

# Razpoznavanje Vzorcev

Poročilo tretje laboratorijske vaje

Avtor: Bor Starčič

Mentorja: izr. prof. dr. Simon Dobrišek, asist. dr. Klemen Grm

Datum: December 17, 2023

# Contents

<b>1</b>	<b>Naloge</b>	<b>3</b>
<b>2</b>	<b>Navkrižno preverjanje</b>	<b>3</b>
<b>3</b>	<b>Testi statistične analize</b>	<b>4</b>
3.1	Test Kolmogorova in Smirnova . . . . .	4
3.2	Test Andersona in Darlinga . . . . .	5
3.3	Test vsote rangov . . . . .	5
<b>4</b>	<b>Povzetek</b>	<b>6</b>

# 1 Naloge

1. Navkrižno preverjanje
2. Testi statistične analize

## 2 Navkrižno preverjanje

Značilke, ki sem jih uporabil: [4, 8, 12, 20, 21, 28, 29, 36].

Natančnosti za posamezne delitve so se gibale od 49.86% do 58.26%, kar kaže na določeno variabilnost v uspešnosti modela pri različnih delitvah podatkov.

Povprečna natančnost klasifikacije je 53.96%, standardni odklon natančnosti je 2.34% in mediana natančnosti je 54.20%. Te vrednosti so v mejah raztrosa v primerjavi z rezultati iz 2. laboratorijske vaje.

```
Natančnost klasifikacije 1: 51.40%
Natančnost klasifikacije 2: 53.07%
Natančnost klasifikacije 3: 55.18%
Natančnost klasifikacije 4: 54.06%
Natančnost klasifikacije 5: 49.86%
Natančnost klasifikacije 6: 53.78%
Natančnost klasifikacije 7: 54.90%
Natančnost klasifikacije 8: 58.26%
Natančnost klasifikacije 9: 54.34%
Natančnost klasifikacije 10: 56.02%
Natančnost klasifikacije 11: 56.58%
Natančnost klasifikacije 12: 50.42%
Natančnost klasifikacije 13: 51.82%
Natančnost klasifikacije 14: 55.74%
Povprečna natančnost klasifikacije: 53.96%
Standardni odklon natančnosti: 2.34%
Mediana natančnosti: 54.20%
```

Figure 1: Rezultati.

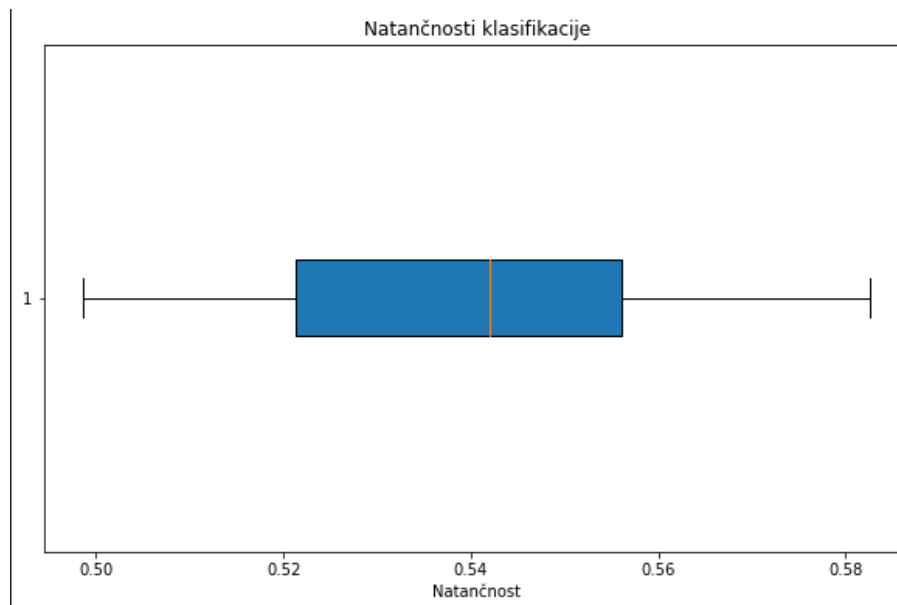


Figure 2: Raztros.

### 3 Testi statistične analize

Naslednji testi so uporabljeni za oceno ujemanja vzorca z določeno distribucijo, preverjanje če vzorci izhajajo iz iste distribucije ter primerjava median vzorcev.

#### 3.1 Test Kolmogorova in Smirnova

KS Statistika: 0.3333333333333333 - vrednost pomeni, da je največja razlika med kumulativnimi distribucijami dveh primerjanih vzorcev približno 33.33 %.

P-vrednost: 0.989010989010989 - vrednost je verjetnost, da bi opazovali tako veliko ali večjo razliko med vzorci, če bi bili ti dejansko vzeti iz iste populacije. Visoka P-vrednost kaže, da obstaja velika verjetnost, da bi opazili razlike med vzorci.

Iz teh dveh vrednosti ne dobimo dovolj dokazov, da bi trdili, da se vzorca statistično značilno razlikujeta

```
✓ ### Test Kolmogorova in Smirnova za primer dveh vzorcev ### ...
KS statistika: 0.3333333333333333
P-vrednost: 0.989010989010989
Ni dovolj dokazov, da bi trdili, da se vzorca statistično značilno razlikujeta.
```

Figure 3: Rezultati.

### 3.2 Test Andersona in Darlinga

Anderson-Darlingov test primerja izračunano statistiko z vrednostmi kritičnih točk, ki so odvisne od stopnje značilnosti. Stopnja značilnosti je prag, da vzorec sledi določeni distribuciji.

Statistika: 0.229 - vrednost je merilo razlike med empirično distribucijo vzorca in normalno distribucijo.

Pri 15%, 10%, 5%, 2.5% in 1.0% stopnjah značilnosti rezultati kažejo, da vzorec sledi normalni porazdelitvi. To pomeni, da pri teh stopnjah značilnosti izračunana statistika ni večja od kritičnih vrednosti, kar kaže na sprejemljivo ujemanje z normalno distribucijo.

```
✓ ### Test Andersona in Darlinga za primer k vzorcev ### ...  
  
Statistika: 0.229  
Pri stopnji značilnosti 15.0% vzorec sledi normalni porazdelitvi.  
Pri stopnji značilnosti 10.0% vzorec sledi normalni porazdelitvi.  
Pri stopnji značilnosti 5.0% vzorec sledi normalni porazdelitvi.  
Pri stopnji značilnosti 2.5% vzorec sledi normalni porazdelitvi.  
Pri stopnji značilnosti 1.0% vzorec sledi normalni porazdelitvi.
```

Figure 4: Rezultati.

### 3.3 Test vsote rangov

Mann-Whitney U statistika: 12.5 - vrednost testne statistike, ki se izračuna na podlagi rangov podatkovnih točk v dveh skupinah. Mann-Whitney U test primerja mediane dveh skupin.

P-vrednost: 0.1 - vrednost je verjetnost, da bi opazovali razliko med rangi dveh skupin. Nizka P-vrednost, kaže, da ni velike verjetnost, da bi opazili razlike med skupinami.

Na podlagi teh dveh podatkov ni statistično značilne razlike med skupinama.

```
✓ ### Test vsote rangov (Mann-Whitney U-test) ### ...  
Mann-Whitney U statistika: 12.5  
P-vrednost: 1.0  
Ni statistično značilne razlike med skupinama.
```

Figure 5: Rezultati.

## 4 Povzetek

Test Andersona-Darlinga, in test vsote rangom (Mann-Whitney U-test) sta primerna za primerjavo testne in učne zbirke podatkov, ker Anderson-Darling primerja komulativne porazdelitve dveh podatkovnih zbirk, Mann-Whitney U-test pa primerja mediani dveh podatkovnih zbirk. Test vsote rangov (Mann-Whitney U-test) se uporablja za ugotovitev če vzorec prihaja iz specifične distribucije in je zato primeren za uporabo tukaj medtem ko pa Pearsonov Test ni primeren.