

Klasifikacija skupa podataka **Online News Popularity**

Lea Petković 163/2016

Uvod u podatke

Zadatak:	Karakteristike atributa:	Broj instanci:	Broj atributa:
Klasifikacija	Celi i realni brojevi	39797	61

Null vrednosti?

Ekstremne vrednosti i elementi van granice?

Ciljni atribut?

Tipovi ostalih atributa?

Analiza i
pretprocesiranje
podataka

Analiza i preprocesiranje podataka (nastavak)

Podela skupa podataka

Redukcija broja atributa

Primena algoritama

C5.0 algoritam

Nad svim atributima

Podela	Test s.		Trening s.	
Ispravno	6 867	57.77%	16 020	57.71%
Pogrešno	5 019	42.23%	11 738	42.29%

Podela	Test s.		Trening s.	
Ispravno	8 588	72.75%	20 259	72.77%
Pogrešno	3 216	27.25%	7 581	27.23%

Nad redukovanim brojem
atributa

CART algoritam

Nad svim atributima

Podela	Test s.		Trening s.	
Ispravno	4 819	40.59%	11 147	40.14%
Pogrešno	7 054	59.41%	16 624	59.86%

Podela	Test s.		Trening s.	
Ispravno	4 767	39.59%	11 053	40.04%
Pogrešno	7 274	60.41%	16 550	59.96%

Nad redukovanim brojem
atributa

K najbližih suseda (KNN)

Nad svim atributima

Podela	Test s.		Trening s.	
Ispravno	5 622	46.77%	13 077	47.34%
Pogrešno	6 399	53.23%	14 546	52.66%

Podela	Test s.		Trening s.	
Ispravno	5 534	46.19%	13 008	47.02%
Pogrešno	6 448	53.81%	14 054	52.98%

Nad redukovanim brojem
atributa

Metod potpornih vektora (SVM)

Nad svim atributima

Podela	Test s.		Trening s.	
Ispravno	5 589	47.22%	13 067	46.99%
Pogrešno	6 246	52.78%	14 742	53.01%

Podela	Test s.		Trening s.	
Ispravno	4 784	40.17%	10 938	39.44%
Pogrešno	7 126	59.83%	16 796	60.56%

Nad redukovanim brojem
atributa

Algoritam slučajne šume (Random Decision Forest)

Nad svim atributima:

Preciznost nad trening podacima:	0,856
Preciznost nad test podacima:	0,346



Problemi?

Da li je
PCA
algoritam
pomogao?

Šta jeste?

Koji
algoritam
je
najbolji?

Zaključak

HVALA NA
PAŽNJI!