

Nekaj lastnosti podatkov o razvrščanju člankov v tematske skupine

Gregor Majcen (63070199)

25. februar 2012

1 Uvod

Prva domača naloga je namenjena, da se seznanimo s tipom podatkov, ki jih bomo tudi v nadalnje uporabljali in pa tudi malo o statistiki le teh. Odgovoriti je potrebno na 8 vprašanj, med katerimi so 3 lastna.

2 Rezultati

Koliko primerov vsebujejo podatki?

2000 primerov

Koliko atributov vsebujejo podatki?

10000 atributov

Kakšnega tipa so atributi?

zvezni

Kakšen delež elementov matrike ima vrednost različno od 0?

0.412115%

Koliko je vseh različnih oznak (razredov) v podatkih?

82 različnih oznak

Koliko neničelnih atributov ima primer, ki ima največ atributov enakih 0?

21 atributov

Koliko neničelnih atributov ima primer, ki ima največ atributov različnih od 0?

79 atributov

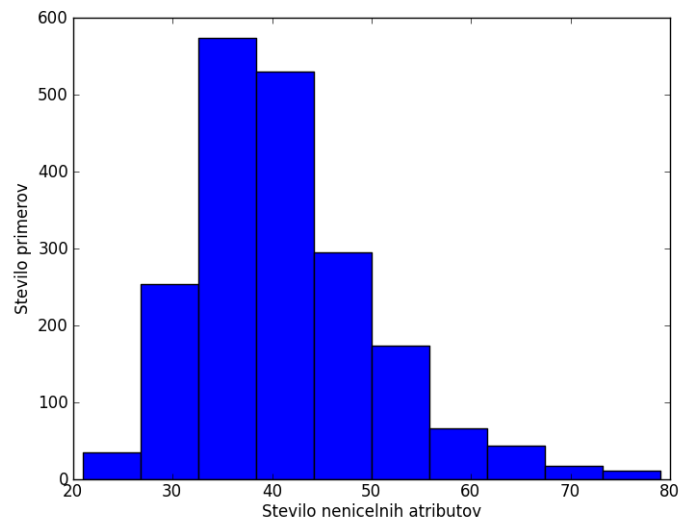
Koliko je primerov, ki imajo manj kot 30 neničelnih atributov?

1878 primerov

Koliko je atributov, ki imajo povsod vrednost enako 0?

2349 atributov

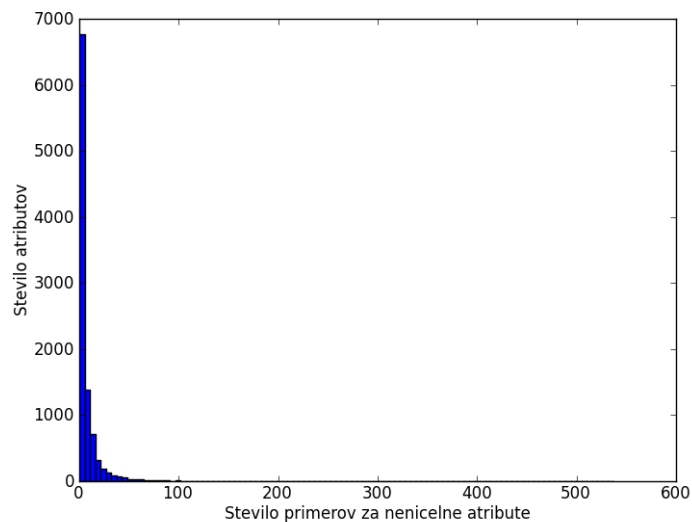
Koliko atributov ima vrednost različno od 0 za posamezen primer?



Slika 1: Histogram, ki prikazuje koliko je nenicelnih atributov za posamezen primer

Z grafa je lepo opazno, da je največ primerov s približno 35 do 45 nenicelnimi atributi.

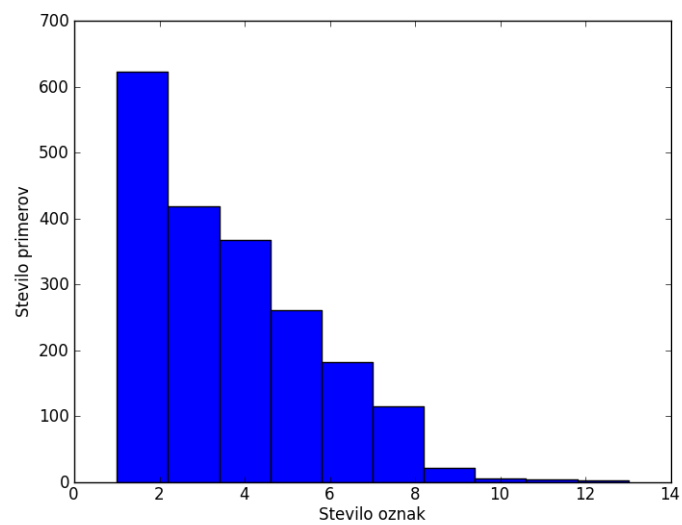
V koliko primerih atribut zavzame neničelne vrednosti?



Slika 2: Histogram, ki prikazuje v koliko primerih atribut zavzame neničelne vrednosti

Z grafa je lepo opazno, da je zelo malo izkoriščenih atributov, saj je vse zelo blizu ničle.

S koliko različnimi oznakami so označeni primeri?



Slika 3: Histogram, ki prikazuje s koliko različnimi oznakami so označeni primeri

Z grafa je lepo opazno, da sta dve oznaki za primer kar najbolj pogosti, nato pa vseskozi manj pogosto.

3 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelal sam.