

**UNIVERZITET U SARAJEVU**  
**ELEKTROTEHNIČKI FAKULTET SARAJEVO**

# **DOMAĆA ZADAĆA 1**

## **Zadatak 3**

### **MAŠINSKO UČENJE**

**Odsjek: Računarstvo i Informatika**

**Datum: 18.11.2019**

**Studenti:**

- **Mašović Haris, 1689/17993**
- **Muminović Amir, 1661/17744**

### Zadatak 3 (1.5 bodova)

Za model koji ispunjava minimalne performanse, generišite pravila klasifikacije tako da je coverage  $\geq 80\%$

U zadatku 1, više modela je zadovoljilo minimalne tražene performanse. Za izradu trećeg zadatka odabran je prvi model (konceptualno najjednostavniji), koji je podijeljen koristeći holdout metodu i treniran koristeći C5.0 algoritam.

Dijeljenje podataka	Način treniranja modela	Pos pred value ( $\geq 0.65$ )	Balanced Accuracy ( $\geq 0.6$ )	Kappa statistics ( $\geq 0.25$ )	Lower Bound CI ( $\geq 0.8$ )
Holdout	C5.0	0.9264	0.6910	0.3440	0.8021

Prvenstveni model je imao 21 pravilo (pravilo definišemo kao putanja od početnog do krajnjeg odnosno leaf čvora) od kojeg 3 pravila nisu imala traženu tačnost. Shodno tome, obavljeno je čišćenje stabla kako bi se poboljšala tačnost. Kako bi se to postiglo sa C5.0 dodane su određene kontrole varijable.

```
ctrl = C5.0Control(  
  bands = 20,  
  winnow = TRUE,  
  CF = 0.19,  
  minCases = 5,  
  fuzzyThreshold = TRUE,  
)
```

Specifično promijenjen je interval pouzdanosti sa vrijednosti 0.25 na 0.19. Bands varijabla je postavljena na 20 što specifikira broj grupa u koja će pravila biti podijeljena. Fuzzy threshold je logička varijabla koja uključuje napredne metode podjele podataka je aktivirana. Minimalni broj pravila je postavljen na 5. Winnow je opcija koja omogućava selekciju atributa.

Sa ovim postavkama dobiju se sljedeći rezultati::

-----  
Class specified by attribute 'outcome'

Read 430 cases (21 attributes) from undefined.data

Rules:

Rule 1: (205/15, lift 1.1)  
WorkLifeBalance in {Best, Better, Good}  
StockOptionLevel in {1, 2}  
-> class No [0.923]

Rule 2: (311/25, lift 1.1)  
OverTime = No  
-> class No [0.917]

Rule 3: (162/14, lift 1.1)  
WorkLifeBalance = Better  
TotalWorkingYears > 8  
-> class No [0.909]

Rule 4: (128/11, lift 1.1)  
WorkLifeBalance in {Best, Better, Good}  
JobSatisfaction = 4  
-> class No [0.908]

Rule 5: (8, lift 5.5)  
WorkLifeBalance in {Best, Good}  
JobSatisfaction in {1, 2, 3}  
OverTime = Yes  
StockOptionLevel in {0, 3}  
-> class Yes [0.900]

Rule 6: (5, lift 5.2)  
WorkLifeBalance = Bad  
OverTime = Yes  
-> class Yes [0.857]

Rule 7: (20/3, lift 5.0)  
TotalWorkingYears <= 8  
JobSatisfaction in {1, 2, 3}  
OverTime = Yes  
StockOptionLevel in {0, 3}  
-> class Yes [0.818]

Default class: No

```

Evaluation on training data (430 cases):

      Rules
-----
      No      Errors

      7    46(10.7%)  <<

      (a)  (b)  <-classified as
      ----  ----
      356    3   (a): class No
      43    28   (b): class Yes

Attribute usage:

79.53% OverTime
79.30% WorkLifeBalance
53.72% StockOptionLevel
42.33% TotalWorkingYears
35.81% JobSatisfaction

Evaluation on test data (393 cases):

      Rules
-----
      No      Errors

      7    67(17.0%)  <<

      (a)  (b)  <-classified as
      ----  ----
      309    15   (a): class No
      52    17   (b): class Yes

```

**Možemo vidjeti da je kreirano 7 pravila i da svako od tih pravila ima pokrivenost veću ili jednaku 80%**