	True	False	True	False
31	False	False	False	False	False
32	False	True	True	True	False
33	False	False	False	False	False
34	False	False	False	False	False
35	False	False	False	False	False
36	False	False	False	False	False
37	False	False	False	False	False
38	False	True	True	True	False
39	False	False	False	False	False
40	False	False	False	False	False
41	False	False	True	False	False
42	False	False	True	False	False
43	False	False	False	False	False
44	False	False	False	False	False
45	False	True	True	False	False
46	False	False	False	False	False
47	False	True	False	False	False
48	False	True	False	True	False

49	False	False	False	False	False
50	False	False	False	False	False
51	False	True	False	True	False
52	False	False	False	False	False
53	False	True	False	True	False
54	False	False	False	False	False
55	False	False	False	False	False
56	False	False	False	False	False
57	False	True	False	True	False
58	False	True	False	False	False
59	False	True	False	False	False
60	False	False	False	False	False
61	False	False	False	False	False
62	False	False	False	False	False
63	False	False	False	False	False
64	False	True	True	True	False
65	False	False	False	False	False
66	False	True	True	False	False
67	False	False	False	False	False
68	False	False	False	False	False
69	False	False	False	False	False
70	False	False	False	False	False
71	False	False	False	False	False
72	False	False	False	False	False
73	False	True	False	False	False
74	False	True	False	True	False
75	False	True	False	True	False
76	False	False	False	False	False
77	False	False	True	True	False
78	False	False	False	False	False
79	False	False	False	False	False
80	False	True	True	False	False
81	False	False	False	False	False
82	False	False	False	False	False
83	False	True	False	False	False
84	False	True	False	False	False
85	False	False	False	False	False
86	False	True	True	True	False
87	False	False	True	False	False
88	False	False	False	False	False
89	False	False	False	False	False
90	False	False	False	False	False
91	False	True	True	True	False
92	False	True	False	True	False
93	False	False	False	False	False
94	False	False	False	False	False
95	False	True	False	True	False
96	False	True	True	True	False
97	False	True	False	True	False
98	False	False	True	False	False
99	False	False	False	False	False
100	False	False	False	False	False
101	False	False	False	False	False

102	False	False	False	False	False
103	False	False	False	False	False
104	False	False	False	False	False
105	False	False	False	False	False
106	False	False	False	False	False
107	False	False	False	False	False
108	False	False	True	False	False
109	False	False	False	False	False
110	False	False	False	False	False
111	False	False	False	False	False
112	False	True	False	True	False
113	False	False	False	False	False
114	False	False	False	False	False
115	False	False	False	True	False
116	False	False	False	False	False
117	False	True	False	False	False
118	False	False	True	False	False
119	False	False	False	False	False
120	False	True	True	True	False

Tabel clasificări eronate pentru modelul cel mai eficient

Random forest are cea mai mare acuratete globala

```

naive_bayes decision_tree knn random_forest svm
0          False          True  False          False  False
1          False          False  False          False  False
2          False          False  False          False  False
3          False          False  False          False  False
4          False          True   False          True   False
...          ...          ...   ...          ...   ...
2584        False          False  False          False  False
2585        False          False  False          False  False
2586        False          False  False          False  False
2587        False          False  False          False  False
2588        False          True   False          True   False

[2589 rows x 5 columns]

Previzionate gresit:
1991

Previzionate corect:
598

```