

Stabla odlučivanja

DECISION TREE LEARNING

Stabla odlučivanja (1)

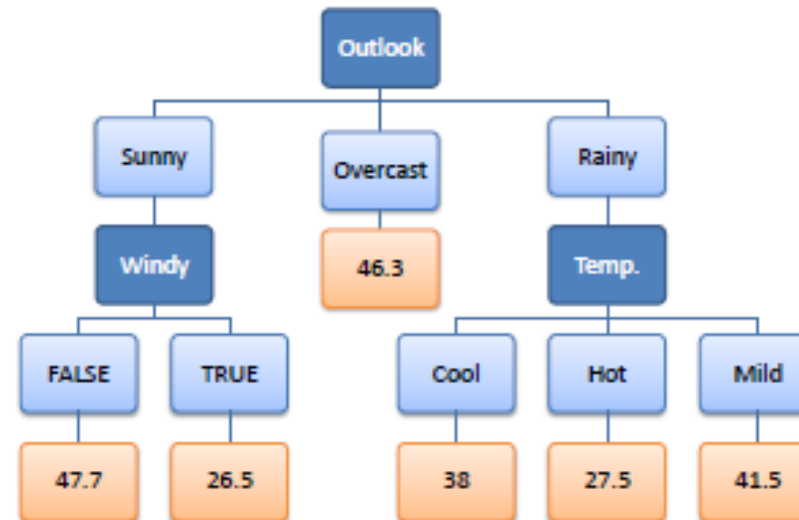
- ▶ Tip nadgledanih algoritama (eng. *supervised learning*)
- ▶ Najčešće se koriste kod problema klasifikacije
- ▶ Cilj: Model koji predviđa izlaznu vrednost na osnovu ulaznih vrednosti nekoliko parametara
- ▶ Ulazne vrednosti, kao i izlazne, mogu biti kategoričkog ili kontinualnog tipa
- ▶ Svaki čvor odgovara ulaznom parametru, svaka grana (koja potiče iz posmatranog čvora) odgovara nekoj vrednosti tog parametra
- ▶ Listovi predstavljaju izlazne parametre u zavisnosti od vrednosti ulaznih parametara na datom putu kroz stablo, od korena do posmatranog lista (to su obeležja klase, tzv. *class label*)
- ▶ Obeležja klase: da ili ne, + ili -, itd.

Stabla odlučivanja (2)

- ▶ Prema tipu izlazne vrednosti:
 - ▶ Klasifikaciona stabla
 - ▶ Regresivna stabla
- ▶ Kod klasifikacionih stabala izlazna promenljiva je kategoričkog tipa
- ▶ Kod regresivnih stabala izlazna promenljiva je kontinualnog tipa
- ▶ Oba tipa stabala dele prostor za predviđanje u nepreklapajuće regije

Stabla odlučivanja (3)

Predictors				Target
Outlook	Temp.	Humidity	Windy	Hours Played
Rainy	Hot	High	False	26
Rainy	Hot	High	True	30
Overcast	Hot	High	False	48
Sunny	Mild	High	False	46
Sunny	Cool	Normal	False	62
Sunny	Cool	Normal	True	23
Overcast	Cool	Normal	True	43
Rainy	Mild	High	False	36
Rainy	Cool	Normal	False	38
Sunny	Mild	Normal	False	48
Rainy	Mild	Normal	True	48
Overcast	Mild	High	True	62
Overcast	Hot	Normal	False	44
Sunny	Mild	High	True	30



Primena klasifikacije

- ▶ Medicinska dijagnostika
- ▶ Upotreba kreditnih kartica
- ▶ Detekcija prevara u elektronskom poslovanju
- ▶ Otkrivanje virusa i neželjenog sadržaja, filtriranje spam e-mailova
- ▶ Sistem preporučivanja na internetu (eng. Recommender Systems)
- ▶ Investiciono bankarstvo
- ▶ Prepoznavanje govora
- ▶ Prepoznavanje slika
- ▶ Prepoznavanje rukopisa

Izgradnja stabla

- ▶ Stabla se grade od korena, ka listovima, tzv. "pohlepnim" pristupom, tj. rekurzivnim binarnim deljenjem
- ▶ Na početku, sve instance pripadaju jednoj regiji, a prostor se sukcesivno deli na regione
- ▶ Za deljenje kažemo da je pohlepno zato što se pri svakom koraku najbolja podela određuje na osnovu stanja u posmatranom koraku, odnosno ne uzima se u obzir kako će se podela izvršiti u narednim koracima i koja bi podela mogla dovesti do boljih rezultata u narednim koracima
- ▶ Postupak podele se nastavlja dok se ne dosegne do definisanog kriterijuma za zaustavljanje deljenja
- ▶ Kod oba tipa stabala se na kraju ovog postupka dobija kompletno razgranato stablo, gde su listovi dostigli definisani kriterijum zaustavljanja
- ▶ Na tačnost predviđanja stabla najviše utiče metod kojim se vrši račvanje u čvorovima
- ▶ Kriterijum podele je različit u zavisnosti od toga da li se gradi regresivno ili klasifikaciono stablo
- ▶ Izbor algoritma se vrši na osnovu tipa stabla koji se gradi

Izbor najboljeg atributa

- ▶ Kako izabrati "najbolji atribut"?
 - ▶ **Random**: slučajnim izborom izabrati bilo koji atribut
 - ▶ **Least-values**: izabrati atribut koji ima najmanji broj mogućih vrednosti
 - ▶ **Most-values**: izabrati atribut sa najvećim brojem mogućih vrednosti
 - ▶ **Max-Gain**: izabrati atribut koji ima najveću očekivanu dobit informacija (information gain)

Dobitak informacija (eng. *Information Gain*)

- ▶ Kako se određuje information gain?
 - ▶ **cilj:** pokušati izabrati atribut koji će realizovati najmanjim podstablama koja počinju njegovim naslednicima
 - ▶ koristi teoriju informacija (eng. *information theory*)
- ▶ Koliko da/ne pitanja očekujete da pitate radi određivanja broja koji sam zamislio u opsegu od 1 do 100?
- ▶ Rezultat: 7
- ▶ Sa svakim da/ne pitanjem u optimalnom stablu odlučivanja najviše $\frac{1}{2}$ od preostalih elemenata može biti eliminisana
- ▶ $\log_2 100 = 6.64$
- ▶ Ako je dat skup S veličine $|S|$, očekivan broj pokušaja radi određivanja konkretnog elementa je: $\log_2 |S|$
- ▶ Neka ova vrednost bude vrednost informacije saznanja, a da nismo ni morali da postavljamo pitanja

Entropija (1)

- ▶ Entropija skupa primera, S , za binarnu klasifikaciju (ishod ima samo dve klase) je:

$$\text{Entropy}(S) = -p_1 \log_2(p_1) - p_0 \log_2(p_0)$$

gde je p_1 procenat pozitivnih primera u S , a p_0 procenat negativnih

- ▶ Ako su svi primeri isti (pripadaju istoj klasi) entropija je nula (koristimo $0 * \log(0) = 0$)
- ▶ Ako su svi primeri podjednako raspoređeni ($p_1 = p_0 = 0.5$), entropija ima maksimalnu vrednost 1
- ▶ Entropija se može posmatrati kao prosečan broj bitova potreban za kodiranje klase primera iz skupa S , gde su češćim slučajevima kompresijom dodeljeni kraći kodovi.
- ▶ Za problem gde postoji više klasa ishoda (multi-class), entropija se računa kao:

$$\text{Entropy}(S) = \sum_{i=1}^c -p_i \log_2(p_i)$$

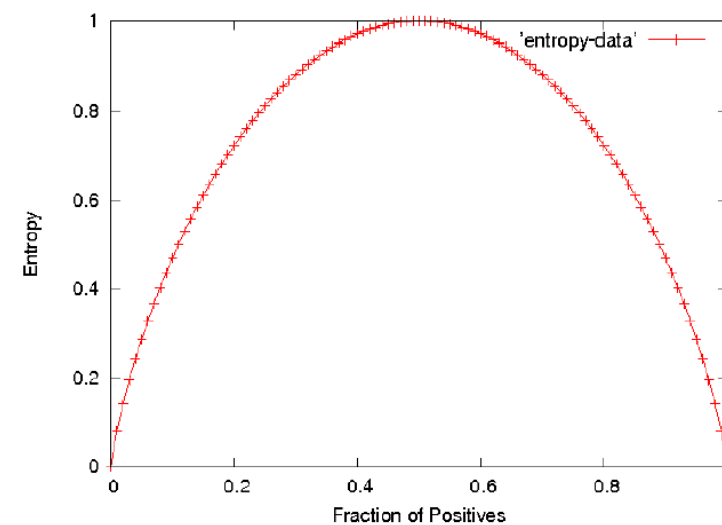
gde je c broj klasa.

Entropija (2)

- ▶ Ako ima n jednako verovatnih poruka, onda je verovatnoća p svake poruke $1/n$, a informacija koja se prenosi porukom je: $-\log_2(p) = \log_2(n)$
- ▶ To znači ako ima 16 poruka, onda je $\log_2(16) = 4$, i treba nam 4 bita
- ▶ Uopšteno, ako nam je data raspodela verovatnoća: $P = (p_1, p_2, \dots, p_n)$, onda se informacija koju nosi ova distribucija zove entropija za P :

$$I(P) = -(p_1 \cdot \log_2(p_1) + p_2 \cdot \log_2(p_2) + \dots + p_n \cdot \log_2(p_n))$$

- ▶ Na primer: ako je $P (0.5, 0.5)$ onda je $I(P) = 1$, ako je $P (0.67, 0.33)$, onda je $I(P) = 0.92$, ako je $P (1, 0)$, onda je $I(P) = 0$.
- ▶ Što je uniformnija raspodela, više informacija nosi.
- ▶ Entropija bacanja fer novčića je jedan bit po bacanju.



Over-fitting problem

- ▶ Kod kompletno razgranatih stabala problem je over-fitting podataka
- ▶ Over-fitting podataka se javlja kada model savršeno opisuje testni skup, ali to ne znači da ima dobru tačnost kod procene podataka koji nisu bili u testnom skupu
- ▶ Često se radi "orezivanje stabla" - odsecanje nekih terminalnih čvorova
- ▶ Izbor čvorova koji se odsecaju se vrši upotrebom validacije kako bi se otkrilo koje podstablo daje najmanju grešku

Važnost stabala odlučivanja

- ▶ Stabla odlučivanja su dobra za grafičko predstavljanje i jednostavno interpretiranje posmatranog modela, bez obzira da li koristimo klasifikaciono ili regresivno stablo
- ▶ Prednost: možemo koristiti i ukoliko pojedini atributi imaju nedostajuće vrednosti
- ▶ Nedostatak: slabija tačnost predikcije u poređenju sa drugim pristupima mašinskog učenja

Primer 1 - Koja životinja leže jaja?

Nezavisni atributi					Atributi odluke
Životinja	Toplokrvna	Pernata	Ima krzno	Pliva	Leže jaja
noj	da	da	ne	ne	da
krokodil	ne	ne	ne	da	da
gavran	da	da	ne	ne	da
albatros	da	da	ne	ne	da
delfin	da	ne	ne	da	ne
koala	da	ne	da	ne	ne

Primer 1 - Rešenje (1)

$$Entropy(S) = \sum_{i=1}^c -p_i \cdot \log_2(p_i)$$

$$Gain(S, F) = Entropy(S) - \sum_{v \in Values(F)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$Entropy(4Y, 2N) = -\frac{4}{6} \cdot \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \cdot \log_2\left(\frac{2}{6}\right) = 0.91829$$

- Sada je potrebno pronaći Information Gain za sve attribute: toplokrvna, pernata, ima krzno, pliva.

Primer 1 - Rešenje (2)

$$Gain(S, F) = Entropy(S) - \sum_{v \in Values(F)} \frac{|S_v|}{|S|} Entropy(S_v)$$

- ▶ Za atribut "toplokrvni": vrednosti su [DA, NE]

$$S = [4 \text{ DA}, 2 \text{ NE}]$$

$$S_{DA} = [3 \text{ DA}, 2 \text{ NE}] \quad E(S_{DA}) = 0.97095$$

$$S_{NE} = [1 \text{ DA}, 0 \text{ NE}] \quad E(S_{NE}) = 0 \quad (\text{svi članovi pripadaju istoj klasi})$$

$$Gain(S, \text{toplokrvna}) = 0.91829 - \left[\frac{5}{6} \cdot 0.97095 + \frac{1}{6} \cdot 0 \right] = 0.10916$$

- ▶ Za atribut "pernata": vrednosti su [DA, NE]

$$S = [4 \text{ DA}, 2 \text{ NE}]$$

$$S_{DA} = [3 \text{ DA}, 0 \text{ NE}] \quad E(S_{DA}) = 0$$

$$S_{NE} = [1 \text{ DA}, 2 \text{ NE}] \quad E(S_{NE}) = 0.91829$$

$$Gain(S, \text{pernata}) = 0.91829 - \left[\frac{3}{6} \cdot 0 + \frac{3}{6} \cdot 0.91829 \right] = 0.45914$$

Primer 1 - Rešenje (3)

$$Gain(S, F) = Entropy(S) - \sum_{v \in Values(F)} \frac{|S_v|}{|S|} Entropy(S_v)$$

- ▶ Za atribut "ima krzno": vrednosti su [DA, NE]

$$S = [4 \text{ DA}, 2 \text{ NE}]$$

$$S_{DA} = [0 \text{ DA}, 1 \text{ NE}] \quad E(S_{DA}) = 0$$

$$S_{NE} = [4 \text{ DA}, 1 \text{ NE}] \quad E(S_{NE}) = 0.7219$$

$$Gain(S, \text{ima krzno}) = 0.91829 - \left[\frac{1}{6} \cdot 0 + \frac{5}{6} \cdot 0.7219 \right] = 0.3167$$

- ▶ Za atribut "pliva": vrednosti su [DA, NE]

$$S = [4 \text{ DA}, 2 \text{ NE}]$$

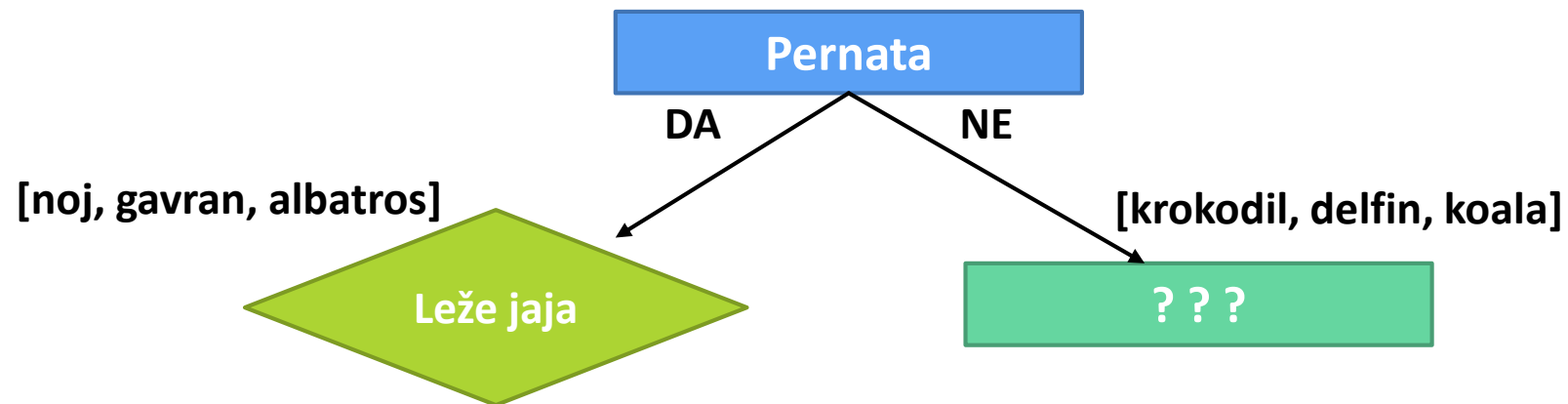
$$S_{DA} = [1 \text{ DA}, 1 \text{ NE}] \quad E(S_{DA}) = 1 \quad (\text{isti broj članova obe klase})$$

$$S_{NE} = [3 \text{ DA}, 1 \text{ NE}] \quad E(S_{NE}) = 0.81127$$

$$Gain(S, \text{pliva}) = 0.91829 - \left[\frac{2}{6} \cdot 1 + \frac{4}{6} \cdot 0.81127 \right] = 0.04411$$

Primer 1 - Rešenje (4)

- ▶ $Gain(S, toplokrvna) = 0.10916$ $Gain(S, pernata) = 0.45914$
- ▶ $Gain(S, ima\ krzno) = 0.31670$ $Gain(S, pliva) = 0.04411$
- ▶ $Gain(S, pernata)$ je maksimum od 4 atributa, pa je izabran kao koreni čvor



Primer 1 - Rešenje (5)

Životinja	Toplo-krvna	Perje	Krzno	Pliva	Leže jaja
krokodil	NE	NE	NE	DA	DA
delfin	DA	NE	NE	DA	NE
koala	DA	NE	DA	NE	NE

- ▶ Sada ponavljamo postupak:
S : [krokodil, delfin, koala]
S : [1+, 2-]

$$Entropy(S) = \sum_{i=1}^c -p_i \cdot \log_2(p_i)$$

$$Entropy(S) = -\left(\frac{1}{3}\right) \cdot \log_2\left(\frac{1}{3}\right) - \left(\frac{2}{3}\right) \cdot \log_2\left(\frac{2}{3}\right) = 0.91829$$

Primer 1 - Rešenje (6)

- ▶ Za atribut "toplokrvna": vrednosti su [DA, NE]

$$S = [1 \text{ DA}, 2 \text{ NE}]$$

$$S_{\text{DA}} = [0 \text{ DA}, 2 \text{ NE}] \quad E(S_{\text{DA}}) = 0$$

$$S_{\text{NE}} = [1 \text{ DA}, 0 \text{ NE}] \quad E(S_{\text{NE}}) = 0$$

$$\text{Gain}(S, \text{toplokrvna}) = 0.91829 - \left[\frac{2}{3} \cdot 0 + \frac{1}{3} \cdot 0 \right] = \mathbf{0.91829}$$

- ▶ Za atribut "krzno": vrednosti su [DA, NE]

$$S = [1 \text{ DA}, 2 \text{ NE}]$$

$$S_{\text{DA}} = [0 \text{ DA}, 1 \text{ NE}] \quad E(S_{\text{DA}}) = 0$$

$$S_{\text{NE}} = [1 \text{ DA}, 1 \text{ NE}] \quad E(S_{\text{NE}}) = 1$$

$$\text{Gain}(S, \text{krzno}) = 0.91829 - \left[\frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 \right] = \mathbf{0.25162}$$

- ▶ Za atribut "pliva": vrednosti su [DA, NE]

$$S = [1 \text{ DA}, 2 \text{ NE}]$$

$$S_{\text{DA}} = [1 \text{ DA}, 1 \text{ NE}] \quad E(S_{\text{DA}}) = 1$$

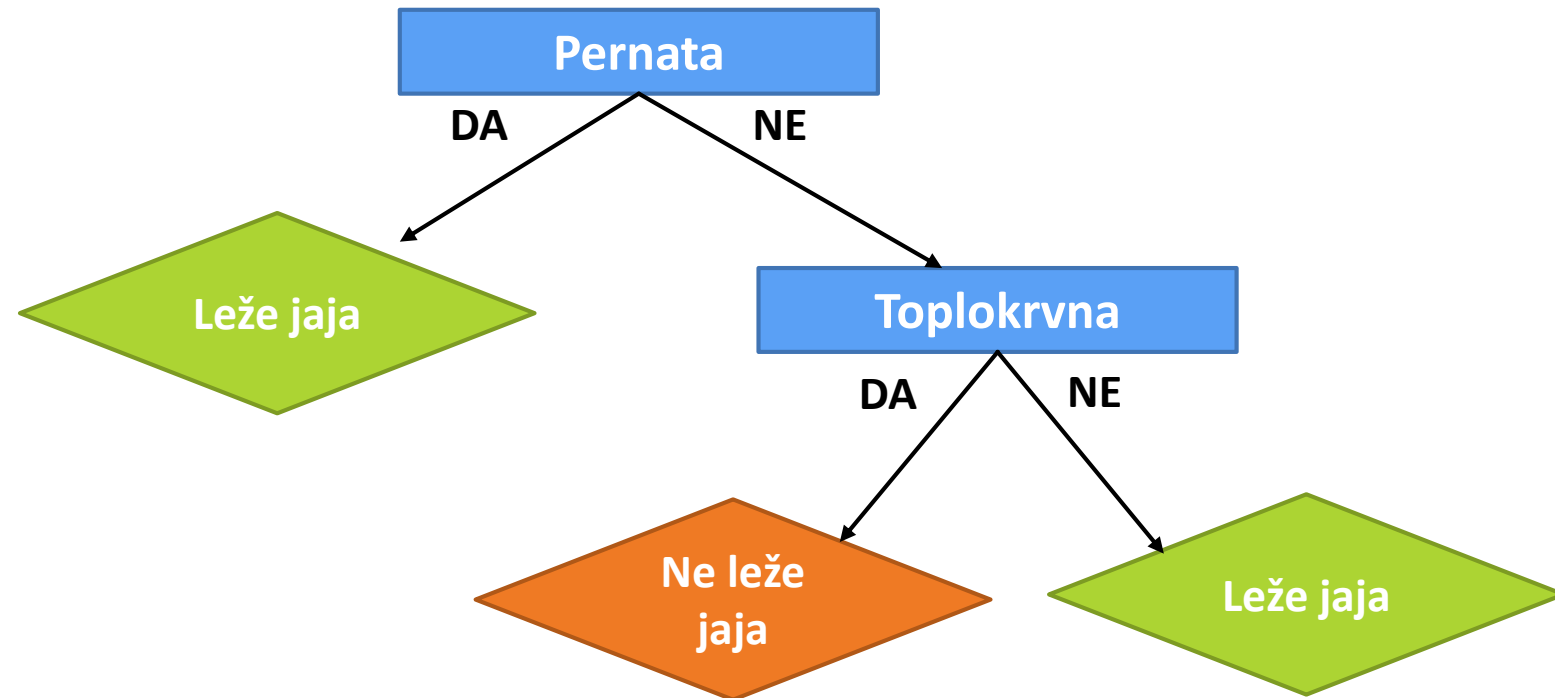
$$S_{\text{NE}} = [0 \text{ DA}, 1 \text{ NE}] \quad E(S_{\text{NE}}) = 0$$

$$\text{Gain}(S, \text{pliva}) = 0.91829 - \left[\frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 0 \right] = \mathbf{0.25162}$$

- ▶ **Maks je Gain (S, toplokrvna)**

Primer 1 - Rešenje (7)

- ▶ Konačno stablo odlučivanja će biti:

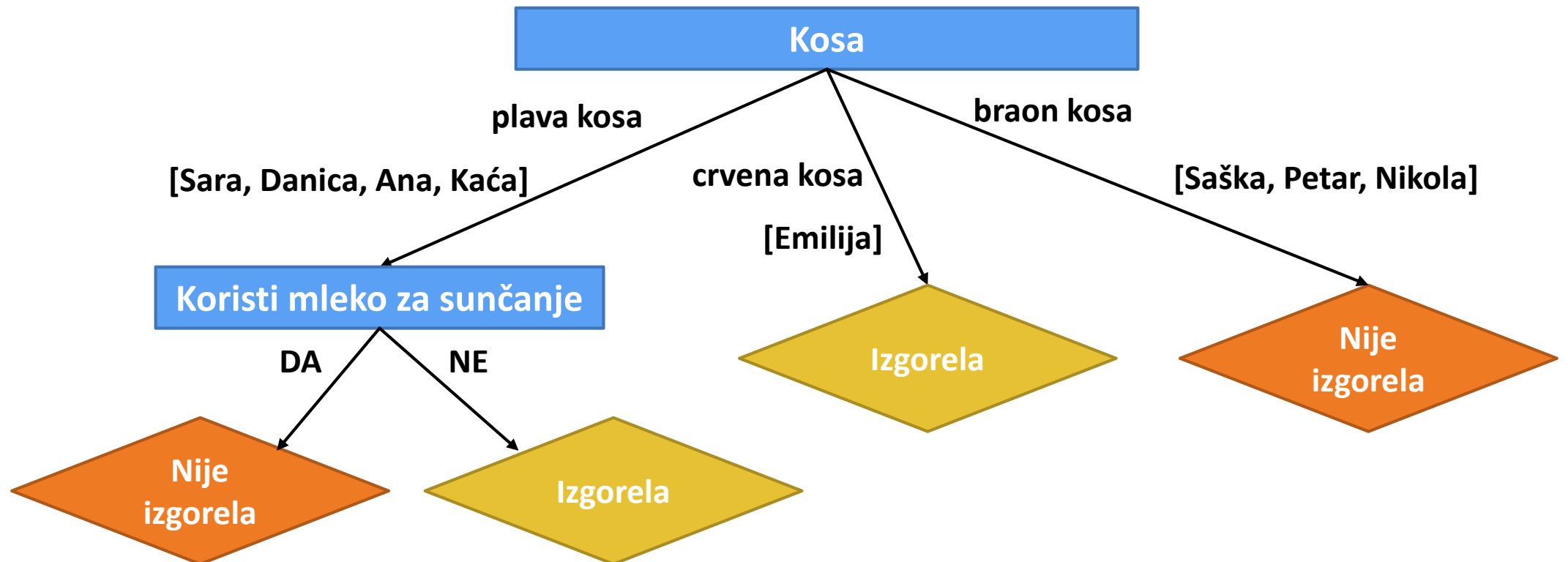


Primer 2 - Da li će osoba izgoreti na suncu?

Nezavisni atributi (ulaz)					Atributi odluke
Ime	Kosa	Visina	Težina	Koristi mleko za sunč.	Izgoreo/la od sunca
Sara	plava	prosečna	lak	ne	da
Danica	plava	visoka	prosečna	da	ne
Saška	braon	niska	prosečna	da	ne
Ana	plava	niska	prosečna	ne	da
Emilija	crvenokosa	prosečna	teška	ne	da
Petar	braon	visoka	teška	ne	ne
Nikola	braon	prosečna	teška	ne	ne
Kaća	plava	niska	lak	da	ne

Primer 2 - Rešenje

- ▶ Konačno stablo odlučivanja će biti:



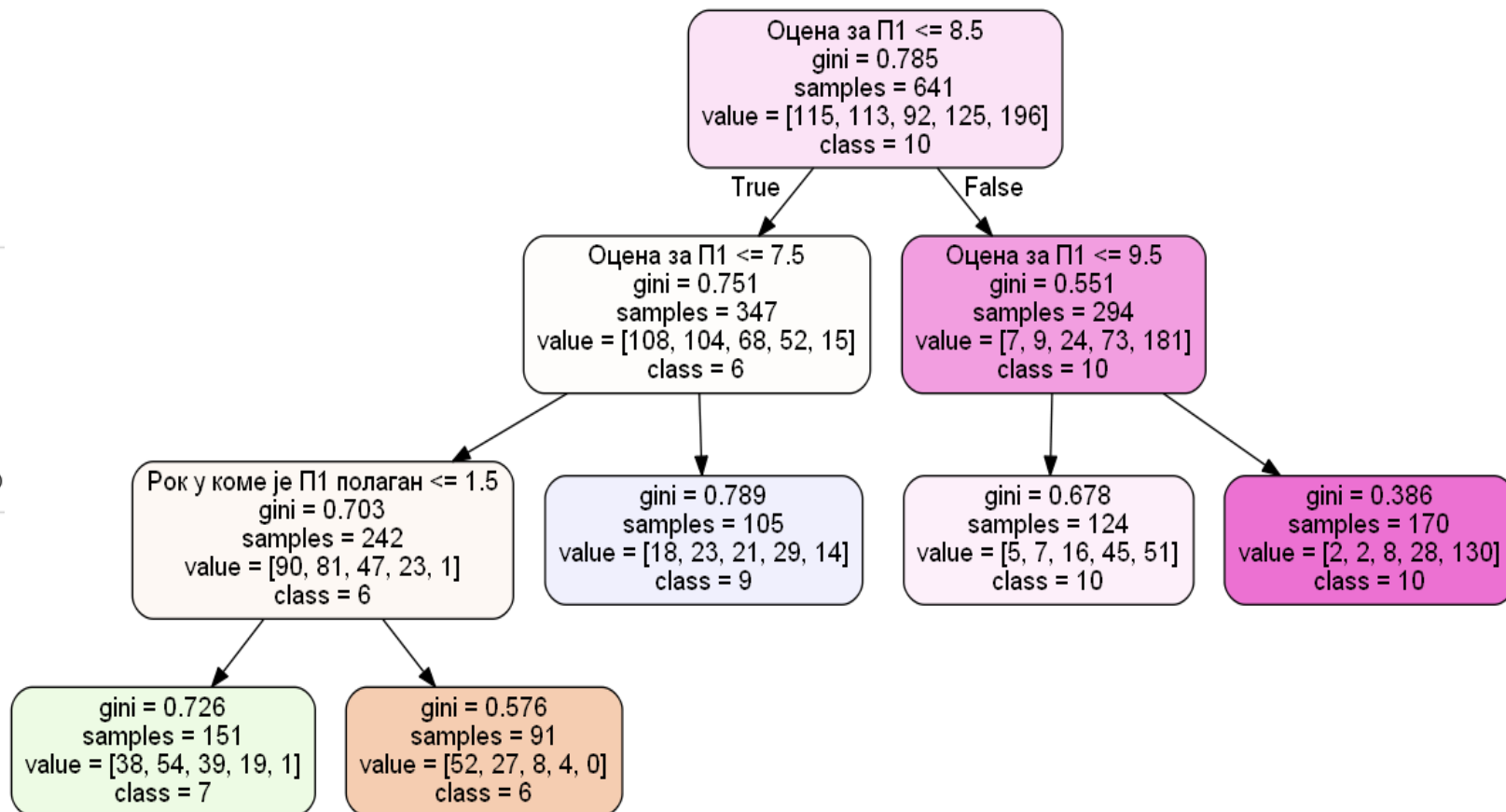
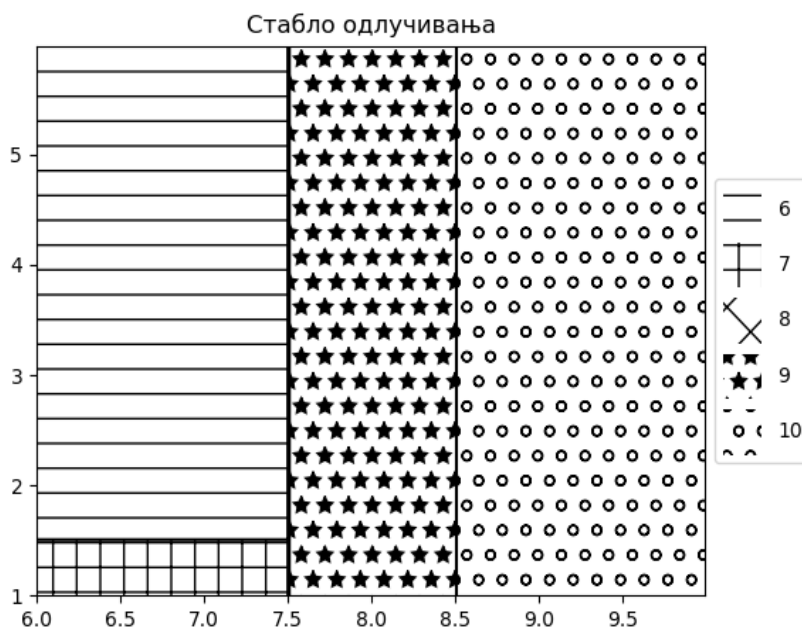
Eksperiment 1

- ▶ Analiza na podacima studenata koji su polagali ispite Programiranje 1 i Programiranje 2
- ▶ Podaci su izmešani i na slučajan način podeljeni u 3 skupa:
 - ▶ skup podataka za učenje
 - ▶ skup podataka za testiranje
 - ▶ skup podataka za validaciju modela
- ▶ Kategorijalni parametar koji predstavlja rok u kome je ispit polagan je pretvoren u numerički podatak, tako da su raniji rokovi manjih vrednosti, a kasniji većih vrednosti (ispit polagan u januarskom roku 1, u februarskom 2, itd.)

Eksperiment 1 - rezultati

- ▶ Za analizu podataka izabrana je vrednost parametra maksimalne dubine stabla 3, kao vrednost pri kojoj se javlja najbolja procena predviđanja za validacioni skup podataka
- ▶ Rezultati:
 - ▶ 42.75 % slučajeva gde je predviđena tačna ocena na P2 u odnosu na P1
 - ▶ 36.96 % slučajeva gde je predviđena ocena za jednu ocenu više ili niže od stvarno dobijene
 - ▶ 13.04 % slučajeva gde je predviđena ocena za dve viša ili niža
 - ▶ 7.25 % slučajeva gde je predviđanje preko dve ocene više ili niže

Експеримент 1 - резултати



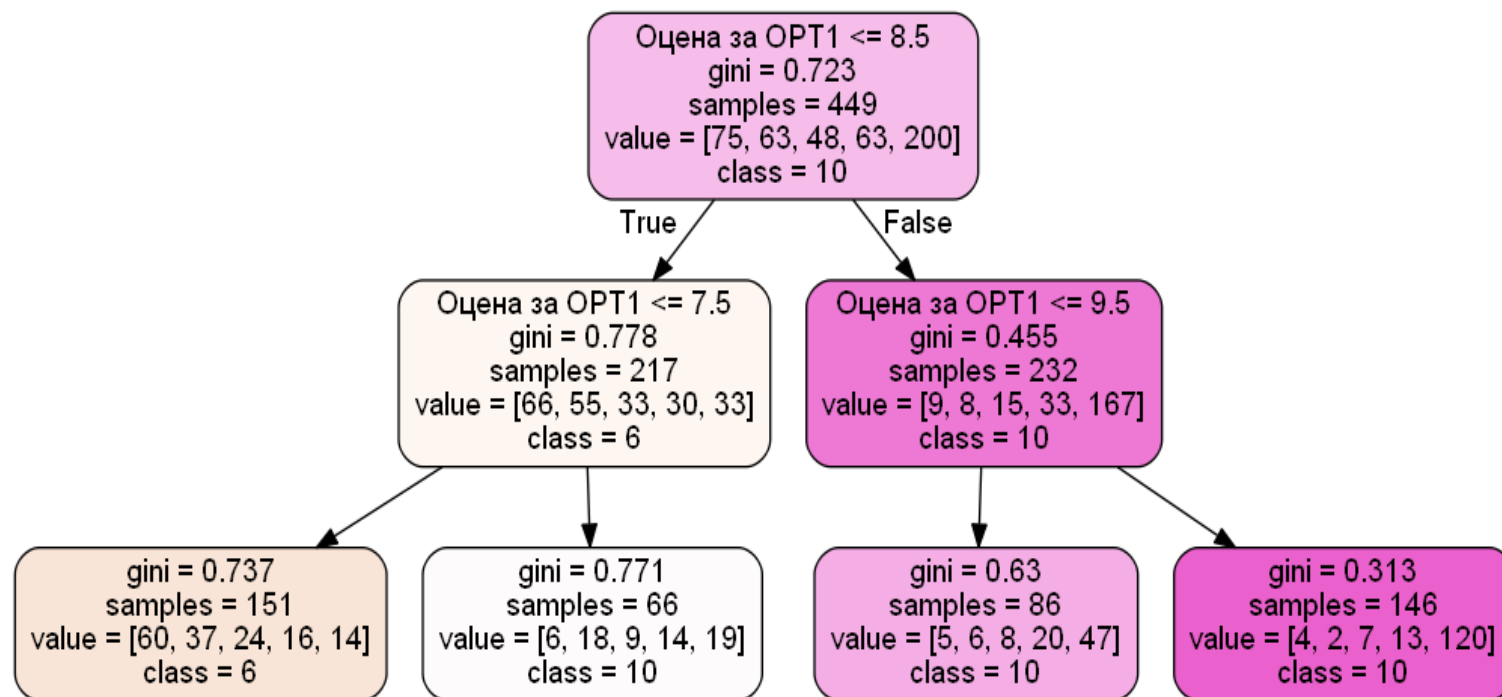
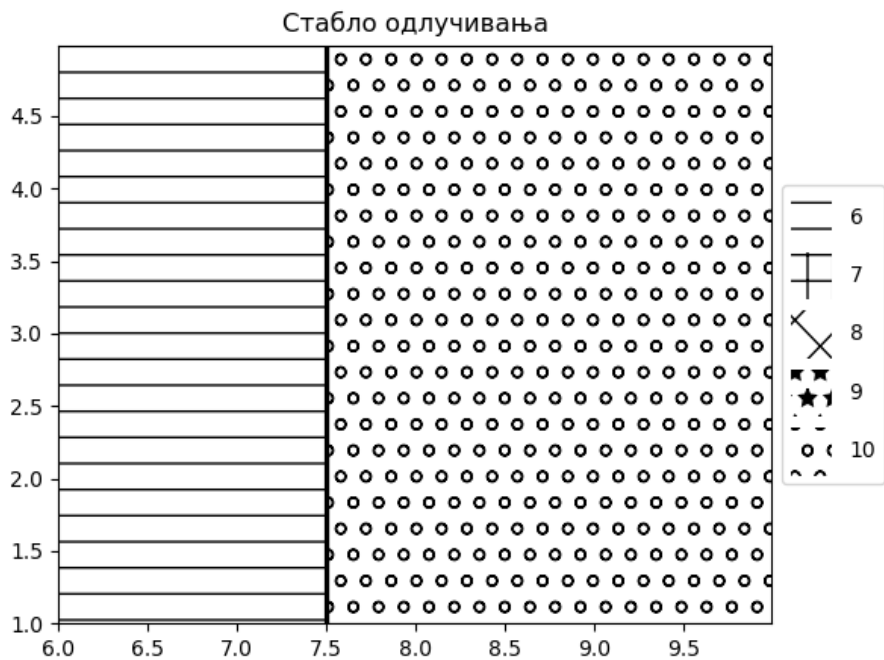
Eksperiment 2

- ▶ Analiza na podacima studenata koji su polagali ispite ORT 1 i ORT 2
- ▶ Podaci su izmešani i na slučajan način podeljeni u 3 skupa:
 - ▶ skup podataka za učenje
 - ▶ skup podataka za testiranje
 - ▶ skup podataka za validaciju modela
- ▶ Kategorijalni parametar koji predstavlja rok u kome je ispit polagan je pretvoren u numerički podatak, tako da su raniji rokovi manjih vrednosti, a kasniji većih vrednosti (ispit polagan u junskom roku 1, u julskom 2, itd.)

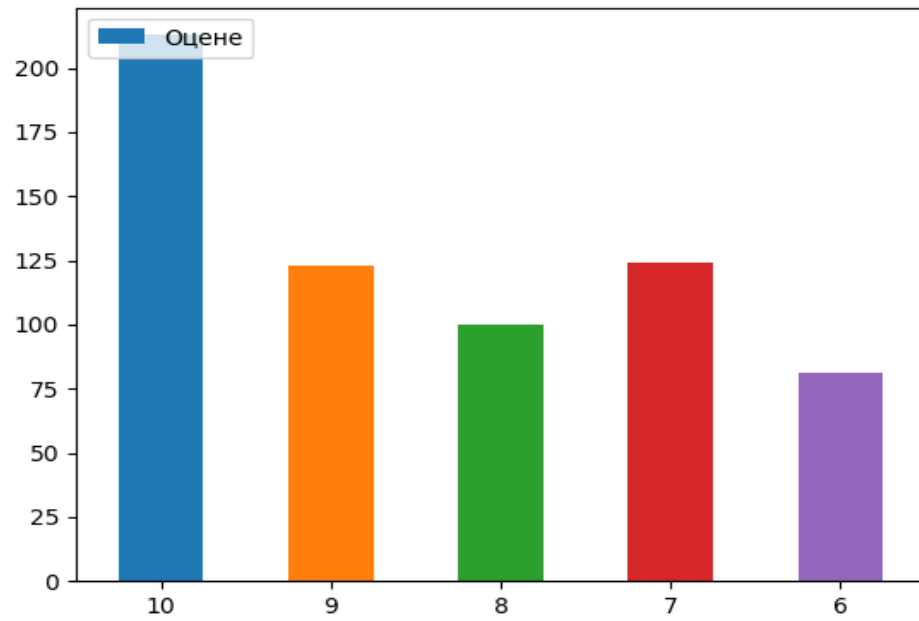
Eksperiment 2 - rezultati

- ▶ Za analizu podataka izabrana je vrednost parametra maksimalne dubine stabla 2, kao vrednost pri kojoj se javlja najbolja procena predviđanja za validacioni skup podataka
- ▶ Rezultati:
 - ▶ 60.82 % slučajeva gde je predviđena tačna ocena na ORT2 u odnosu na ORT1
 - ▶ 16.5 % slučajeva gde je predviđena ocena za jednu ocenu više ili niže od stvarno dobijene
 - ▶ 10.31 % slučajeva gde je predviđena ocena za dve viša ili niža
 - ▶ 12.37 % slučajeva gde je predviđanje preko dve ocene više ili niže

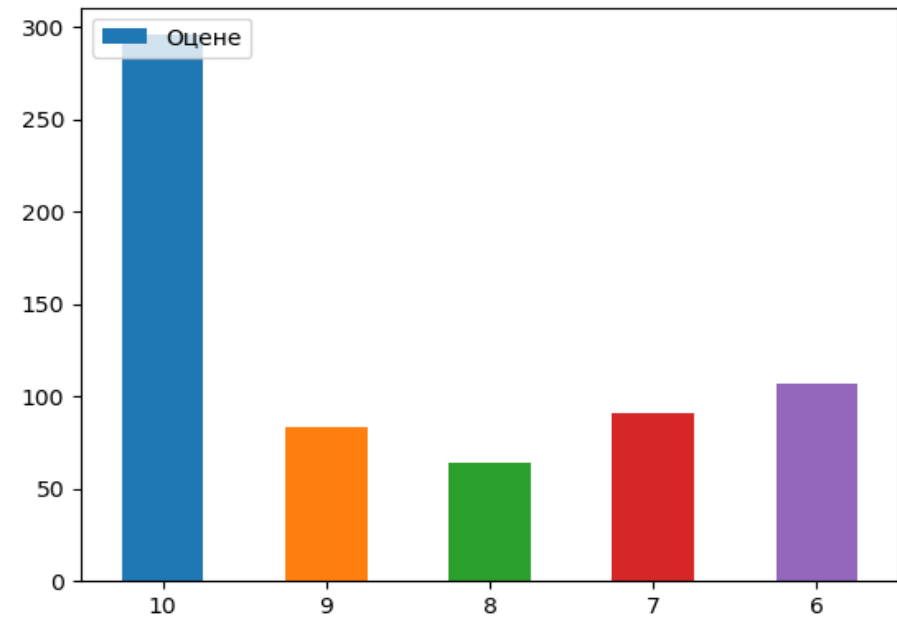
Експеримент 2 - резултати



Vizuelizacija



Дистрибуција оцена предмета Основи рачунарске технике 1



Дистрибуција оцена предмета Основи рачунарске технике 2

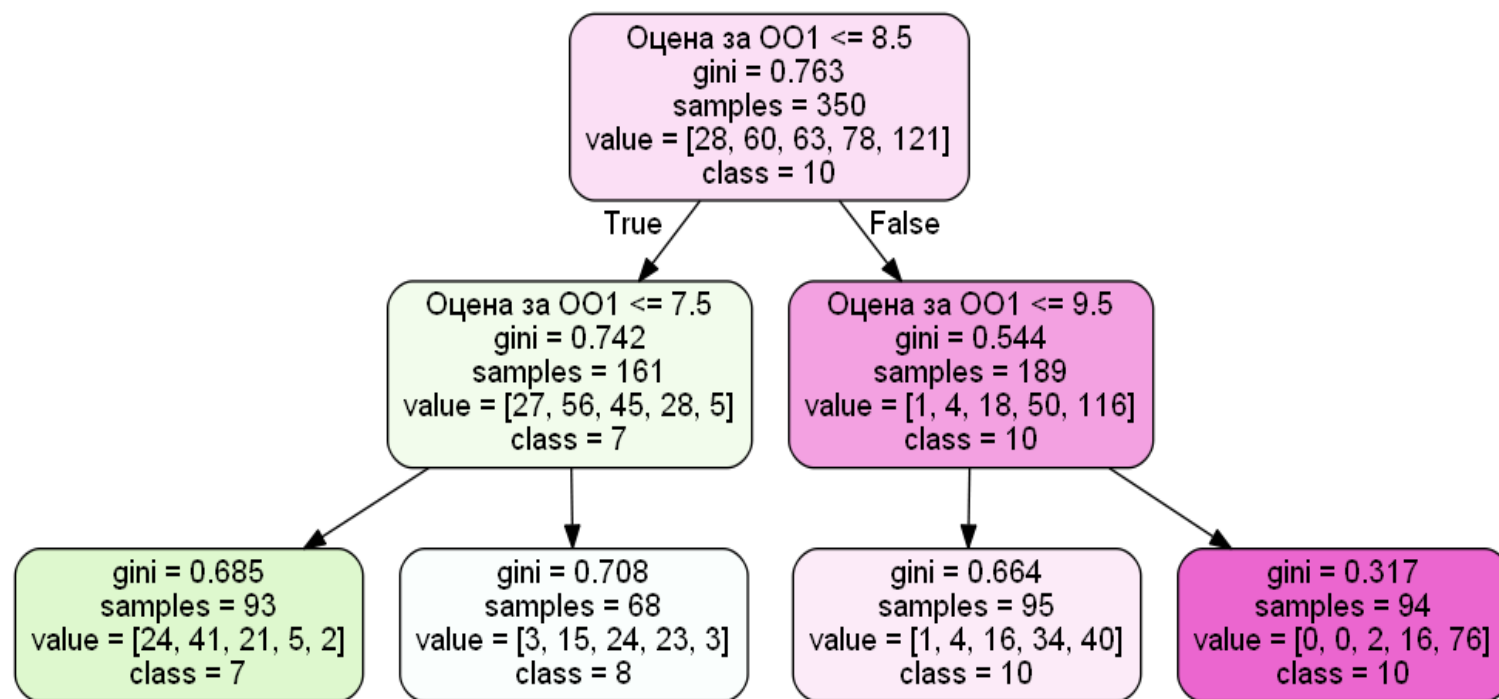
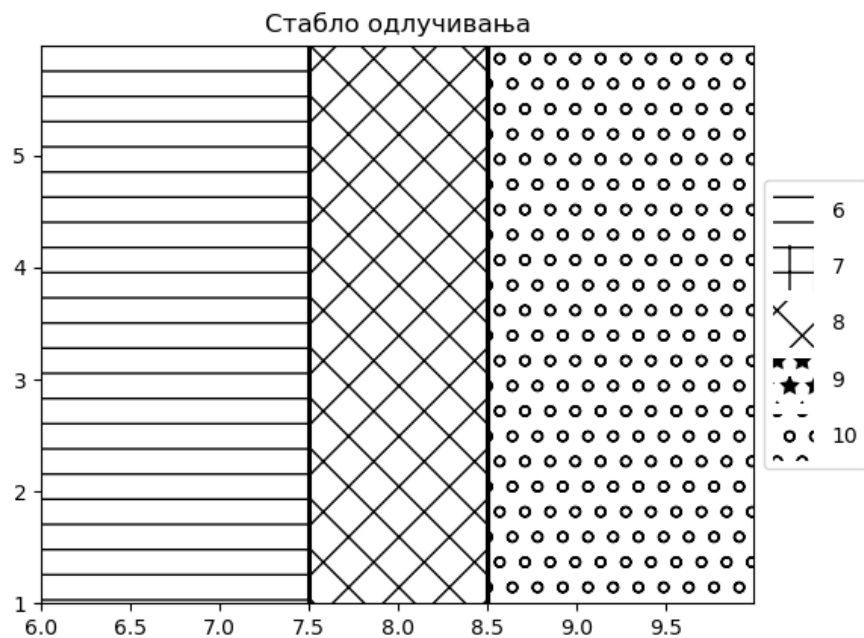
Eksperiment 3

- ▶ Analiza na podacima studenata koji su polagali ispite OOP 1 i OOP 2
- ▶ Podaci su izmešani i na slučajan način podeljeni u 3 skupa:
 - ▶ skup podataka za učenje
 - ▶ skup podataka za testiranje
 - ▶ skup podataka za validaciju modela
- ▶ Kategorijalni parametar koji predstavlja rok u kome je ispit polagan je pretvoren u numerički podatak, tako da su raniji rokovi manjih vrednosti, a kasniji većih vrednosti (ispit polagan u januarskom roku 1, u februarskom 2, itd.)

Eksperiment 3 - rezultati

- ▶ Za analizu podataka izabrana je vrednost parametra maksimalne dubine stabla 2, kao vrednost pri kojoj se javlja najbolja procena predviđanja za validacioni skup podataka
- ▶ Rezultati:
 - ▶ 53.95 % slučajeva gde je predviđena tačna ocena na OOP2 u odnosu na OOP1
 - ▶ 39.47 % slučajeva gde je predviđena ocena za jednu ocenu više ili niže od stvarno dobijene
 - ▶ 5.26 % slučajeva gde je predviđena ocena za dve viša ili niža
 - ▶ 1.32 % slučajeva gde je predviđanje preko dve ocene više ili niže

Експеримент 3 - резултати



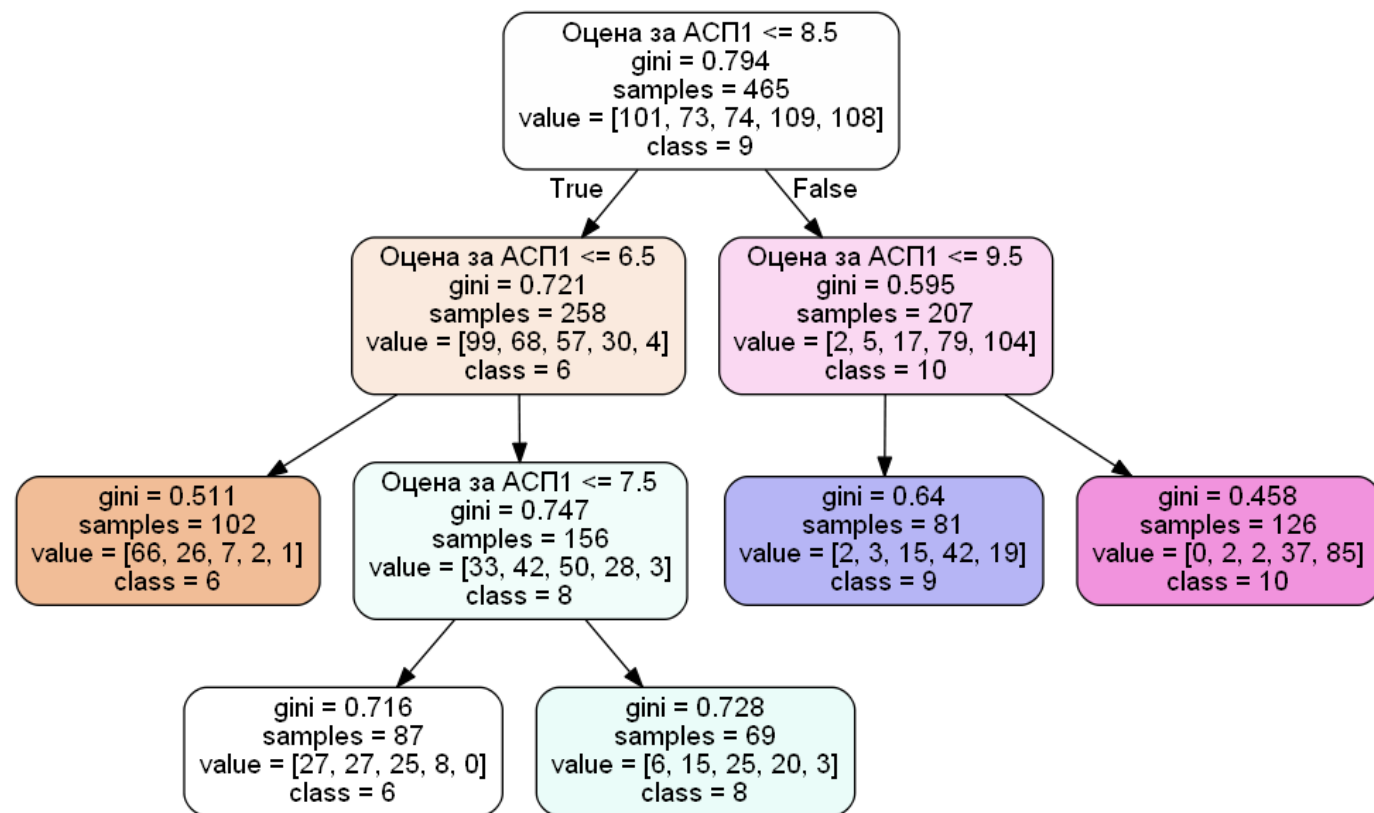
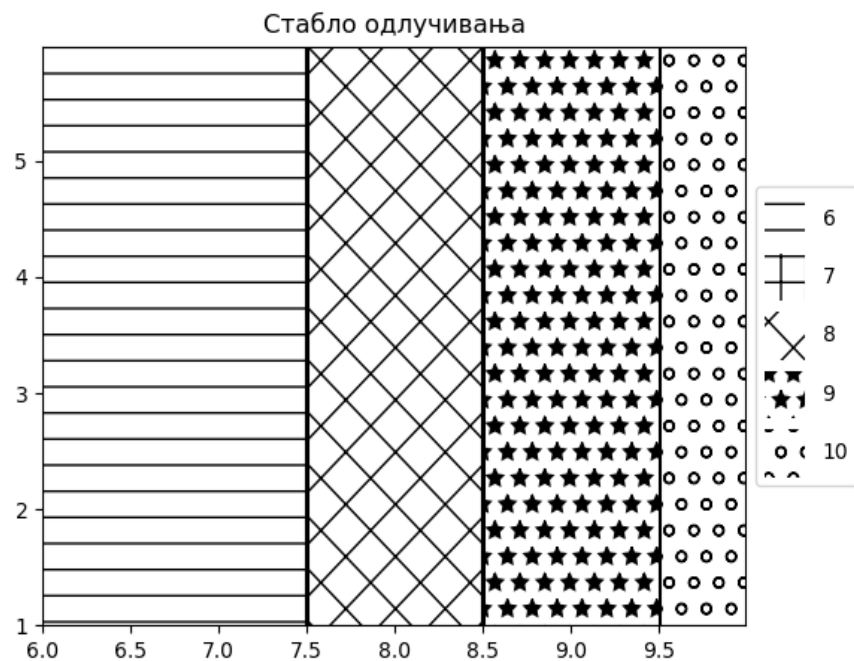
Eksperiment 4

- ▶ Analiza na podacima studenata koji su polagali ispite ASP1 i ASP2
- ▶ Podaci su izmešani i na slučajan način podeljeni u 3 skupa:
 - ▶ skup podataka za učenje
 - ▶ skup podataka za testiranje
 - ▶ skup podataka za validaciju modela
- ▶ Kategorijalni parametar koji predstavlja rok u kome je ispit polagan je pretvoren u numerički podatak, tako da su raniji rokovi manjih vrednosti, a kasniji većih vrednosti (ispit polagan u junskom roku 1, u julskom 2, itd.)

Eksperiment 4 - rezultati

- ▶ Za analizu podataka izabrana je vrednost parametra maksimalne dubine stabla 4, kao vrednost pri kojoj se javlja najbolja procena predviđanja za validacioni skup podataka
- ▶ Rezultati:
 - ▶ 60 % slučajeva gde je predviđena tačna ocena na ASP2 u odnosu na ASP1
 - ▶ 28 % slučajeva gde je predviđena ocena za jednu ocenu više ili niže od stvarno dobijene
 - ▶ 10 % slučajeva gde je predviđena ocena za dve viša ili niža
 - ▶ 2 % slučajeva gde je predviđanje preko dve ocene više ili niže

Експеримент 4 - резултати



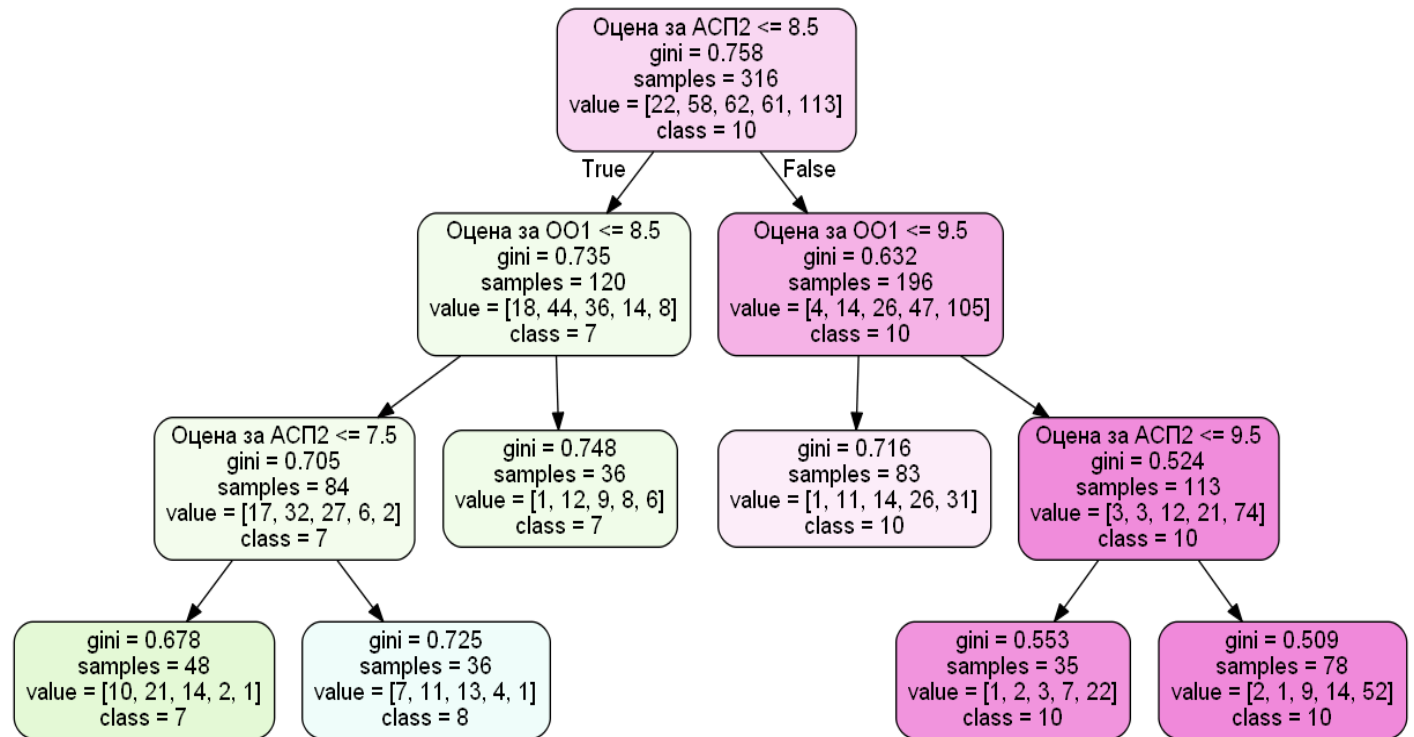
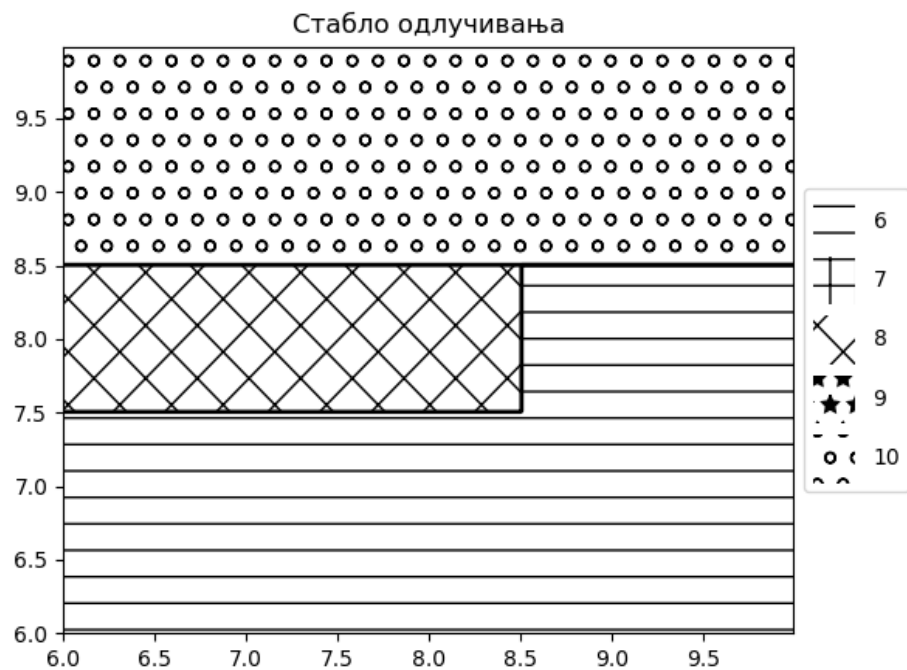
Eksperiment 5

- ▶ Analiza na podacima studenata koji su polagali ispite ASP2 i OOP1 i njihov uticaj na OS1
- ▶ Podaci su izmešani i na slučajan način podeljeni u 3 skupa:
 - ▶ skup podataka za učenje
 - ▶ skup podataka za testiranje
 - ▶ skup podataka za validaciju modela
- ▶ Kategorijalni parametri koji predstavljaju rok u kome su ispiti polagani su pretvoreni u numerički podatak, tako da su raniji rokovi manjih vrednosti, a kasniji većih vrednosti (ispit polagan u prvom roku 1, u drugom roku 2, trećem 3 itd.)

Eksperiment 5 - rezultati

- ▶ Za analizu podataka izabrana je vrednost parametra maksimalne dubine stabla 5, kao vrednost pri kojoj se javlja najbolja procena predviđanja za validacioni skup podataka
- ▶ Rezultati:
 - ▶ 45.59 % slučajeva gde je predviđena tačna ocena na OS1 u odnosu na ocene iz ASP2 i OOP1
 - ▶ 35.29 % slučajeva gde je predviđena ocena za jednu ocenu više ili niže od stvarno dobijene
 - ▶ 14.71 % slučajeva gde je predviđena ocena za dve viša ili niža
 - ▶ 4.41 % slučajeva gde je predviđanje preko dve ocene više ili niže

Експеримент 5 - резултати



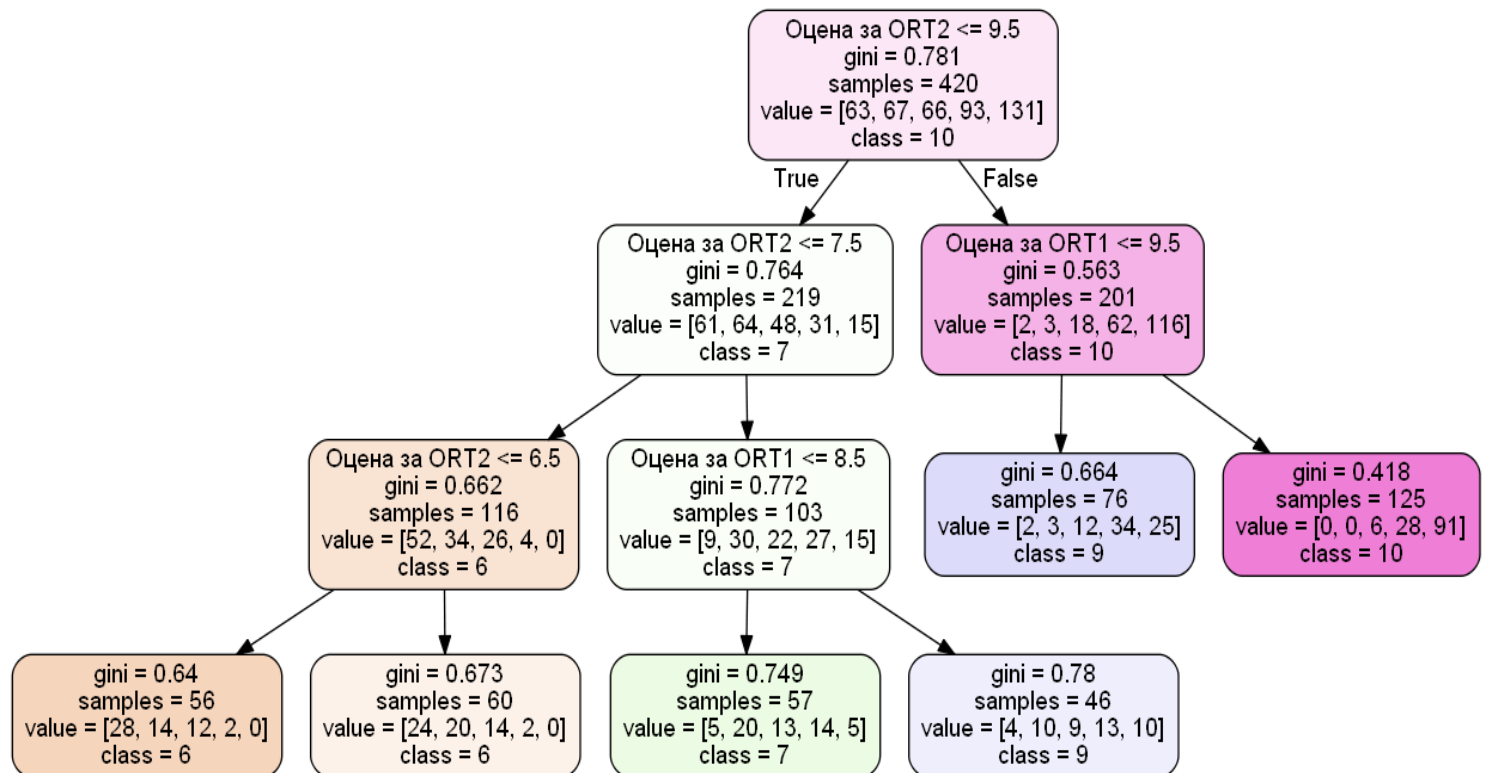
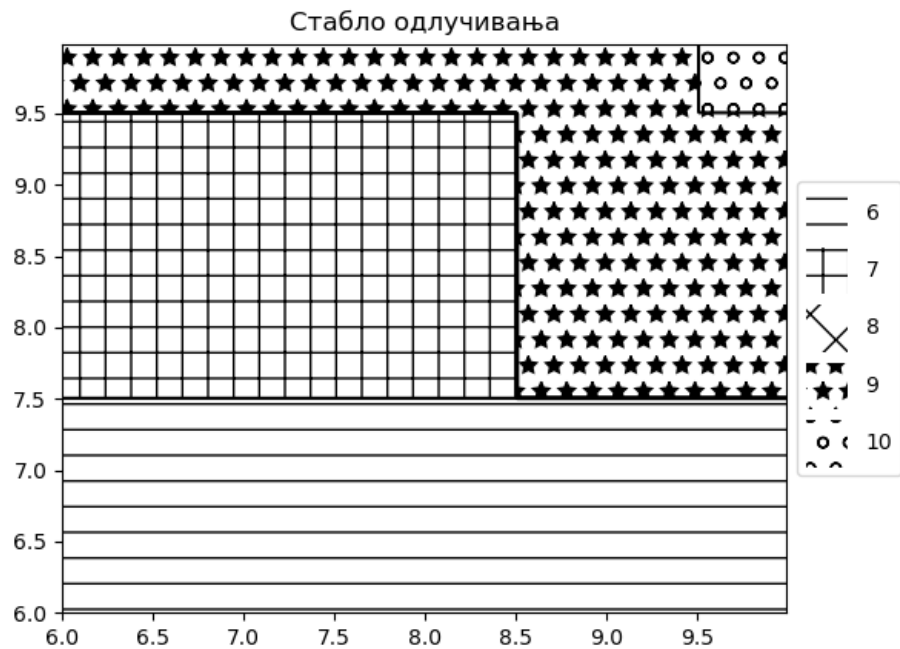
Eksperiment 6

- ▶ Analiza na podacima studenata koji su polagali ispite ORT1 i ORT 2 i njihov uticaj na AR
- ▶ Podaci su izmešani i na slučajan način podeljeni u 3 skupa:
 - ▶ skup podataka za učenje
 - ▶ skup podataka za testiranje
 - ▶ skup podataka za validaciju modela
- ▶ Kategorijalni parametri koji predstavljaju rok u kome su ispiti polagani su pretvoreni u numerički podatak, tako da su raniji rokovi manjih vrednosti, a kasniji većih vrednosti (isпит polagan u prvom roku 1, u drugom roku 2, trećem 3 itd.)

Eksperiment 6 - rezultati

- ▶ Za analizu podataka izabrana je vrednost parametra maksimalne dubine stabla 4, kao vrednost pri kojoj se javlja najbolja procena predviđanja za validacioni skup podataka
- ▶ Rezultati:
 - ▶ 52.75 % slučajeva gde je predviđena tačna ocena na AR u odnosu na ocene iz ORT1 i ORT2
 - ▶ 34.07 % slučajeva gde je predviđena ocena za jednu ocenu više ili niže od stvarno dobijene
 - ▶ 8.79 % slučajeva gde je predviđena ocena za dve viša ili niža
 - ▶ 4.39 % slučajeva gde je predviđanje preko dve ocene više ili niže

Експеримент 6 - резултати



Eksperiment 7

- ▶ Analiza na podacima studenata koji su polagali ispite P1 i OOP 2 i njihov uticaj na KDP
- ▶ Podaci su izmešani i na slučajan način podeljeni u 3 skupa:
 - ▶ skup podataka za učenje
 - ▶ skup podataka za testiranje
 - ▶ skup podataka za validaciju modela
- ▶ Kategorijalni parametri koji predstavljaju rok u kome su ispiti polagani su pretvoreni u numerički podatak, tako da su raniji rokovi manjih vrednosti, a kasniji većih vrednosti (ispit polagan u prvom roku 1, u drugom roku 2, trećem 3 itd.)

Eksperiment 7 - rezultati

- ▶ Za analizu podataka izabrana je vrednost parametra maksimalne dubine stabla 3, kao vrednost pri kojoj se javlja najbolja procena predviđanja za validacioni skup podataka
- ▶ Rezultati:
 - ▶ 38.71 % slučajeva gde je predviđena tačna ocena na KDP u odnosu na ocene iz P1 i OOP2
 - ▶ 40.32 % slučajeva gde je predviđena ocena za jednu ocenu više ili niže od stvarno dobijene
 - ▶ 9.68 % slučajeva gde je predviđena ocena za dve viša ili niža
 - ▶ 11.29 % slučajeva gde je predviđanje preko dve ocene više ili niže

Експеримент 7 - резултати

