

Klasifikacija pečuraka

Tatjana Radovanović

Matematički fakultet, Univerzitet u Beogradu

August 27, 2019

1 Podaci

2 Klasifikacija

Osnovno o podacima

- <https://archive.ics.uci.edu/ml/datasets/Mushroom>
- 8416 redova sa 23 atributa
- Svi atributi su kategorički
- Fizičke karakteristike pečuraka

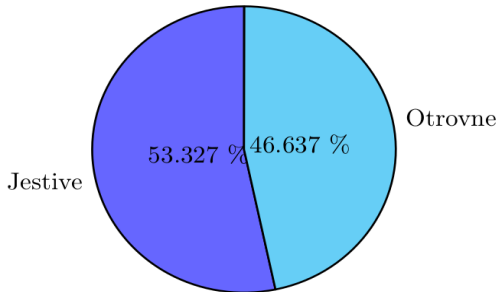


Figure: Odnos otrovnih i jestivih pečuraka

Pretprocesiranje podataka

- Vrednosti atributa 'Edible' postavljeni na 0 (POISONOUS) i 1 (EDIBLE)
- Vrednosti atributa 'Bruises' postavljeni na 0 (NO) i 1 (Bruises)
- Nedostajuće vrednosti označene sa '?' i sve se odnose na atribut 'Stalk-root'
- Isključen atribut 'Stalk-root' jer ima približno 29.47% nedostajućih vrednosti
- Atribut 'Veil-type' ima samo jednu vrednost na celom skupu pa je i on uklonjen
- Binarno kodiranje

- StandardScaler
- Srednja vrednost 0, standardna devijacija 1

$$z = \frac{x - u}{s}$$

- u - srednja vrednost, s - standardna devijacija
- MinMaxScaler
- Opseg (0, 1)

- Ciljna promenljiva y je atribut 'Edible', a x su ostali atributi
- Podela na trening i test skup
- Stratifikacija po y
- Unakrsna validacija

- Drvo odlučivanja
- k najbližih suseda
- Naivni Bajesov klasifikator
- Veštačke neuronske mreže
- Metod potpornih vektora

Drvo odlučivanja

- Kriterijum podele entropija
- Maksimalna dubina 4
- Maksimalan broj listova 8
- Preciznost za klasu 0 je približno 1, a za klasu 1 približno 0.997
- Odziv za klasu 0 je 0.9966, a za klasu 1 približno 1

Table: Matrica konfuzije za drvo odlučivanja

	0	1
0	1174	4
1	0	1347

K najbližih suseda

- Menhetn rastojanje
- Težina suseda 'uniform'
- Broj suseda 1
- Preciznost i odziv za obe klase je 1

Table: Matrica konfuzije za k najbliih suseda

	0	1
0	1178	0
1	0	1347

Naivni Bajesov klasifikator

- Nije upotrebljena unakrsna validacija
- Preciznost za klasu 0 približno 0.9937, a za klasu 1 približno 0.9477
- Odziv za klasu 0 je približno 0.9372, a za klasu 1 približno 0.9948

Table: Matrica konfuzije za Naivni Bajesov klasifikator

	0	1
0	1104	74
1	7	1340

- Rešavač za optimizaciju težina stohističkog opadajućeg gradijenta
- Maksimalan broj iteracija je 500
- Indetička aktivaciona funkcija
- Stopa učenja pri ažuriranju je 'adaptive'
- Inicijalna stopa učenja je 1
- Preciznost za klasu 0 približno 1, a za klasu 1 približno 0.9993
- Odziv za klasu 0 je približno 0.9992, a za klasu 1 približno 1

Table: Matrica konfuzije za neuronsku mrežu

	0	1
0	1177	1
1	0	1347

Metod potpornih vektora

- Parametar regularizacije 0.1
- Funkcija kernela 'linear'
- Preciznost i odziv za obe klase je 1

Table: Matrica konfuzije za potporne vektore

	0	1
0	1178	0
1	0	1347

Hvala na pažnji!