

# Prva domača naloga

Anže Pečar (63060257)

3. marec 2012

## 1 Uvod

Naša naloga je bila izračunati kako informativne so značilke za posamezen razred. Podatke smo najprej binarilizirali, nato pa smo s pomočjo permutacijskega izračunali medsebojno informacijo. Delali smo na zmanjšanem naboru podatkov, ki jih uporablja tekmovanje *JRS 2012 Data Mining Competition: Topical Classification of Biomedical Research Papers*.

## 2 Metode

### 2.1 Količina medsebojne informacije

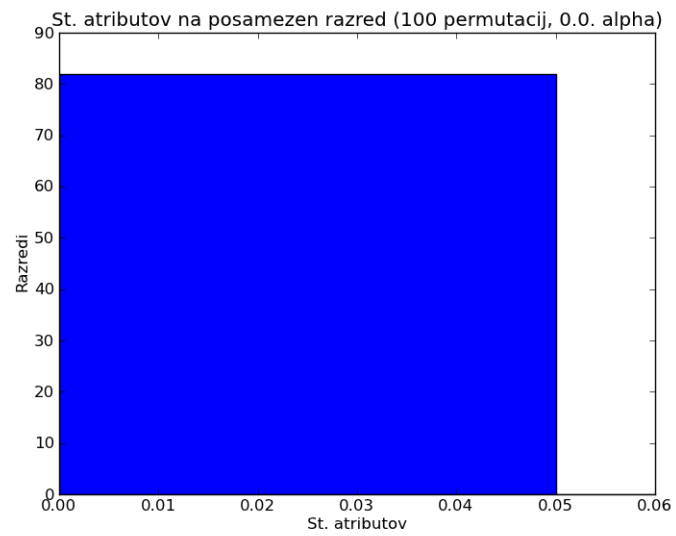
Količina medsebojne informacije

$$H(C) = -\left(\frac{20}{50} \log_2 \frac{20}{50} + \frac{30}{50} \log_2 \frac{30}{50}\right) =$$

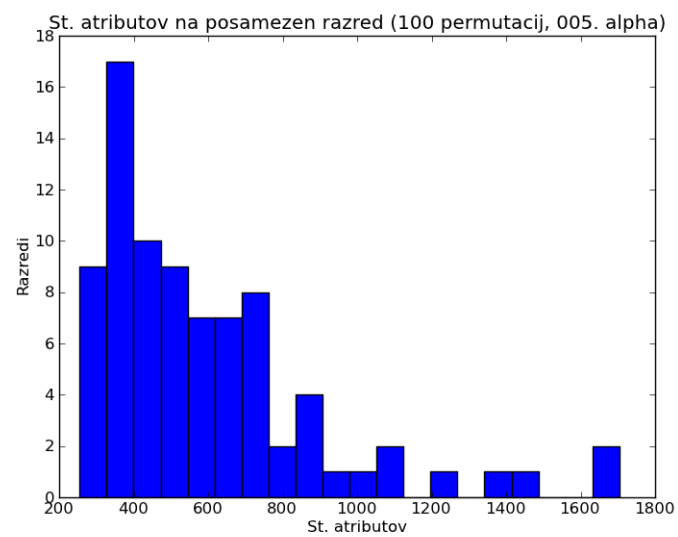
## 3 Rezultati

### 3.1 Grafi

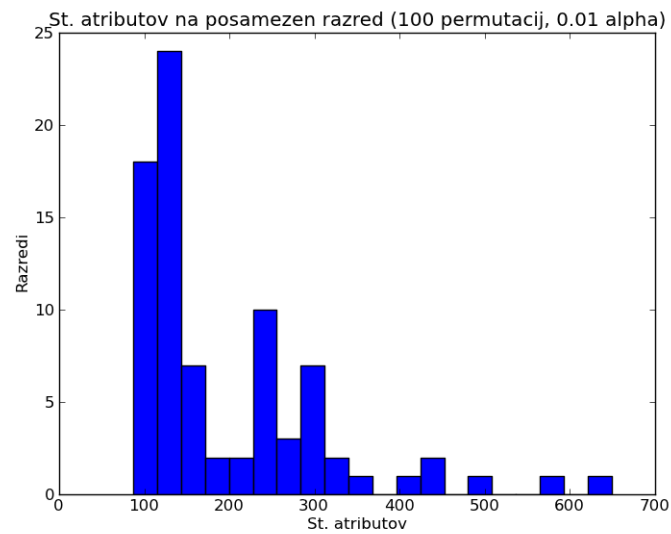
Testni primeri:



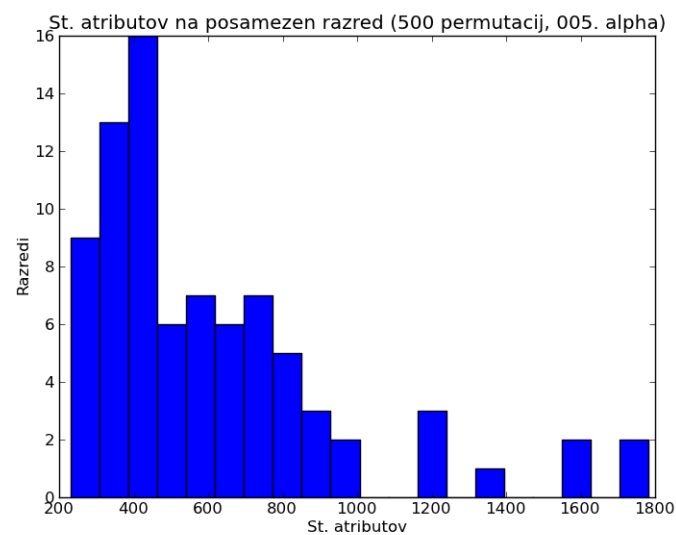
Slika 1: TEST



Slika 2:  $\alpha = 0.05$



Slika 3: Prikaz atributov, ki imajo vrednost različno od 0 za posamezen primer



Slika 4: 500

### 3.2 Hitrost izvajanja

10 permutacij:

real 8m20.437s

user 8m18.875s

sys 0m0.628s

100 permutacij:

```
real 73m46.716s
user 73m36.088s
sys 0m2.380s
```

500 permutacij:

```
real 377m52.873s
user 377m18.379s
sys 0m7.308s
```

## 4 Izjava o izdelavi domače naloge

Domačo nalogo in pripadajoče programe sem izdelal sam.