



Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko
Koroška cesta 46
2000 Maribor, Slovenija



Projektna naloga

Pri predmetu MOS

Avtor: Jaša Jernej Rakun Kokalj

Smer študija: ITK

Študijsko leto: 2022/2023

1 Ime izbranega klasifikatorja

KNN – k najbližjih sosedov

1.1 Kratek opis

Algoritem k-najbližjih sosedov, znan tudi kot KNN ali k-NN, je neparametričen klasifikator z nadzorovanim učenjem, ki uporablja bližino za klasifikacije ali napovedi glede združevanja posamezne podatkovne točke.



1.2 Osnovne značilnosti algoritma

KNN deluje po načelu, ki predvideva, da vsaka podatkovna točka, ki spada blizu druga drugi, spada v isti razred. Z drugimi besedami, razvrsti novo podatkovno točko na podlagi podobnosti.

2 Podatkovna zbirka

<https://archive.ics.uci.edu/ml/datasets/car+evaluation>

<https://archive.ics.uci.edu/ml/machine-learning-databases/car/car.data>

2.1 Opis izbrane podatkovne zbirke

Car Evaluation Database je bila izpeljan iz preprostega hierarhičnega odločitvenega modela, prvotno razvitega za predstavitev DEX, M. Bohanec, V. Rajkovic: Ekspertni sistem za odločanje. Sistemica 1(1), str. 145-157, 1990.).

Model ocenjuje avtomobile glede na naslednjo strukturo koncepta:

CAR sprejemljivost .

-Buying- skupna cena

-Maint- cena vzdrževanja

-Door- število vrat

-Persons- zmožljivost oseb v smislu oseb za prenašanje

-lug_boot- velikost prtljažnika

-Safety- ocenjena varnost avtomobila

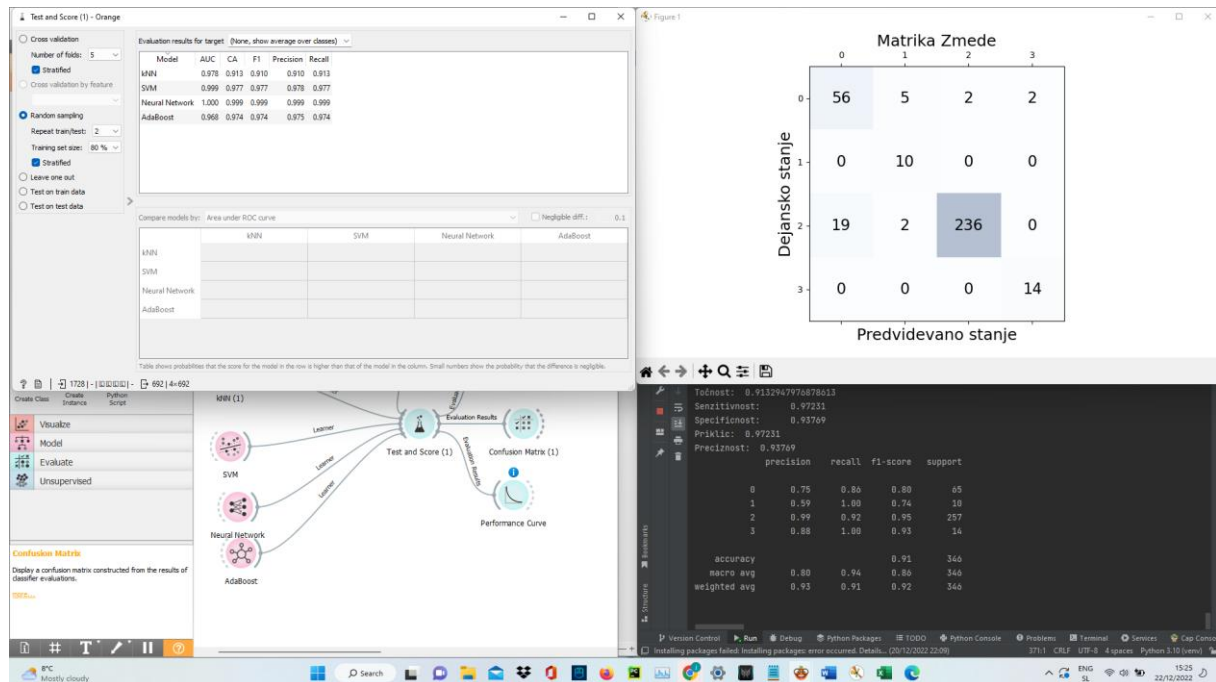
2.2 Vsebina podatkovne zbirke

Podatkovna zbirka vsebuje opisne informacije o 1729 avtomobilih. Avtomobili so razen atributov door(vrata) in persons (max. Število potnikov) opisani z besedami in ne številkami. Podatkovna zbirka je v obliki PDF dokumenta priložena poročilu.

2.3 Namen podatkovne zbirke

Namen podatkovne zbirke je oceniti ustreznost avtomobila glede na podane parametre. Ustreznost se ocenjuje opisno kot: nesprejemljivo, sprejemljivo, dobro, odlično.

3 Predstavitev rezultatov



Z uporabo orodja Orange sem nad isto podatkovno zbirko z uporabo kNN metode dobil 91.3% točnost kar je tudi sovpadalo z mojo implementacijo kNN metode kjer je dobljena točnost klasifikacije znašala 91.329%.

Senzitivnost: 0.97231

Specifičnost: 0.93769

Priklic: 0.97231

Preciznost: 0.93769

precision recall f1-score support

0 0.75 0.86 0.80 65

1 0.59 1.00 0.74 10

2 0.99 0.92 0.95 257

3 0.88 1.00 0.93 14

accuracy 0.91 346

macro avg 0.80 0.94 0.86 346

weighted avg 0.93 0.91 0.92 346

Z uporabo orodja Orange so bili rezultati:

kNN

Area under control: 0.9784124466805205 oz. 97.841%

Točnost: 0.9132947976878613 oz. 91.329%

F1: 0.9096419229468806 oz. 90.964%

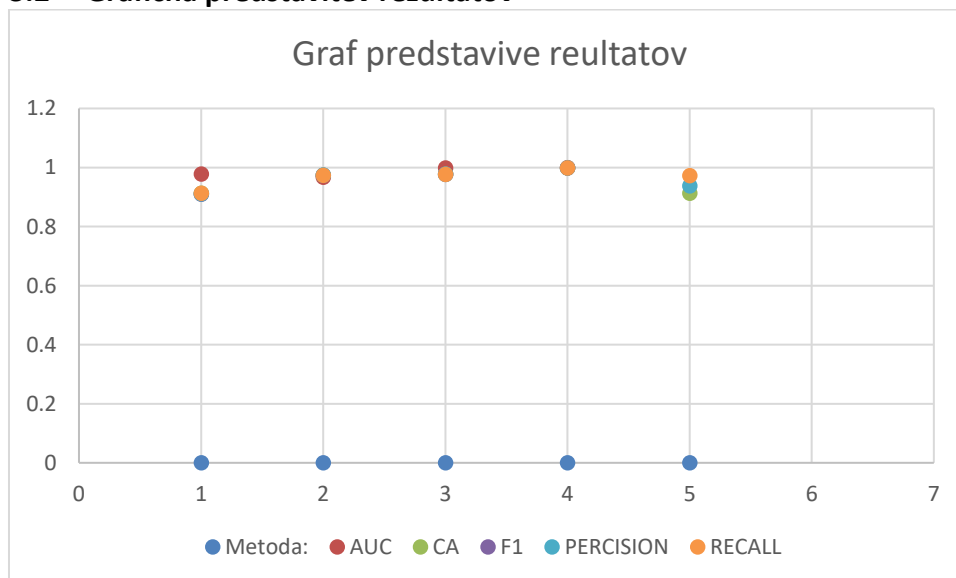
Precision: 0.9104036530692784 oz. 91.040%

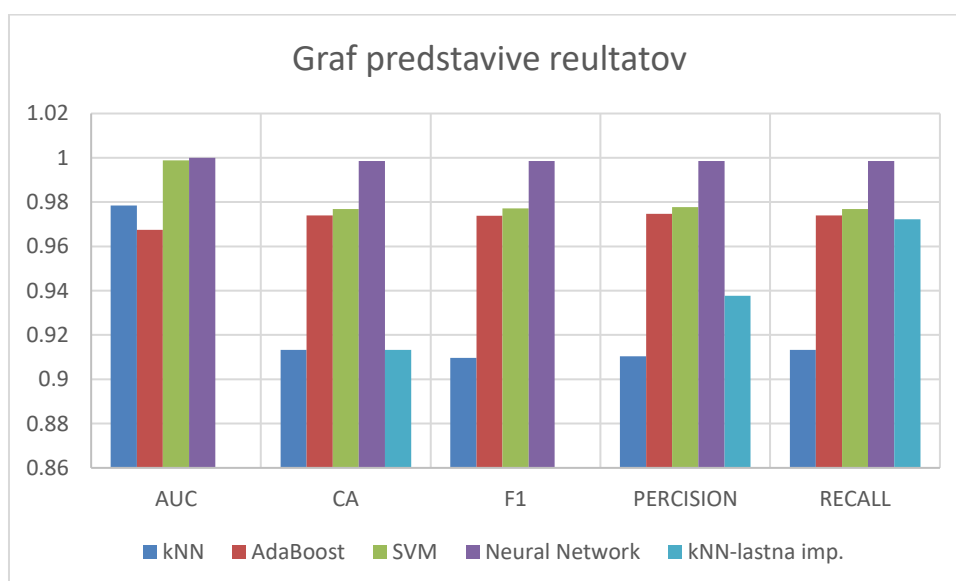
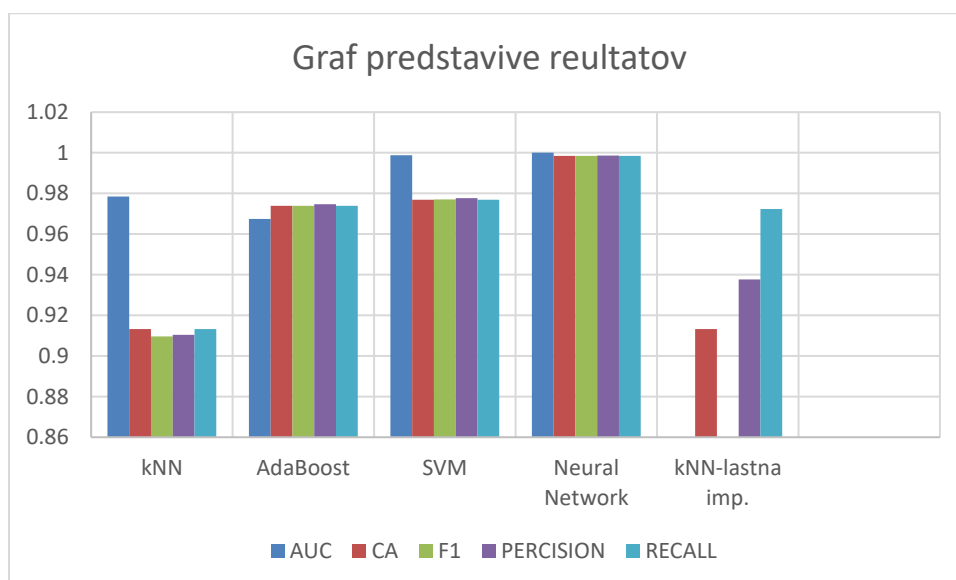
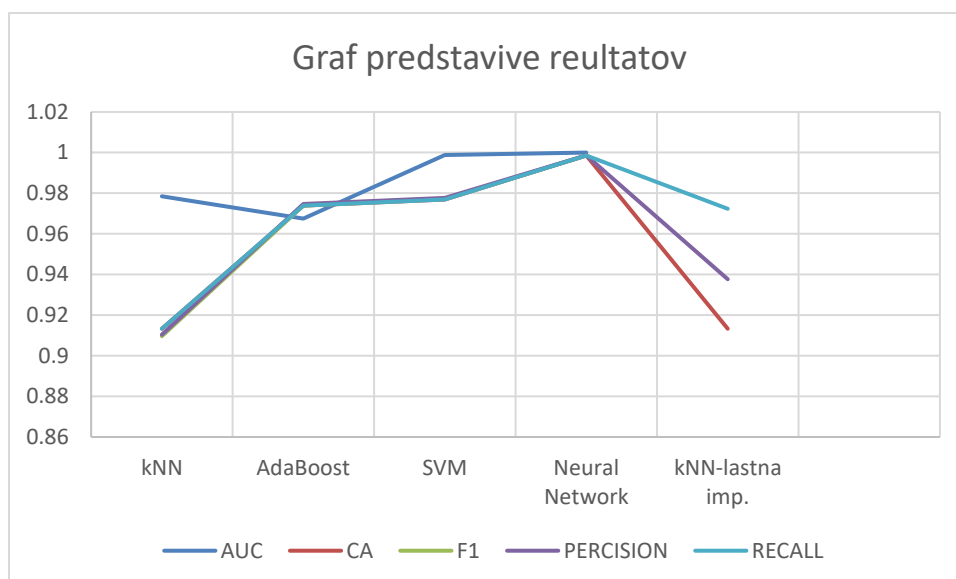
Recall: 0.9132947976878613 oz. 91.329%

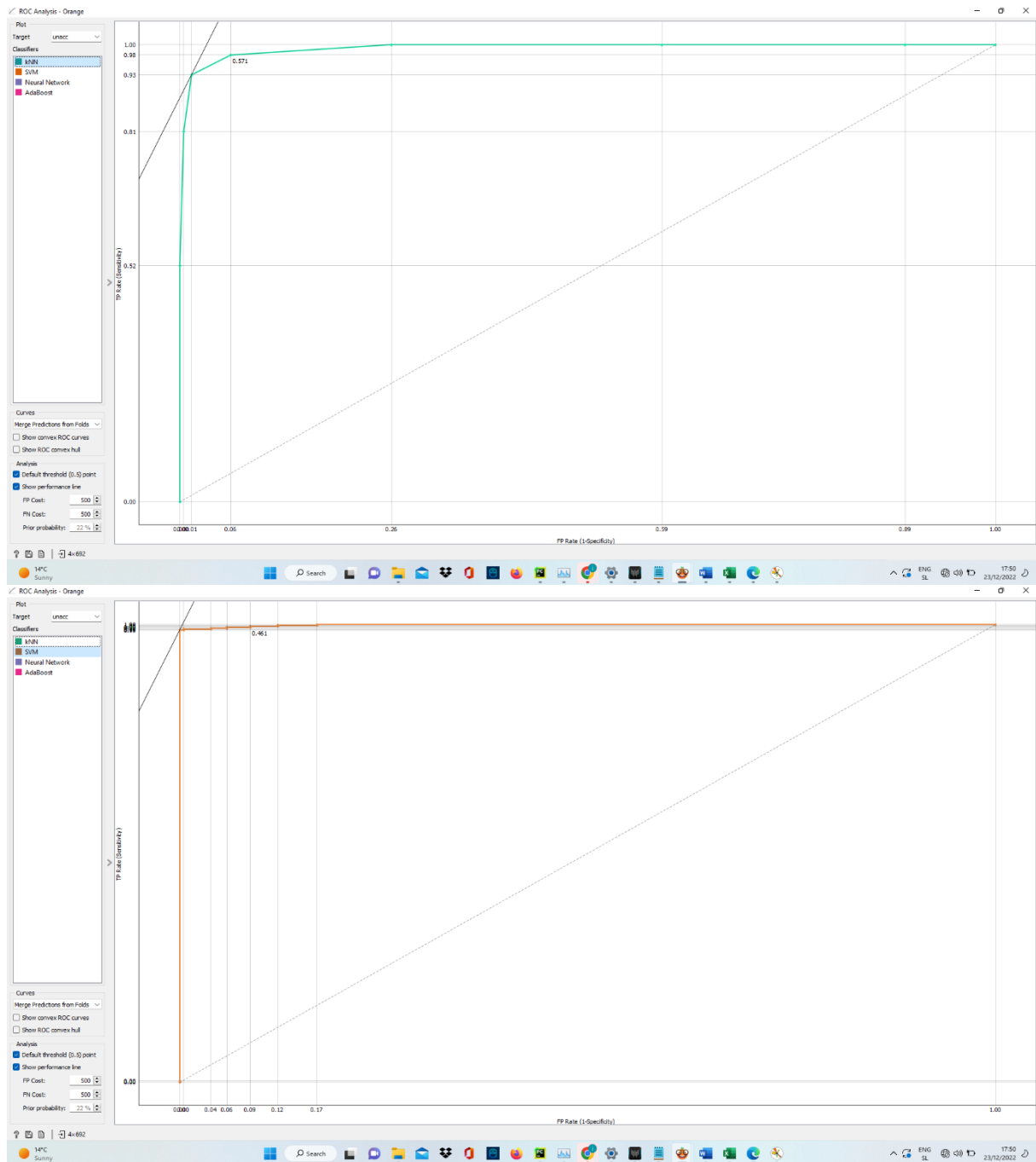
3.1 Tabelarična predstavitev rezultatov

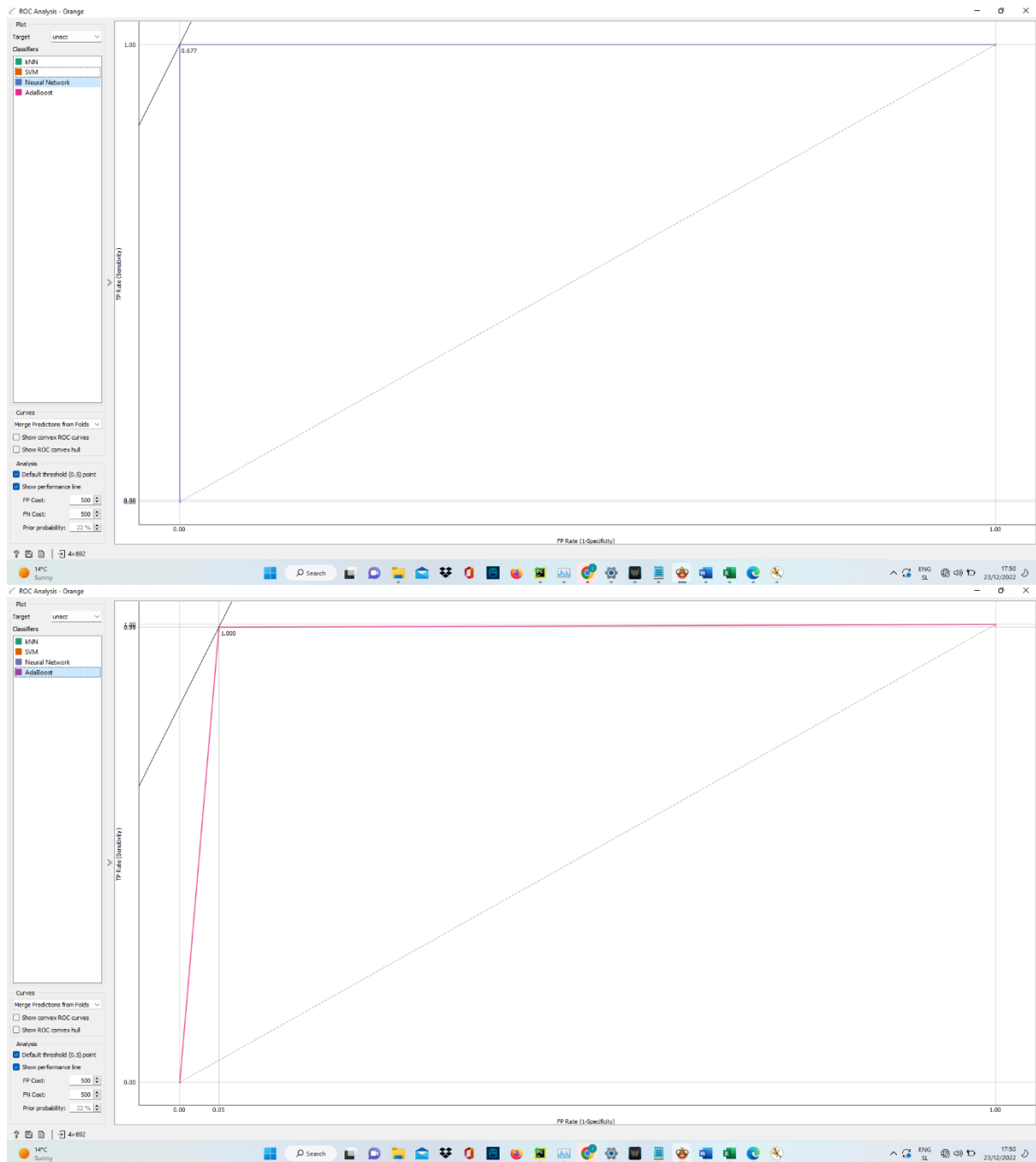
Metoda:	AUC	CA	F1	PERCISION	RECALL
kNN	0.978412447	0.913294798	0.909641923	0.910403653	0.913295
AdaBoost	0.967502461	0.973988439	0.973889959	0.974764157	0.973988
SVM	0.998815123	0.976878613	0.977141037	0.977718771	0.976879
Neural Network	1	0.998554913	0.998565236	0.998604744	0.998555
kNN-lastna imp.		0.913294798		0.93769	0.97231

3.2 Grafična predstavitev rezultatov









Confusion Matrix (1) - Orange

Learners				
kNN				
SVM				
Neural Network				
AdaBoost				

		Predicted				Σ
		acc	good	unacc	vgood	
Actual	acc	132	3	17	2	154
	good	15	12	0	1	28
	unacc	11	0	473	0	484
	vgood	8	3	0	15	26
Σ		166	18	490	18	692

Confusion Matrix (1) - Orange

Learners				
kNN				
SVM				
Neural Network				
AdaBoost				

		Predicted				Σ
		acc	good	unacc	vgood	
Actual	acc	146	3	2	3	154
	good	0	27	0	1	28
	unacc	4	1	479	0	484
	vgood	2	0	0	24	26
	Σ	152	31	481	28	692

Confusion Matrix (1) - Orange

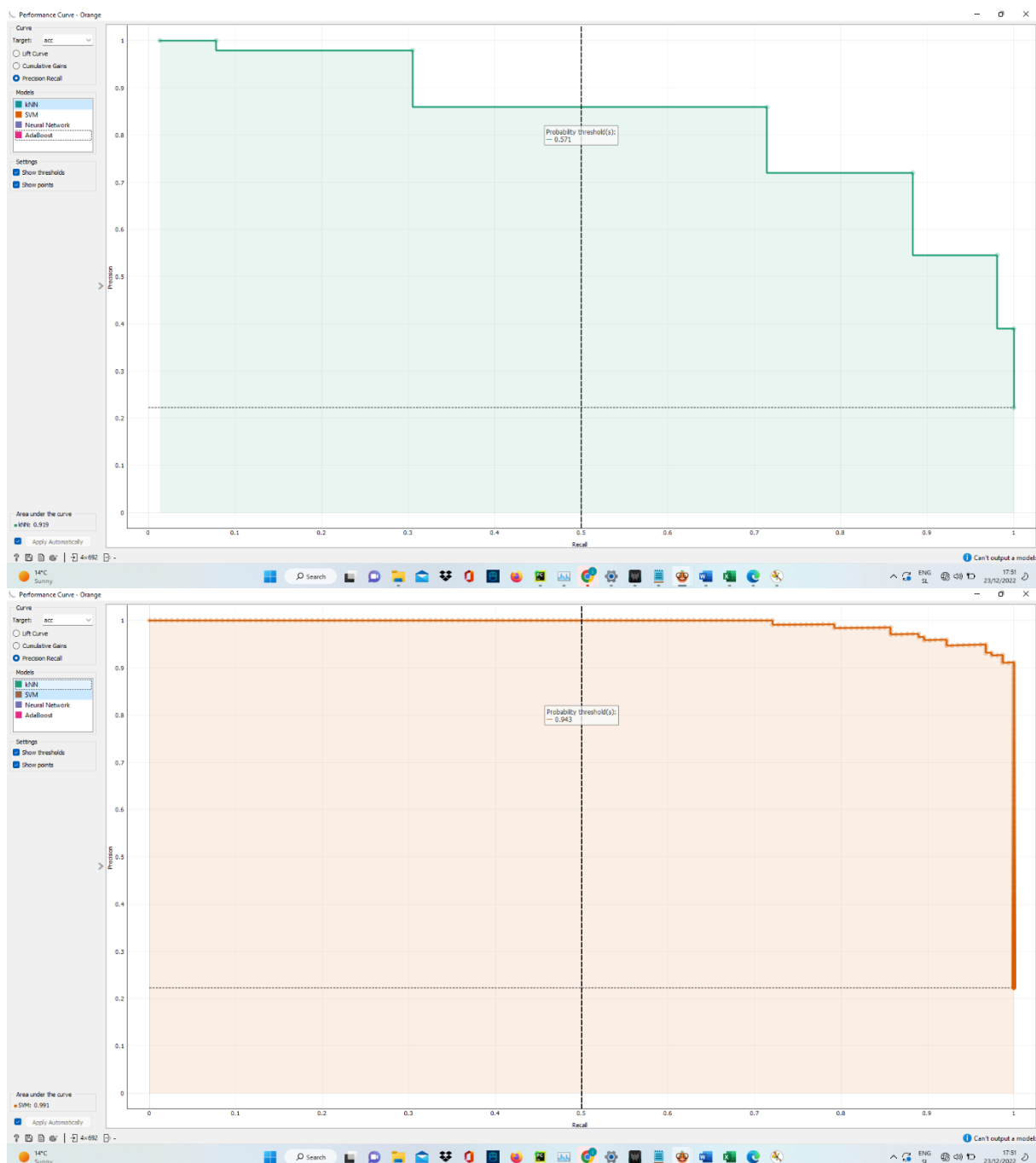
Learners				
kNN				
SVM				
Neural Network				
AdaBoost				

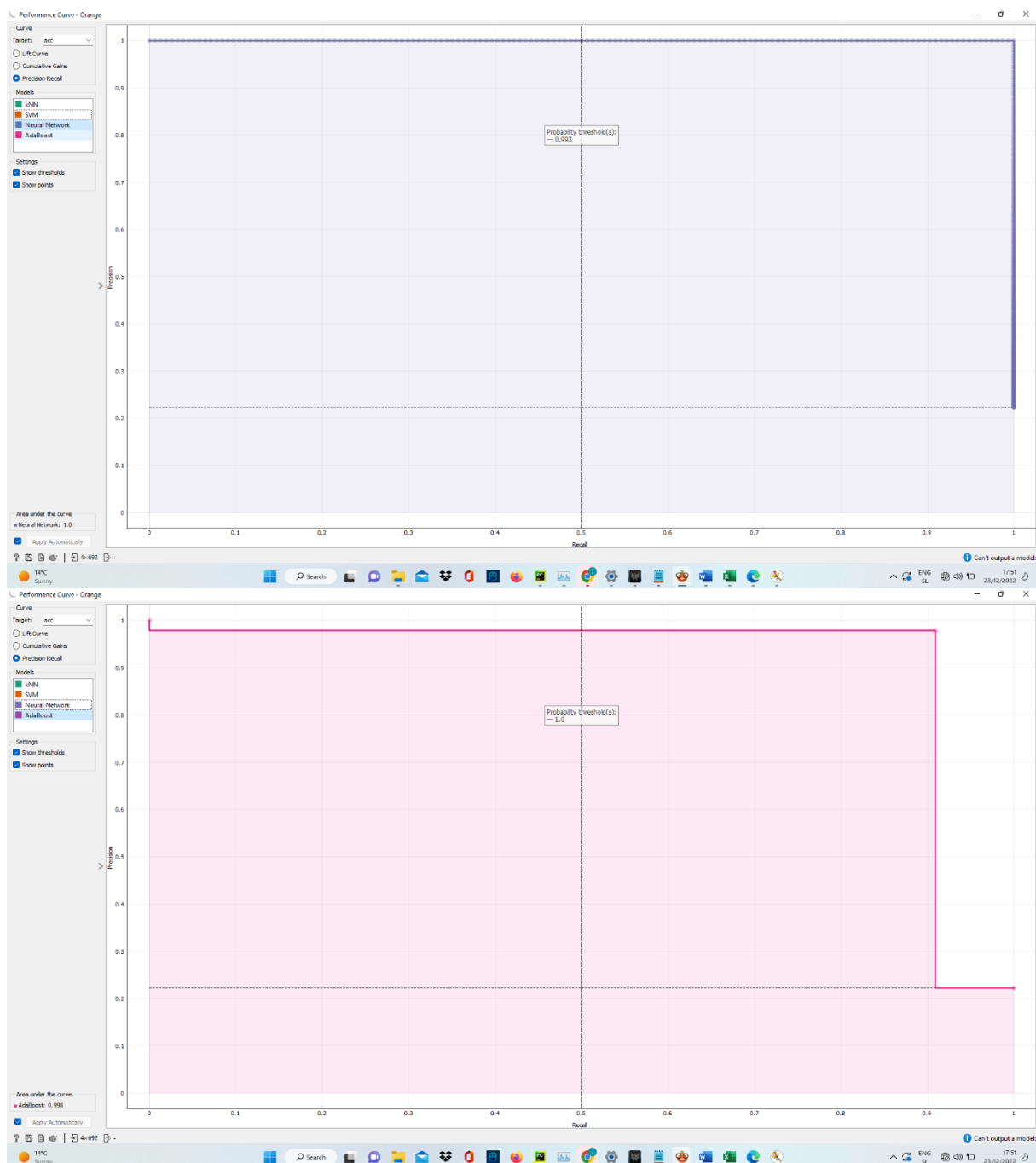
		Predicted				Σ
		acc	good	unacc	vgood	
Actual	acc	153	1	0	0	154
	good	0	28	0	0	28
	unacc	0	0	484	0	484
	vgood	0	0	0	26	26
	Σ	153	29	484	26	692

Confusion Matrix (1) - Orange

Learners				
kNN				
SVM				
Neural Network				
AdaBoost				

		Predicted				Σ
		acc	good	unacc	vgood	
Actual	acc	140	4	10	0	154
	good	1	27	0	0	28
	unacc	2	1	481	0	484
	vgood	0	0	0	26	26
	Σ	143	32	491	26	692





3.3 Interpretacija rezultatov

Točnost

Točnost = (true positives + true negatives) / (true positives + true negatives + false positives + false negatives)

Moja implementacija kNN v primerjavi z kNN algoritmom orodja Orange deluje podobno dobro (razlike so minimalne- za približno 0.03%) torej se rezultati na testni množici razlikujejo v približno enem do dveh primerih, vse ostale enako oceni.

Preciznost

$\text{Preciznost} = \text{True positives} / (\text{True positives} + \text{False positives})$

Preciznost je delež ustreznih primerkov med pridobljenimi primerki, medtem ko je priklic (znan tudi kot občutljivost) delež ustreznih primerkov, ki so bili pridobljeni.

Moja implementacija kNN v primerjavi z kNN algoritmom orodja Orange deluje glede preciznosti bolje (za približno 2-3% bolje).

Priklic (Recall)

$\text{Priklic} = \text{true positives} / (\text{true positives} + \text{false negatives})$

Priklic je delež ustreznih primerkov, ki so bili pridobljeni. Tako natančnost kot priklic torej temeljita na ustreznosti.

Moja implementacija kNN v primerjavi z kNN algoritmom orodja Orange deluje glede priklica bolje (za približno 6% bolje).

Senzitivnost (Sensitivity)

V kontekstu ocenjevanja modela strojnega učenja je občutljivost (znana tudi kot resnična pozitivna stopnja ali priklic) merilo deleža pozitivnih primerov, ki jih je model pravilno identificiral. Izračuna se kot število resničnih pozitivnih napovedi modela, deljeno s skupnim številom pozitivnih primerov v naboru podatkov.

$\text{Občutljivost} = \text{resnično pozitivni} / (\text{resnični pozitivni} + \text{lažni negativni})$

Senzitivnost pri moji implementaciji kNN je znašala: 0.9406528189910979

Senzitivnost kNN pri orodju orange, pa je znašala: 0.9446935724962631

Glede senzitivnosti deluje moja implementacija kNN v primerjavi z kNN orodja Orange slabše (za približno 0.4% slabše) a se še vedno vrednosti (94,47%) kNN orodja Orange zadovoljivo približa.

4 Zaključek

Moja implementacija algoritma kNN deluje dobro (93.8% točnost, 97.2% Priklic, 93.7% Preciznost). V primerjavi z kNN orodja Orange deluje celo bolje.

V primerjavi z ostalimi naprednejšimi algoritmi orodja Orange deluje moja implementacija kNN slabše, zato bi bilo dobro na realnem primeru dobro uporabiti druge npr. Neural network (ki ima točnost klasifikacije 99.9%, preciznost 99.9%, priklic 99.9%) ali SVM (ki ima točnost klasifikacije 97.8%, preciznost 97.8%, priklic 97.7%).

Rezultati so sprejemljivi, a bi jih lahko izboljšali z povečanjem učne množice.