

# NALOGA 1

Zaporniki se na otoku lahko prosto gibljejo. Do kopnega je 8 km in pogosto lahko vidimo v morju morske pse. Nekateri zaporniki so poskušali pobegniti s plavanjem ponoči, a samo nekaterim uspe doseči kopno. V tabeli so opisi nekaterih poskusov pobega.

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

Uporabi naivni Bayes z m-oceno ( $m=5$ ) in oceni verjetnost, da je bil zapornik dober plavalec, če ni dosegel kopna med polno luno ter so ga med pobegom opazili pazniki.

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

Uporabi naivni Bayes z m-oceno ( $m=5$ ) in oceni verjetnost, da je bil zapornik dober plavalec, če ni dosegel kopna med polno luno ter so ga med pobegom opazili pazniki.

$$P(\text{plavalec} = \text{"DA"} \mid \text{kopno} = \text{"NE"}, \text{luna} = \text{"DA"}, \text{pazniki} = \text{"DA"}) = ?$$

$$P(\text{plavalec} = \text{"DA"}) * \frac{P(\text{plavalec} = \text{"DA"} \mid \text{kopno} = \text{"NE"})}{P(\text{plavalec} = \text{"DA"})} * \frac{P(\text{plavalec} = \text{"DA"} \mid \text{luna} = \text{"DA"})}{P(\text{plavalec} = \text{"DA"})} * \frac{P(\text{plavalec} = \text{"DA"} \mid \text{pazniki} = \text{"DA"})}{P(\text{plavalec} = \text{"DA"})}$$

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"}) = \frac{N_k + 1}{N + n_0} =$$



apriorna verjetnost  
razreda "DA"

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"}) = \frac{N_k + 1}{N + n_0} = \frac{6 + 1}{6 + 6}$$



število učnih primerov

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"}) = \frac{N_k + 1}{N + n_0} = \frac{\quad}{6 + 2}$$



število razredov

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

število primerov iz razreda "DA"



$$P(\text{plavalec} = \text{"DA"}) = \frac{N_k + 1}{N + n_0} = \frac{4 + 1}{6 + 2}$$

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"}) = \frac{N_k + 1}{N + n_0} = \frac{4 + 1}{6 + 2} = 0.625$$

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"} \mid \text{kopno} = \text{"NE"}) = \frac{N_{k,i} + m * P_k}{N_i + m} =$$



pogojna verjetnost  
razreda "DA" pri vrednosti  
atributa kopno = "NE"



# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"} \mid \text{kopno} = \text{"NE"}) = \frac{N_{k,i+m} * P_k}{N_i + m} = \frac{1}{4}$$



število primerov z vrednostjo  
atributa kopno = "NE"

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"} \mid \text{kopno} = \text{"NE"}) = \frac{N_{k,i} + m * P_k}{N_i + m} = \frac{4}{4 + 5}$$



parameter m-ocene  
(podan v besedilu naloge)

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

število primerov iz razreda "DA" z vrednostjo atributa kopno="NE"



$$P(\text{plavalec} = \text{"DA"} | \text{kopno} = \text{"NE"}) = \frac{N_{k,i} + m * P_k}{N_i + m} = \frac{3 + 5}{4 + 5}$$

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

apriorna verjetnost razreda "DA"



$$P(\text{plavalec} = \text{"DA"} | \text{kopno} = \text{"NE"}) = \frac{N_{k,i} + m * P_k}{N_i + m} = \frac{3 + 5 * 0.625}{4 + 5} =$$

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"} \mid \text{kopno} = \text{"NE"}) = \frac{N_{k,i} + m * P_k}{N_i + m} = \frac{3 + 5 * 0.625}{4 + 5} = 0.6805$$

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"} \mid \text{luna} = \text{"DA"}) = \frac{N_{k,i} + m * P_k}{N_i + m} = \frac{2 + 5 * 0.625}{4 + 5} = 0.5694$$

# NALOGA 1

Dober plavalec	Dežuje	Polna luna	Podnevi videni morski psi	So ga opazili pazniki	Je dosegel kopno
DA	DA	DA	DA	DA	DA
NE	NE	DA	NE	NE	DA
DA	NE	NE	NE	DA	NE
NE	NE	DA	NE	DA	NE
DA	NE	NE	DA	NE	NE
DA	DA	DA	DA	NE	NE

$$P(\text{plavalec} = \text{"DA"} \mid \text{pazniki} = \text{"DA"}) = \frac{N_{k,i} + m * P_k}{N_i + m} = \frac{2 + 5 * 0.625}{3 + 5} = 0.6406$$

# NALOGA 1

$$P(\text{plavalec} = \text{"DA"}) * \frac{P(\text{plavalec} = \text{"DA"} | \text{kopno} = \text{"NE"})}{P(\text{plavalec} = \text{"DA"})} * \frac{P(\text{plavalec} = \text{"DA"} | \text{luna} = \text{"DA"})}{P(\text{plavalec} = \text{"DA"})} * \frac{P(\text{plavalec} = \text{"DA"} | \text{pazniki} = \text{"DA"})}{P(\text{plavalec} = \text{"DA"})} =$$

$$= 0.625 * \frac{0.6805}{0.625} * \frac{0.5694}{0.625} * \frac{0.6406}{0.625} = 0.6354$$



To ni  
verjetnost!



# NALOGA 1

$$\begin{aligned} P(\text{plavalec} = \text{"DA"}) * \frac{P(\text{plavalec} = \text{"DA"} | \text{kopno} = \text{"NE"})}{P(\text{plavalec} = \text{"DA"})} * \frac{P(\text{plavalec} = \text{"DA"} | \text{luna} = \text{"DA"})}{P(\text{plavalec} = \text{"DA"})} * \frac{P(\text{plavalec} = \text{"DA"} | \text{pazniki} = \text{"DA"})}{P(\text{plavalec} = \text{"DA"})} = \\ = 0.625 * \frac{0.6805}{0.625} * \frac{0.5694}{0.625} * \frac{0.6406}{0.625} = 0.6354 \end{aligned}$$

$$\begin{aligned} P(\text{plavalec} = \text{"NE"}) * \frac{P(\text{plavalec} = \text{"NE"} | \text{kopno} = \text{"NE"})}{P(\text{plavalec} = \text{"NE"})} * \frac{P(\text{plavalec} = \text{"NE"} | \text{luna} = \text{"DA"})}{P(\text{plavalec} = \text{"NE"})} * \frac{P(\text{plavalec} = \text{"NE"} | \text{pazniki} = \text{"DA"})}{P(\text{plavalec} = \text{"NE"})} = \\ = 0.375 * \frac{0.3194}{0.375} * \frac{0.4306}{0.375} * \frac{0.3594}{0.375} = 0.3515 \end{aligned}$$

# NALOGA 1

---

$$P(\text{plavalec} = \text{"DA"} \mid \text{kopno} = \text{"NE"}, \text{luna} = \text{"DA"}, \text{pazniki} = \text{"DA"}) = \frac{0.6354}{0.6354 + 0.3515} = 0.6438$$

$$P(\text{plavalec} = \text{"NE"} \mid \text{kopno} = \text{"NE"}, \text{luna} = \text{"DA"}, \text{pazniki} = \text{"DA"}) = \frac{0.3515}{0.6354 + 0.3515} = 0.3562$$

# NALOGA 2

Ko je novopečeni gobar prišel iz gozda, je nesel polno košaro gob, čeprav ni vedel, ali so užitne ali ne. Prosil je izkušenega gobarja, naj jih razdeli v tri košare: v prvo vse užitne, v drugo vse strupene (ki so seveda neužitne) in v tretjo nestrupene a neužitne gobe. Novopečeni gobar je sestavil učno množico tako, da je za vsako gobo zapisal barvo klobuka in barvo beta, nato pa je preštel gobe iz vsake podskupine. Za vsako košaro je dobil po eno tabelo:

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

Kakšna je verjetnost, da je goba z rjavim betom in rdečim klobukom užitna, če naivni Bayes uporablja m-oceno in je  $m = 10$ ?

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

Kakšna je verjetnost, da je goba z rjavim betom in rdečim klobukom užitna, če naivni Bayes uporablja m-oceno in je  $m = 10$ ?

$$P(\text{goba} = \text{"užitna"} \mid \text{bet} = \text{"rjav"}, \text{klobuk} = \text{"rdeč"}) = ?$$

$$P(\text{goba} = \text{"užitna"}) * \frac{P(\text{goba} = \text{"užitna"} \mid \text{bet} = \text{"rjav"})}{P(\text{goba} = \text{"užitna"})} * \frac{P(\text{goba} = \text{"užitna"} \mid \text{klobuk} = \text{"rdeč"})}{P(\text{goba} = \text{"užitna"})}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"}) = \frac{\quad}{170 + \quad}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"}) = \frac{\quad}{170 + 3}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"}) = \frac{90 + 3}{170 + 3}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"}) = \frac{90 + 1}{170 + 3} = 0.526$$



# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"} \mid \text{bet} = \text{"rjav"}) = \frac{5}{55 + 5}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"} \mid \text{bet} = \text{"rjav"}) = \frac{5}{55 + 10}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"} \mid \text{bet} = \text{"rjav"}) = \frac{30 + 10}{55 + 10}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"} \mid \text{bet} = \text{"rjav"}) = \frac{30 + 10 * 0.526}{55 + 10} = 0.5425$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"} \mid \text{klobuk} = \text{"rdeč"}) = \frac{10}{45 + 10}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"} \mid \text{klobuk} = \text{"rdeč"}) = \frac{45}{45 + 10}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"} \mid \text{klobuk} = \text{"rdeč"}) = \frac{10 + 5}{45 + 10}$$

# NALOGA 2

<b>UŽITNE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	50	10
Rjav bet	10	15	5

<b>NEUŽITNE STRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	30	5	20
Rjav bet	0	10	0

<b>NEUŽITNE NESTRUPENE</b>	Rdeč klobuk	Rjav klobuk	Bel klobuk
Bel bet	0	0	0
Rjav bet	5	10	0

$$P(\text{goba} = \text{"užitna"} \mid \text{klobuk} = \text{"rdeč"}) = \frac{10 + 10 * 0.526}{45 + 10} = 0.2774$$



# NALOGA 2

$$P(\text{goba} = \text{"užitna"}) * \frac{P(\text{goba} = \text{"užitna"} | \text{bet} = \text{"rjav"})}{P(\text{goba} = \text{"užitna"})} * \frac{P(\text{goba} = \text{"užitna"} | \text{klobuk} = \text{"rdeč"})}{P(\text{goba} = \text{"užitna"})} =$$

$$= 0.526 * \frac{0.5425}{0.526} * \frac{0.2774}{0.526} = 0.2861$$



To ni  
verjetnost!

# NALOGA 2

$$P(\text{goba} = \text{"užitna"}) * \frac{P(\text{goba} = \text{"užitna"} | \text{bet} = \text{"rjav"})}{P(\text{goba} = \text{"užitna"})} * \frac{P(\text{goba} = \text{"užitna"} | \text{klobuk} = \text{"rdeč"})}{P(\text{goba} = \text{"užitna"})} = 0.2861$$

$$P(\text{goba} = \text{"neužitna"}) * \frac{P(\text{goba} = \text{"neužitna"} | \text{bet} = \text{"rjav"})}{P(\text{goba} = \text{"neužitna"})} * \frac{P(\text{goba} = \text{"neužitna"} | \text{klobuk} = \text{"rdeč"})}{P(\text{goba} = \text{"neužitna"})} = 0.0925 * \frac{0.245}{0.0925} * \frac{0.1077}{0.0925} = 0.2853$$

$$P(\text{goba} = \text{"strupena"}) * \frac{P(\text{goba} = \text{"strupena"} | \text{bet} = \text{"rjav"})}{P(\text{goba} = \text{"strupena"})} * \frac{P(\text{goba} = \text{"strupena"} | \text{klobuk} = \text{"rdeč"})}{P(\text{goba} = \text{"strupena"})} = 0.3815 * \frac{0.2125}{0.3815} * \frac{0.6148}{0.3815} = 0.3425$$

$$0.2861 + 0.2853 + 0.3425 = 0.9139$$

# NALOGA 2

---

$$P(\text{goba} = \text{"užitna"} \mid \text{bet} = \text{"rjav"}, \text{klobuk} = \text{"rdeč"}) = \frac{0.2861}{0.9139} = 0.3131$$

$$P(\text{goba} = \text{"neužitna"} \mid \text{bet} = \text{"rjav"}, \text{klobuk} = \text{"rdeč"}) = \frac{0.2853}{0.9139} = 0.3122$$

$$P(\text{goba} = \text{"strupena"} \mid \text{bet} = \text{"rjav"}, \text{klobuk} = \text{"rdeč"}) = \frac{0.3425}{0.9139} = 0.3748$$

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja. Naj bo to naša učna množica (2 atributa: vrsta cekina, simbol gusarja; 2 razreda: pravi/ponarejen; 100 učnih primerov).

- a) Nariši odločitveno drevo, ki vsebuje celotno zgoraj opisano informacijo, torej v vsakem listu mora biti napisano število cekinov, ki ustreza temu listu in verjetnost pravilnega odgovora. Pri gradnji drevesa za izbiro atributa uporabi **informacijski prispevek**,
- b) Z odločitvenim drevesom klasificiraj 101. cekin, ki ga je potapljač kasneje našel na obali, kjer ga je bil izgubil. Cekin je bil srebrn z vtisnjenim simbolom gusarja.

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi	
srebrn	DA		
	NE		
bronast	DA		
	NE		
zlat	DA		
	NE		

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, **da jih je 50 srebrnih, 30 bronastih in 20 zlatih**, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi	
srebrn 50	DA		
	NE		
bronast 30	DA		
	NE		
zlat 20	DA		
	NE		

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. **Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja**, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	
	NE 25	
bronast 30	DA	
	NE	
zlat 20	DA	
	NE	

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, **pri bronastih tretjina** in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	
	NE 25	
bronast 30	DA 10	
	NE 20	
zlat 20	DA	
	NE	



# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in **pri zlatih tri četrtine**. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	
	NE 25	
bronast 30	DA 10	
	NE 20	
zlat 20	DA 15	
	NE 5	

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. **Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja**, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	10
	NE 25	0
bronast 30	DA 10	
	NE 20	
zlat 20	DA 15	
	NE 5	

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. **Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja**, od bronastih ni bil nobeden ponarejen in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	10 15
	NE 25	0 25
bronast 30	DA 10	
	NE 20	
zlat 20	DA 15	
	NE 5	

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, **od bronastih ni bil nobeden ponarejen** in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	10 15
	NE 25	0 25
bronast 30	DA 10	0
	NE 20	0
zlat 20	DA 15	
	NE 5	

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, **od bronastih ni bil nobeden ponarejen** in od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	10 15
	NE 25	0 25
bronast 30	DA 10	0 10
	NE 20	0 20
zlat 20	DA 15	
	NE 5	

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in **od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.**

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	10 15
	NE 25	0 25
bronast 30	DA 10	0 10
	NE 20	0 20
zlat 20	DA 15	10
	NE 5	0

# NALOGA 3

V morju je potapljač naletel na razbitino gusarske ladje in v notranjosti je bil zaboj poln cekinov. Ker je bil zaboj pretežak, je vzel s seboj samo 101 naključno izbranih cekinov. Na obali je pregledal cekine in ugotovil, da jih je 50 srebrnih, 30 bronastih in 20 zlatih, enega pa je izgubil. Pri srebrnih je ugotovil, da ima polovica vtisnjen simbol Črnega gusarja, pri bronastih tretjina in pri zlatih tri četrtine. Ker je posumil, da so nekateri ponarejeni, jih je še stehtal. Od srebrnih jih je bila petina lažjih (ponarejenih) in vsi so imeli vtisnjen simbol gusarja, od bronastih ni bil nobeden ponarejen in **od zlatih polovica in vsi so imeli vtisnjen simbol gusarja.**

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	10 15
	NE 25	0 25
bronast 30	DA 10	0 10
	NE 20	0 20
zlat 20	DA 15	10 5
	NE 5	0 5

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

Niso vsi primeri v istem razredu, zato poiščemo najbolj ustrezno razbitje učne množice



# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Razred}=\text{"ponaredek"}) = 20/100$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$P(\text{Razred}=\text{"ponaredek"}) = 20/100$

$P(\text{Razred}=\text{"pravi"}) = 80/100$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Razred}=\text{"ponaredek"}) = 20/100$$

$$P(\text{Razred}=\text{"pravi"}) = 80/100$$

$$H(\text{Razred}) = -(20/100 * \log_2(20/100) + 80/100 * \log_2(80/100)) = \mathbf{0.722}$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"srebrn"}) = 50/100$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"srebrn"}) = 50/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Vrsta}=\text{"srebrn"}) = 10/50$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"srebrn"}) = 50/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Vrsta}=\text{"srebrn"}) = 10/50$$

$$P(\text{Razred}=\text{"pravi"} \mid \text{Vrsta}=\text{"srebrn"}) = 40/50$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"bronast"}) = 30/100$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"bronast"}) = 30/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Vrsta}=\text{"bronast"}) = 0/30$$



# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"bronast"}) = 30/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Vrsta}=\text{"bronast"}) = 0/30$$

$$P(\text{Razred}=\text{"pravi"} \mid \text{Vrsta}=\text{"bronast"}) = 30/30$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"zlat"}) = 20/100$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"zlat"}) = 20/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Vrsta}=\text{"zlat"}) = 10/20$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Vrsta}=\text{"zlat"}) = 20/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Vrsta}=\text{"zlat"}) = 10/20$$

$$P(\text{Razred}=\text{"pravi"} \mid \text{Vrsta}=\text{"zlat"}) = 10/20$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$H(\text{Razred} | \text{Vrsta}) = -(50/100 * (10/50 * \log_2(10/50) + 40/50 * \log_2(40/50)) + 30/100 * (0/30 * \log_2(0/30) + 30/30 * \log_2(30/30)) + 20/100 * (10/20 * \log_2(10/20) + 10/20 * \log_2(10/20))) = 0.561$$

# NALOGA 3

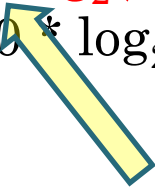
$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$H(\text{Razred} | \text{Vrsta}) = -(50/100 * (10/50 * \log_2(10/50) + 40/50 * \log_2(40/50)) + 30/100 * (0/30 * \log_2(0/30) + 30/30 * \log_2(30/30)) + 20/100 * (10/20 * \log_2(10/20) + 10/20 * \log_2(10/20))) = 0.561$$



$0 * \log_2(0) = 0$ $1 * \log_2(1) = 0$
--

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$H(\text{Razred} | \text{Vrsta}) = -(50/100 * (10/50 * \log_2(10/50) + 40/50 * \log_2(40/50)) + 30/100 * (0/30 * \log_2(0/30) + 30/30 * \log_2(30/30)) + 20/100 * (10/20 * \log_2(10/20) + 10/20 * \log_2(10/20))) = 0.561$$

$$\text{InfGain}(\text{Vrsta}) = 0.722 - 0.561 = \mathbf{0.161}$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Simbol}="DA") = 50/100$$



# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$P(\text{Simbol}=\text{"DA"}) = 50/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Simbol}=\text{"DA"}) = 20/50$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta	Simbol	Razred ponaredek / pravi
srebrn 50	DA 25	10 15
	NE 25	0 25
bronast 30	DA 10	0 10
	NE 20	0 20
zlat 20	DA 15	10 5
	NE 5	0 5

$$P(\text{Simbol}=\text{"DA"}) = 50/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Simbol}=\text{"DA"}) = 20/50$$

$$P(\text{Razred}=\text{"pravi"} \mid \text{Simbol}=\text{"DA"}) = 30/50$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	<b>25</b>	0	25
bronast	30	DA	10	0	10
		NE	<b>20</b>	0	20
zlat	20	DA	15	10	5
		NE	<b>5</b>	0	5

$$P(\text{Simbol}=\text{"NE"}) = 50/100$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	<b>25</b>	<b>0</b>	25
bronast	30	DA	10	0	10
		NE	<b>20</b>	<b>0</b>	20
zlat	20	DA	15	10	5
		NE	<b>5</b>	<b>0</b>	5

$$P(\text{Simbol}=\text{"NE"}) = 50/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Simbol}=\text{"NE"}) = \mathbf{0/50}$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	<b>25</b>	0	<b>25</b>
bronast	30	DA	10	0	10
		NE	<b>20</b>	0	<b>20</b>
zlat	20	DA	15	10	5
		NE	<b>5</b>	0	<b>5</b>

$$P(\text{Simbol}=\text{"NE"}) = 50/100$$

$$P(\text{Razred}=\text{"ponaredek"} \mid \text{Simbol}=\text{"NE"}) = 0/50$$

$$P(\text{Razred}=\text{"pravi"} \mid \text{Simbol}=\text{"NE"}) = \mathbf{50/50}$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

$$H(\text{Razred} | \text{Simbol}) = -(50/100 * (20/50 * \log_2(20/50) + 30/50 * \log_2(30/50)) + 50/100 * (0/50 * \log_2(0/50) + 50/50 * \log_2(50/50))) = 0.485$$

# NALOGA 3

$$H_R = -\sum_k p_k \log p_k$$

$$H_{R|A} = -\sum_j p_j \sum_k p_{k|j} \log p_{k|j}$$

$$\text{InfGain}(A) = H_R - H_{R|A}$$

Vrsta		Simbol		Razred ponaredek / pravi	
srebrn	50	DA	25	10	15
		NE	25	0	25
bronast	30	DA	10	0	10
		NE	20	0	20
zlat	20	DA	15	10	5
		NE	5	0	5

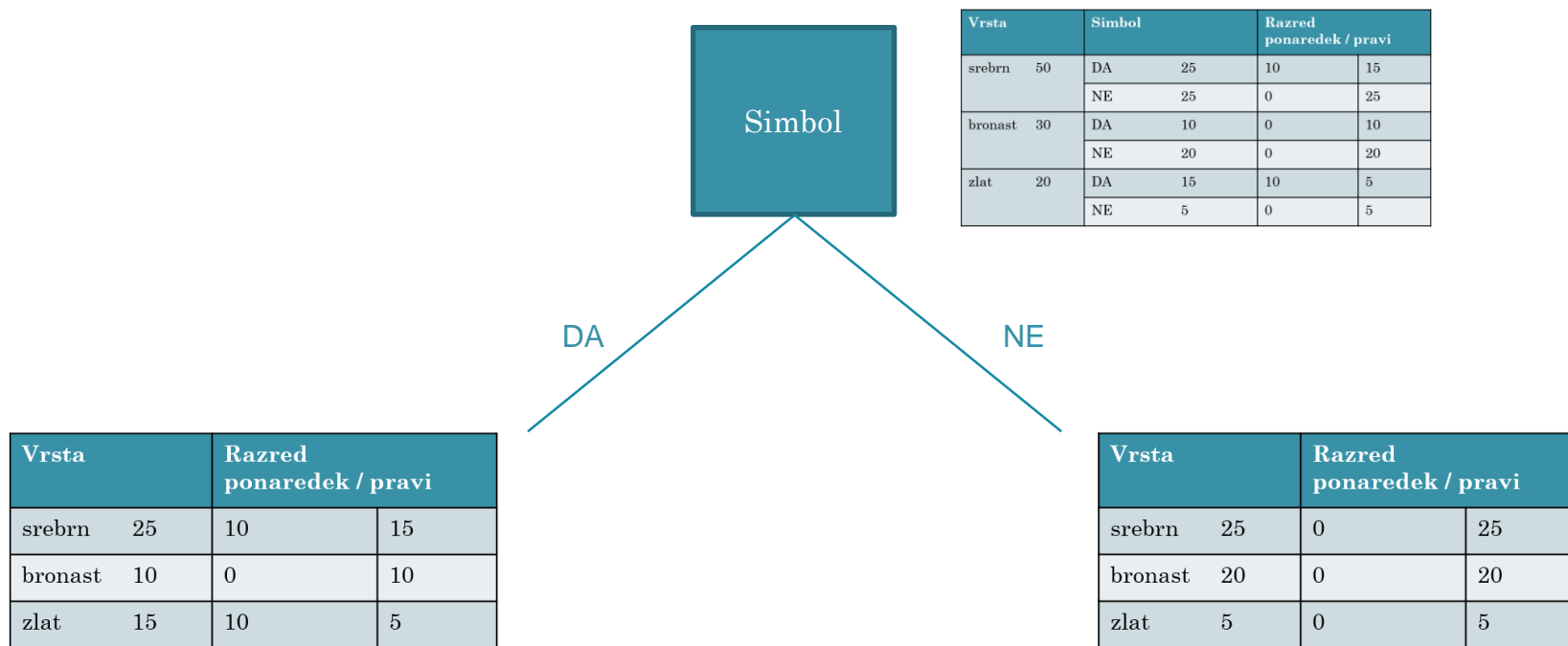
$$H(\text{Razred} | \text{Simbol}) = -(50/100 * (20/50 * \log_2(20/50) + 30/50 * \log_2(30/50)) + 50/100 * (0/50 * \log_2(0/50) + 50/50 * \log_2(50/50))) = 0.485$$

$$\text{InfGain}(\text{Simbol}) = 0.722 - 0.485 = \mathbf{0.237}$$

# NALOGA 3

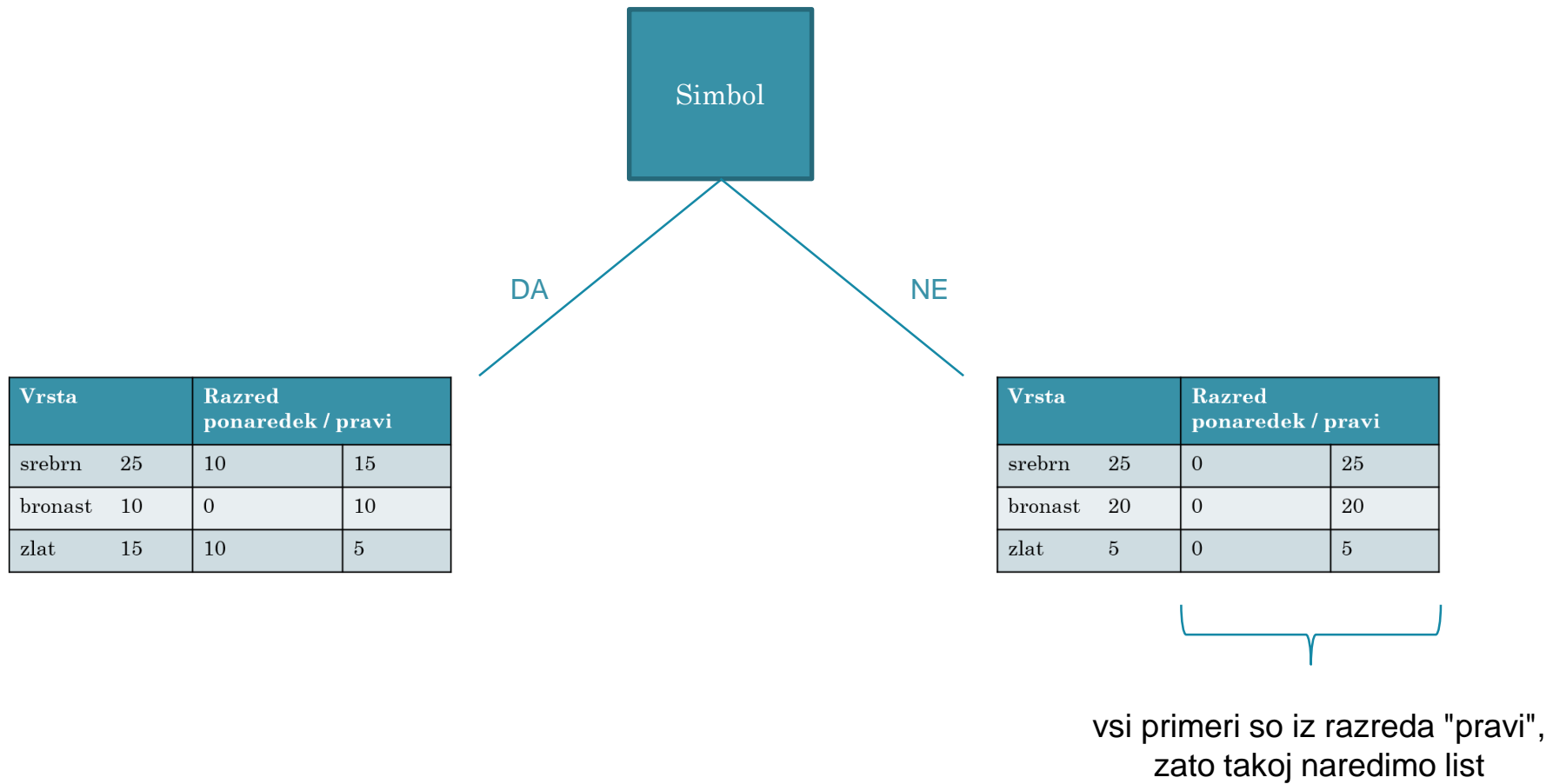
$$\text{InfGain}(\text{Vrsta}) = 0.722 - 0.561 = 0.161$$

$$\text{InfGain}(\text{Simbol}) = 0.722 - 0.485 = 0.237 \leftarrow \text{izberemo}$$

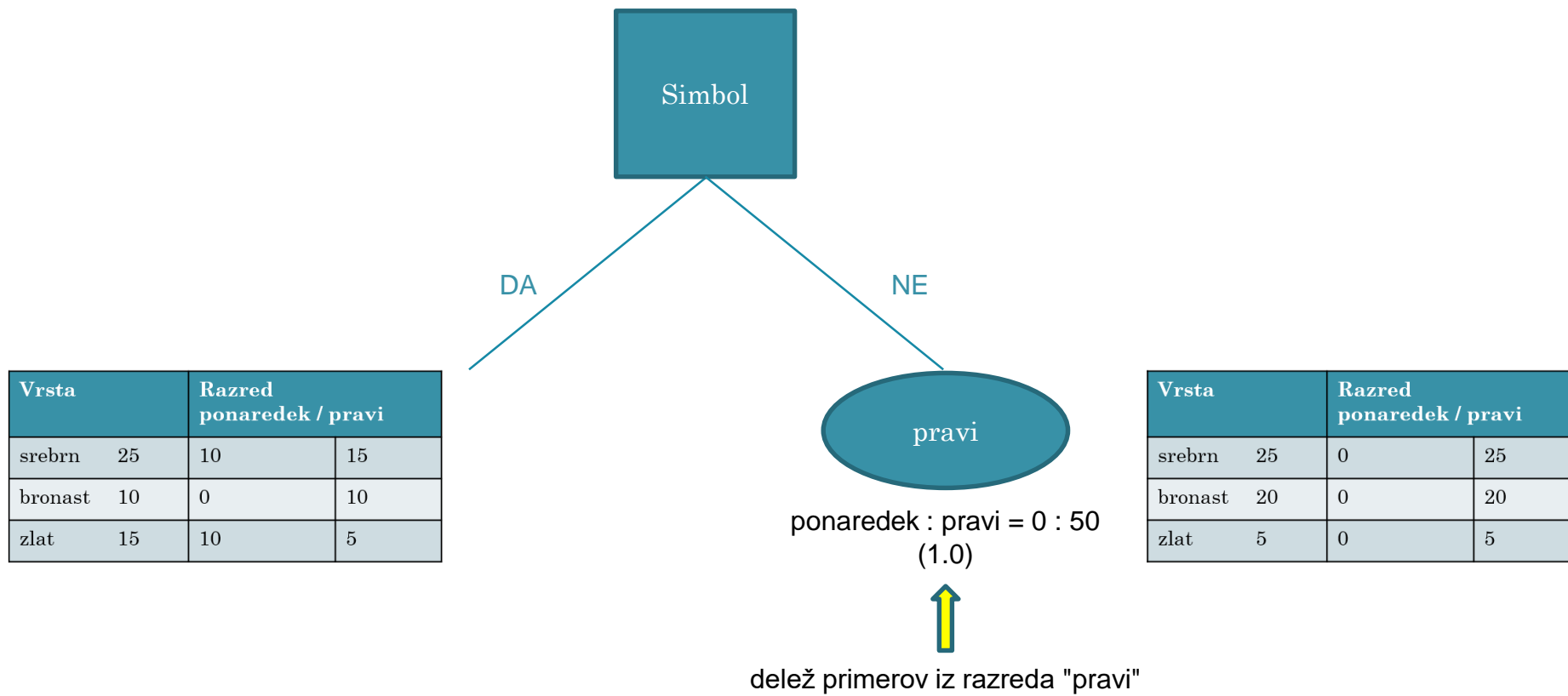




# NALOGA 3



# NALOGA 3

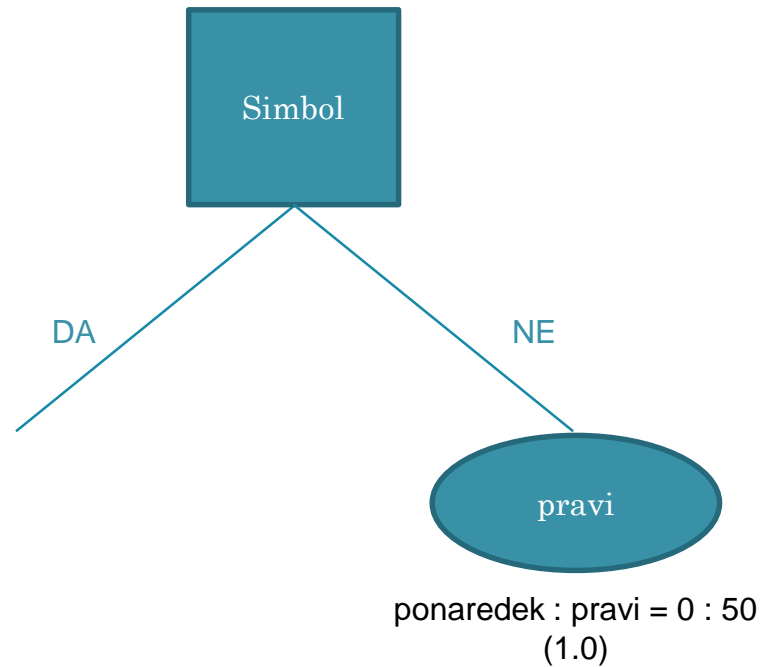


# NALOGA 3

Vrsta		Razred ponaredek / pravi	
srebrn	25	10	15
bronast	10	0	10
zlat	15	10	5

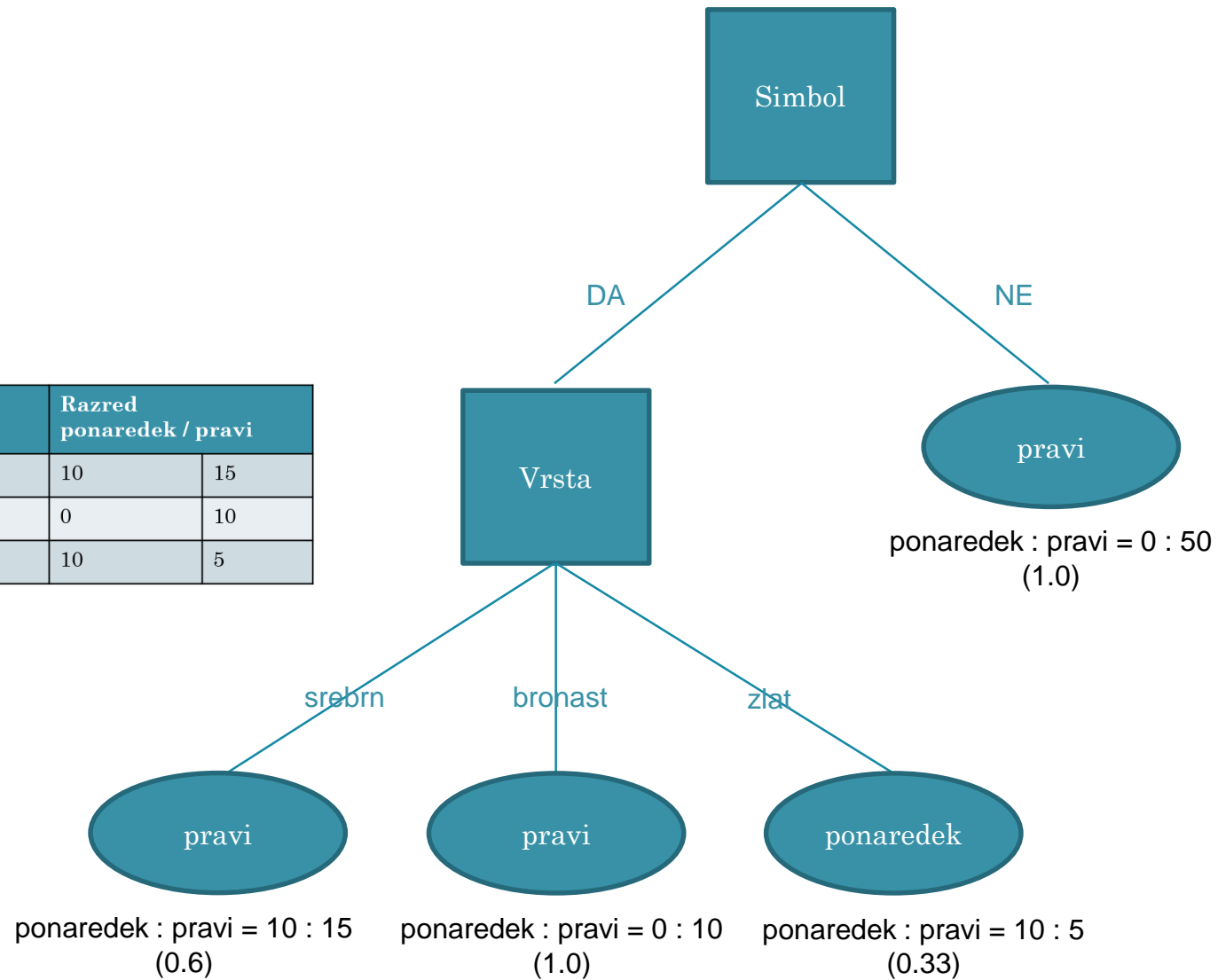


ostal je en sam atribut ("Vrsta"),  
zato takoj naredimo notranje  
vozlišče



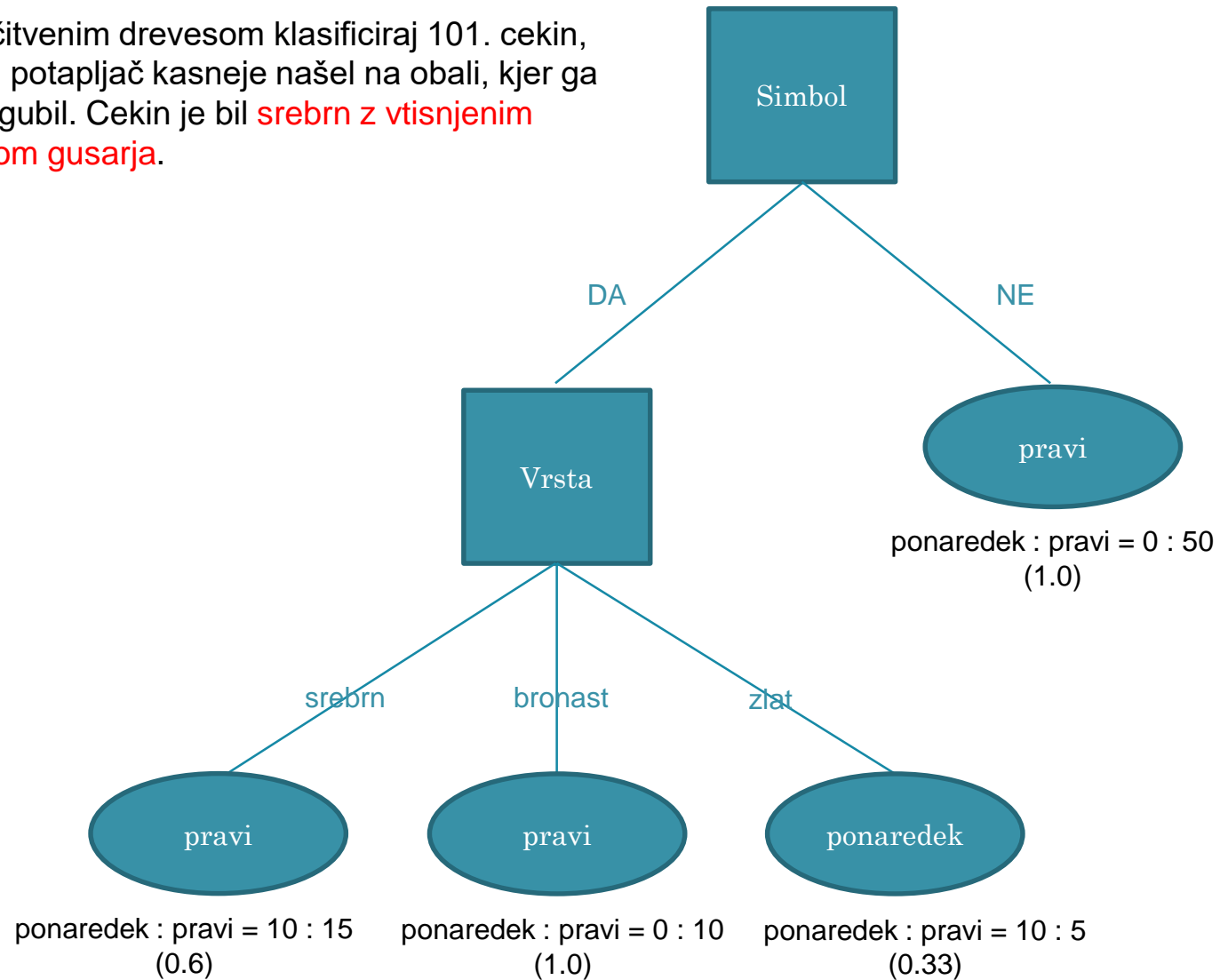
# NALOGA 3

Vrsta		Razred ponaredek / pravi	
srebrn	25	10	15
bronast	10	0	10
zlat	15	10	5



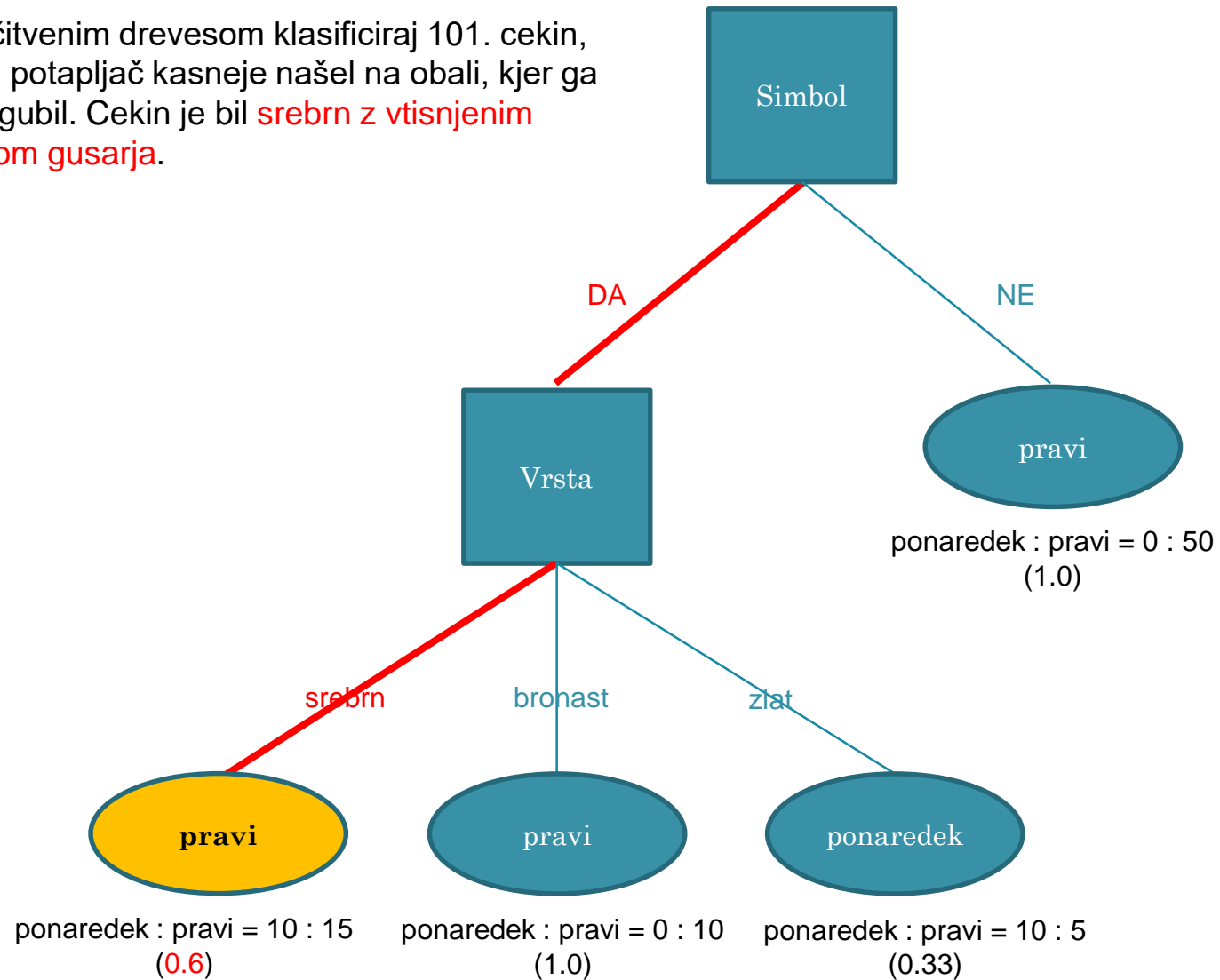
# NALOGA 3

Z odločitvenim drevesom klasificiraj 101. cekin, ki ga je potapljač kasneje našel na obali, kjer ga je bil izgubil. Cekin je bil **srebrn z vtisnjenim simbolom gusarja**.



# NALOGA 3

Z odločitvenim drevesom klasificiraj 101. cekin, ki ga je potapljač kasneje našel na obali, kjer ga je bil izgubil. Cekin je bil **srebrn z vtisnjnim simbolom gusarja**.



# NALOGA 4

Da bi zagovili likvidnost banke, so začeli zbirati podatke o posojilojemalcih in o tem, ali je bilo posojilo vrnjeno. Zbrali so podatke v spodnji tabeli.

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

Direktor banke se je odločil, da bo napovedoval, ali bo posojilo vrnjeno, kar s pomočjo odločitvenega drevesa, ki ga je zgradil z uporabo **gini-indeksa** iz zgornje podatkovne baze o posojilojemalcih.

- Nariši direktorjevo odločitveno drevo.
- Kakšna je napoved za direktorjevo ženo, ki ni v kazenskem postopku, nikoli ni imela kreditov in bi si sposodila 20 000 evrov?

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE



# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

$$P(\text{Vrnjeno}="NE") = 4/8$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

$$P(\text{Vrnjeno}="NE") = 4/8$$

$$P(\text{Spol}="MOŠKI") = 6/8$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

$$P(\text{Vrnjeno}="NE") = 4/8$$

$$P(\text{Spol}="MOŠKI") = 6/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="MOŠKI") = 3/6$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

$$P(\text{Vrnjeno}="NE") = 4/8$$

$$P(\text{Spol}="MOŠKI") = 6/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="MOŠKI") = 3/6$$

$$P(\text{Vrnjeno}="NE" | \text{Spol}="MOŠKI") = 3/6$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

$$P(\text{Vrnjeno}="NE") = 4/8$$

$$P(\text{Spol}="MOŠKI") = 6/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="MOŠKI") = 3/6$$

$$P(\text{Vrnjeno}="NE" | \text{Spol}="MOŠKI") = 3/6$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$P(\text{Spol}="ŽENSKA") = 2/8$$

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

$$P(\text{Vrnjeno}="NE") = 4/8$$

$$P(\text{Spol}="MOŠKI") = 6/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="MOŠKI") = 3/6$$

$$P(\text{Vrnjeno}="NE" | \text{Spol}="MOŠKI") = 3/6$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$P(\text{Spol}="ŽENSKA") = 2/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="ŽENSKA") = 1/2$$

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

$$P(\text{Vrnjeno}="NE") = 4/8$$

$$P(\text{Spol}="MOŠKI") = 6/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="MOŠKI") = 3/6$$

$$P(\text{Vrnjeno}="NE" | \text{Spol}="MOŠKI") = 3/6$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$P(\text{Spol}="ŽENSKA") = 2/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="ŽENSKA") = 1/2$$

$$P(\text{Vrnjeno}="NE" | \text{Spol}="ŽENSKA") = 1/2$$

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

$$P(\text{Vrnjeno}="DA") = 4/8$$

$$P(\text{Vrnjeno}="NE") = 4/8$$

$$P(\text{Spol}="MOŠKI") = 6/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="MOŠKI") = 3/6$$

$$P(\text{Vrnjeno}="NE" | \text{Spol}="MOŠKI") = 3/6$$

$$Gini(\text{Spol}) = \frac{6}{8} * \left( \left( \frac{3}{6} \right)^2 + \left( \frac{3}{6} \right)^2 \right) + \frac{2}{8} * \left( \left( \frac{1}{2} \right)^2 + \left( \frac{1}{2} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$P(\text{Spol}="ŽENSKA") = 2/8$$

$$P(\text{Vrnjeno}="DA" | \text{Spol}="ŽENSKA") = 1/2$$

$$P(\text{Vrnjeno}="NE" | \text{Spol}="ŽENSKA") = 1/2$$



# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Zakreditiran}) = \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) + \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0$$

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Zakreditiran}) = \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) + \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0$$

↑  
P(Zakreditiran = "DA")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Zakreditiran}) = \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) + \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0$$

P(Vrnjeno="DA" | Zakreditiran = "DA")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Zakreditiran}) = \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) + \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0$$



P(Vrnjeno="NE" | Zakreditiran = "DA")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Zakreditiran}) = \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) + \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0$$



P(Zakreditiran = "NE")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Zakreditiran}) = \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) + \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0$$

P(Vrnjeno="DA" | Zakreditiran = "NE")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Zakreditiran}) = \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) + \frac{4}{8} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0$$

P(Vrnjeno="NE" | Zakreditiran = "NE")





# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Znesek}) = \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) + \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Znesek}) = \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) + \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

↑  
P(Znesek = "DA")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Znesek}) = \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) + \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

P(Vrnjeno="DA" | Znesek = "DA")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Znesek}) = \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) + \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

P(Vrnjeno="NE" | Znesek = "DA")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Znesek}) = \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) + \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

↑  
P(Znesek = "NE")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Znesek}) = \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) + \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

↑  
P(Vrnjeno="DA" | Znesek = "NE")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Znesek}) = \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) + \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

↑  
P(Vrnjeno="NE" | Znesek = "NE")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Znesek}) = \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) + \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

↑  
P(Vrnjeno="DA")

↑  
P(Vrnjeno="NE")



# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Kazenski}) = \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) + \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Kazenski}) = \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) + \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

P(Kazenski = "DA")

P(Kazenski = "NE")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Kazenski}) = \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) + \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

P(Vrnjeno="DA" | Kazenski = "DA")

P(Vrnjeno="DA" | Kazenski = "NE")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Kazenski}) = \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) + \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

P(Vrnjeno="NE" | Kazenski = "DA")

P(Vrnjeno="NE" | Kazenski = "NE")

# NALOGA 4

$$Gini(A) = \sum_j p_{.j} \sum_k p_{k|j}^2 - \sum_k p_k^2.$$

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

$$Gini(\text{Kazenski}) = \frac{3}{8} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) + \frac{5}{8} * \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) - \left( \left( \frac{4}{8} \right)^2 + \left( \frac{4}{8} \right)^2 \right) = 0.033$$

↑  
P(Vrnjeno="DA")

↑  
P(Vrnjeno="NE")

# NALOGA 4

Gini(Spol) = 0

Gini(Zakreditiran) = 0

Gini(Znesek) = 0.033 ← izberemo

Gini(Kazenski) = 0.033

Znesek

Posojilo vrnjeno	Spol	Trenutno zakreditiran	Znesek > 100K	V kazenskem postopku
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	NE	DA	NE
NE	ŽENSKA	DA	NE	NE
DA	ŽENSKA	DA	DA	NE
NE	MOŠKI	DA	NE	NE
DA	MOŠKI	NE	NE	DA
NE	MOŠKI	NE	DA	DA
DA	MOŠKI	DA	DA	NE

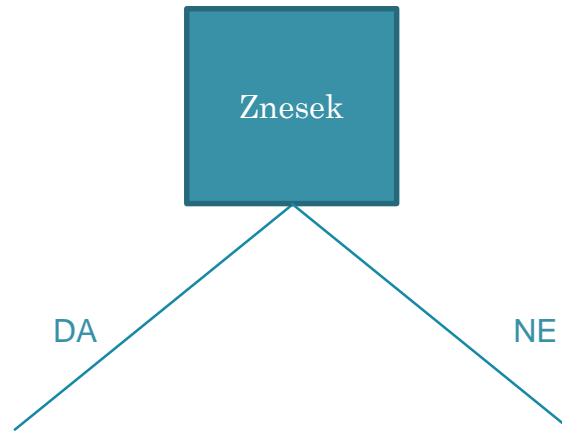
DA

NE

Posojilo vrnjeno	Spol	Trenutno zakreditiran	V kazenskem postopku
NE	MOŠKI	NE	DA
DA	MOŠKI	NE	NE
DA	ŽENSKA	DA	NE
NE	MOŠKI	NE	DA
DA	MOŠKI	DA	NE

Posojilo vrnjeno	Spol	Trenutno zakreditiran	V kazenskem postopku
NE	ŽENSKA	DA	NE
NE	MOŠKI	DA	NE
DA	MOŠKI	NE	DA

# NALOGA 4



Posojilo vrnjeno	Spol	Trenutno zakreditiran	V kazenskem postopku
NE	MOŠKI	NE	DA
DA	MOŠKI	NE	NE
DA	ŽENSKA	DA	NE
NE	MOŠKI	NE	DA
DA	MOŠKI	DA	NE

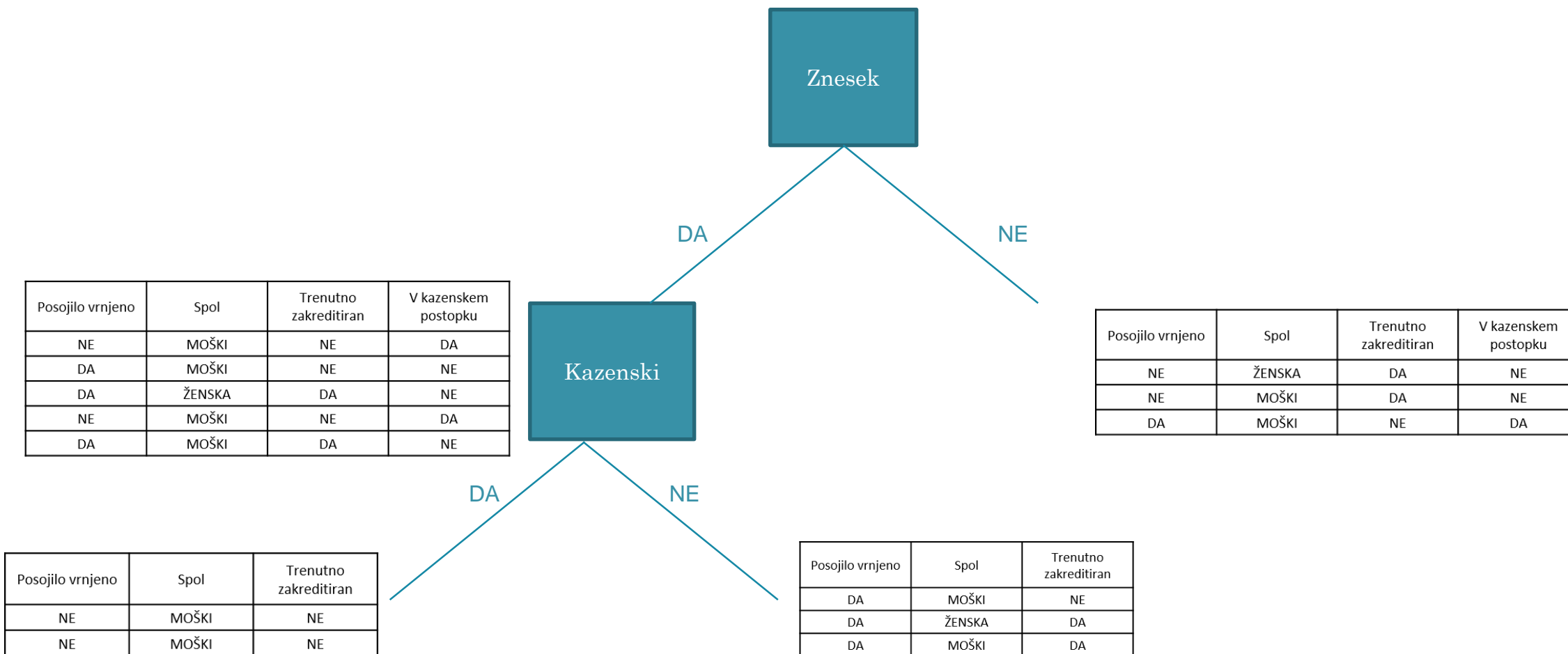
Posojilo vrnjeno	Spol	Trenutno zakreditiran	V kazenskem postopku
NE	ŽENSKA	DA	NE
NE	MOŠKI	DA	NE
DA	MOŠKI	NE	DA

$$Gini(\text{Spol}) = \frac{4}{5} * \left( \left( \frac{2}{4} \right)^2 + \left( \frac{2}{4} \right)^2 \right) + \frac{1}{5} * \left( \left( \frac{1}{1} \right)^2 + \left( \frac{0}{1} \right)^2 \right) - \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) = 0.08$$

$$Gini(\text{Zakreditiran}) = \frac{2}{5} * \left( \left( \frac{2}{2} \right)^2 + \left( \frac{0}{2} \right)^2 \right) + \frac{3}{5} * \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) - \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) = 0.213$$

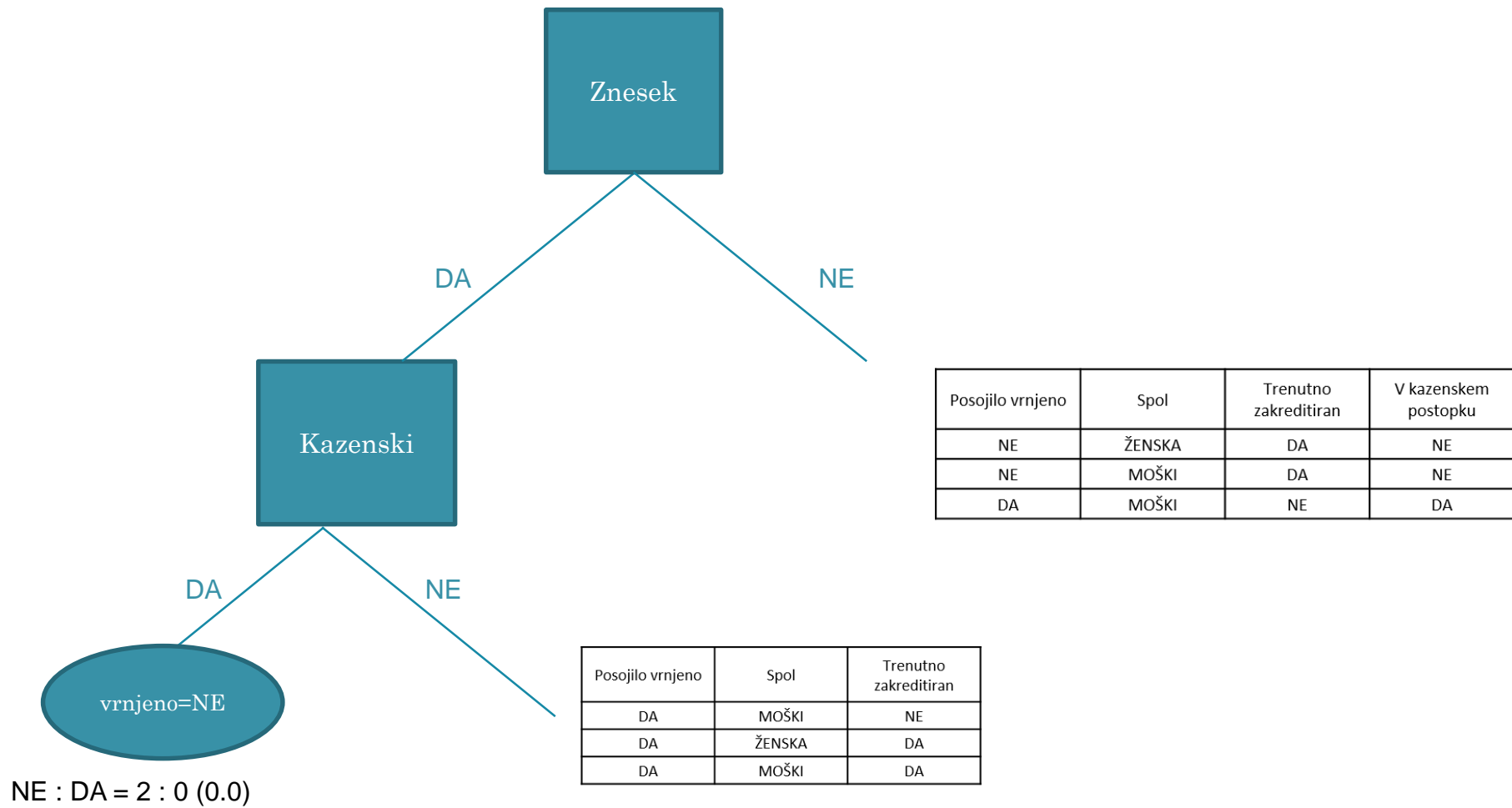
$$Gini(\text{Kazenski}) = \frac{2}{5} * \left( \left( \frac{0}{2} \right)^2 + \left( \frac{2}{2} \right)^2 \right) + \frac{3}{5} * \left( \left( \frac{3}{3} \right)^2 + \left( \frac{0}{3} \right)^2 \right) - \left( \left( \frac{3}{5} \right)^2 + \left( \frac{2}{5} \right)^2 \right) = 0.48 \leftarrow \text{izberemo}$$

# NALOGA 4

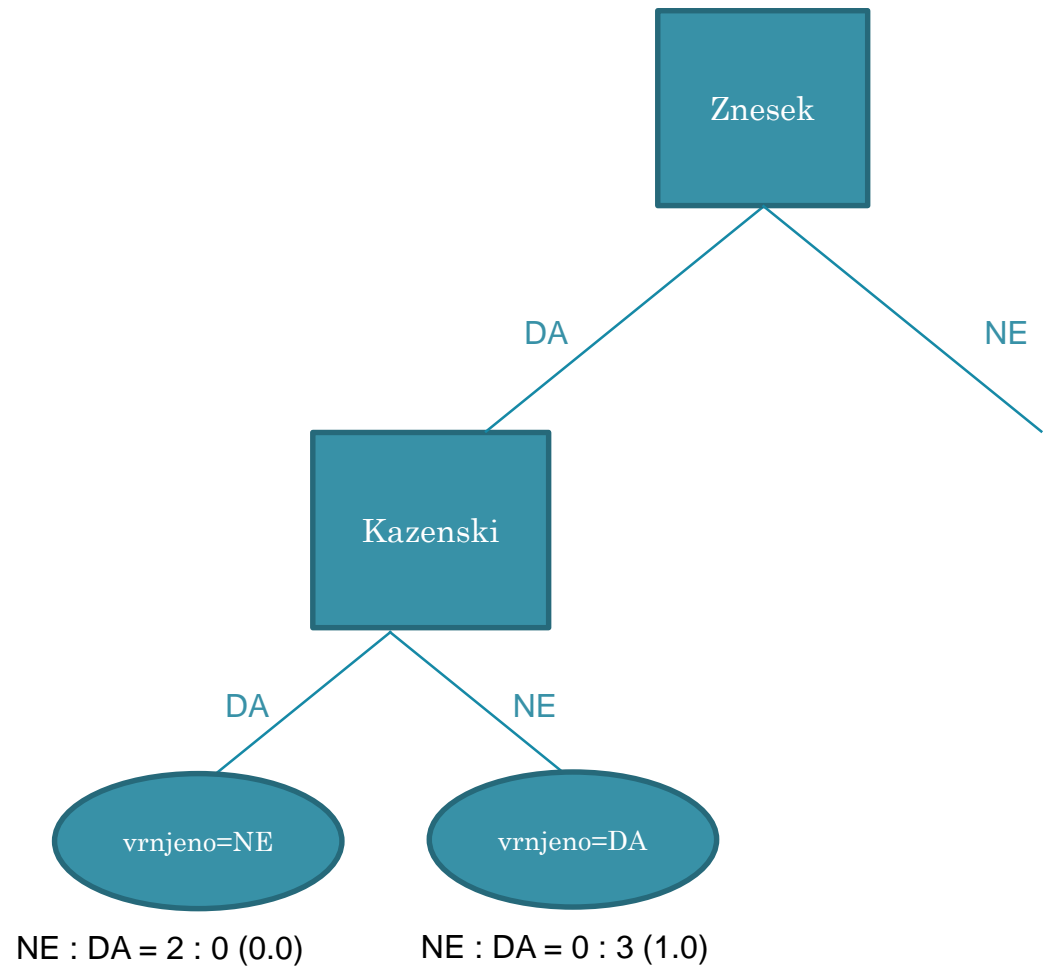




# NALOGA 4

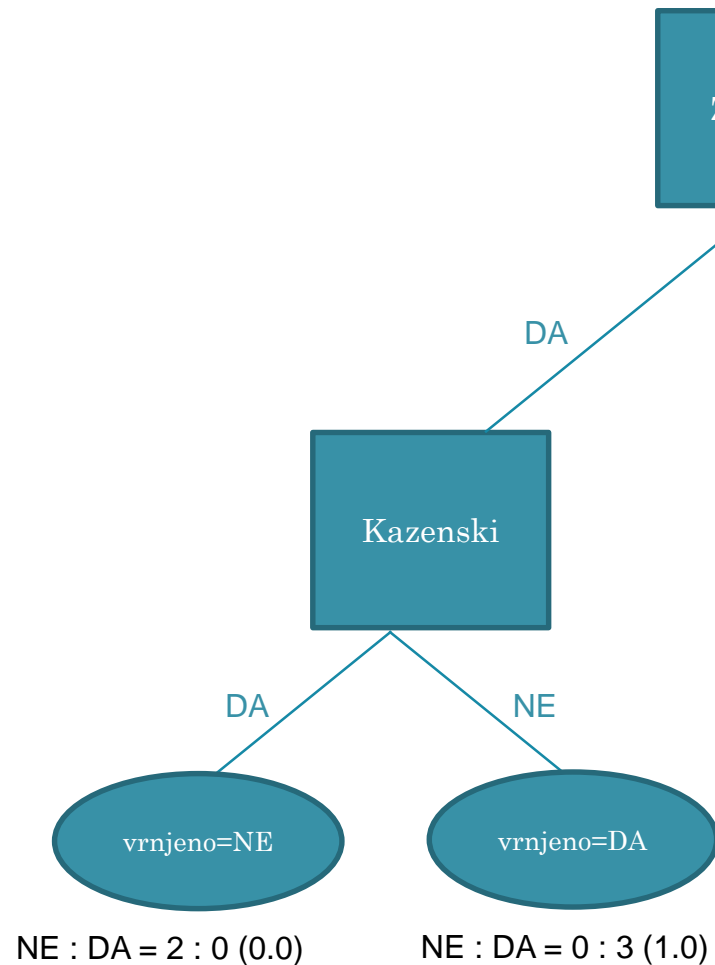


# NALOGA 4



Posojilo vrnjeno	Spol	Trenutno zakreditiran	V kazenskem postopku
NE	ŽENSKA	DA	NE
NE	MOŠKI	DA	NE
DA	MOŠKI	NE	DA

# NALOGA 4



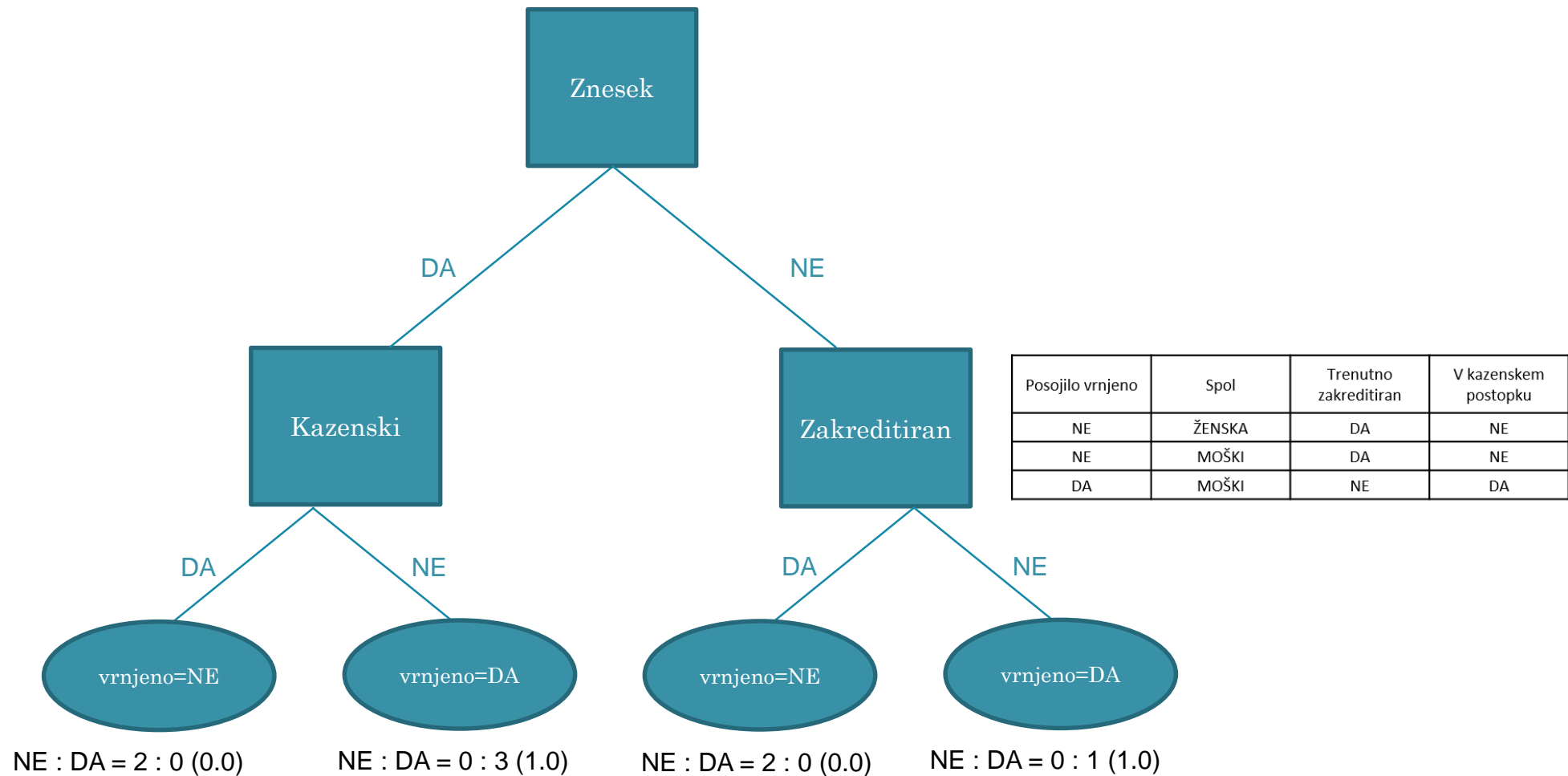
Posojilo vrnjeno	Spol	Trenutno zakreditiran	V kazenskem postopku
NE	ŽENSKA	DA	NE
NE	MOŠKI	DA	NE
DA	MOŠKI	NE	DA

$$Gini(\text{Spol}) = \frac{2}{3} * \left( \left( \frac{1}{2} \right)^2 + \left( \frac{1}{2} \right)^2 \right) + \frac{1}{3} * \left( \left( \frac{0}{1} \right)^2 + \left( \frac{1}{1} \right)^2 \right) - \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) = 0.11$$

$$Gini(\text{Zakreditiran}) = \frac{2}{3} * \left( \left( \frac{0}{2} \right)^2 + \left( \frac{2}{2} \right)^2 \right) + \frac{1}{3} * \left( \left( \frac{1}{1} \right)^2 + \left( \frac{0}{1} \right)^2 \right) - \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) = 0.444$$

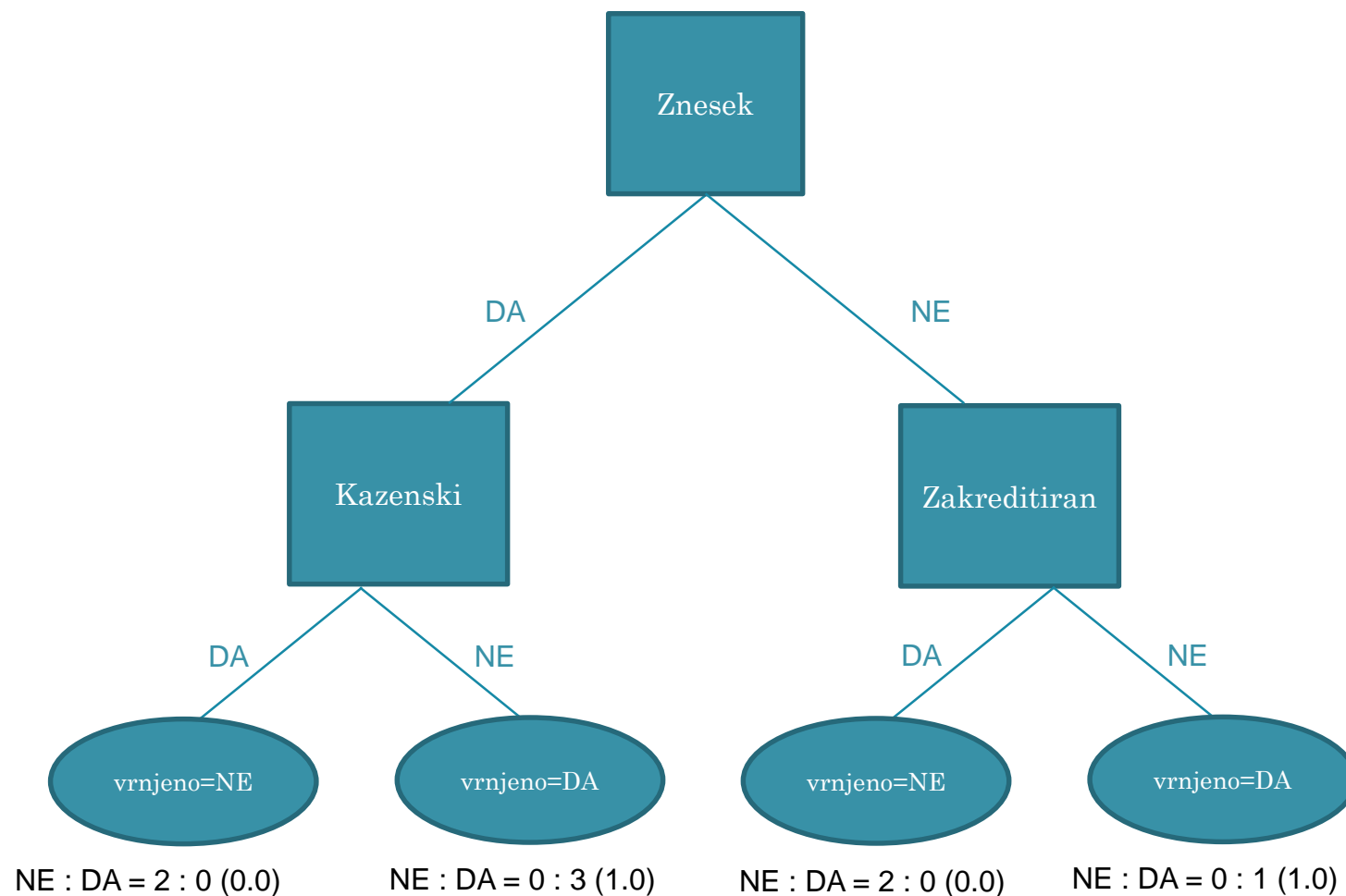
$$Gini(\text{Kazenski}) = \frac{1}{3} * \left( \left( \frac{1}{1} \right)^2 + \left( \frac{0}{1} \right)^2 \right) + \frac{2}{3} * \left( \left( \frac{0}{2} \right)^2 + \left( \frac{2}{2} \right)^2 \right) - \left( \left( \frac{1}{3} \right)^2 + \left( \frac{2}{3} \right)^2 \right) = 0.444$$

# NALOGA 4



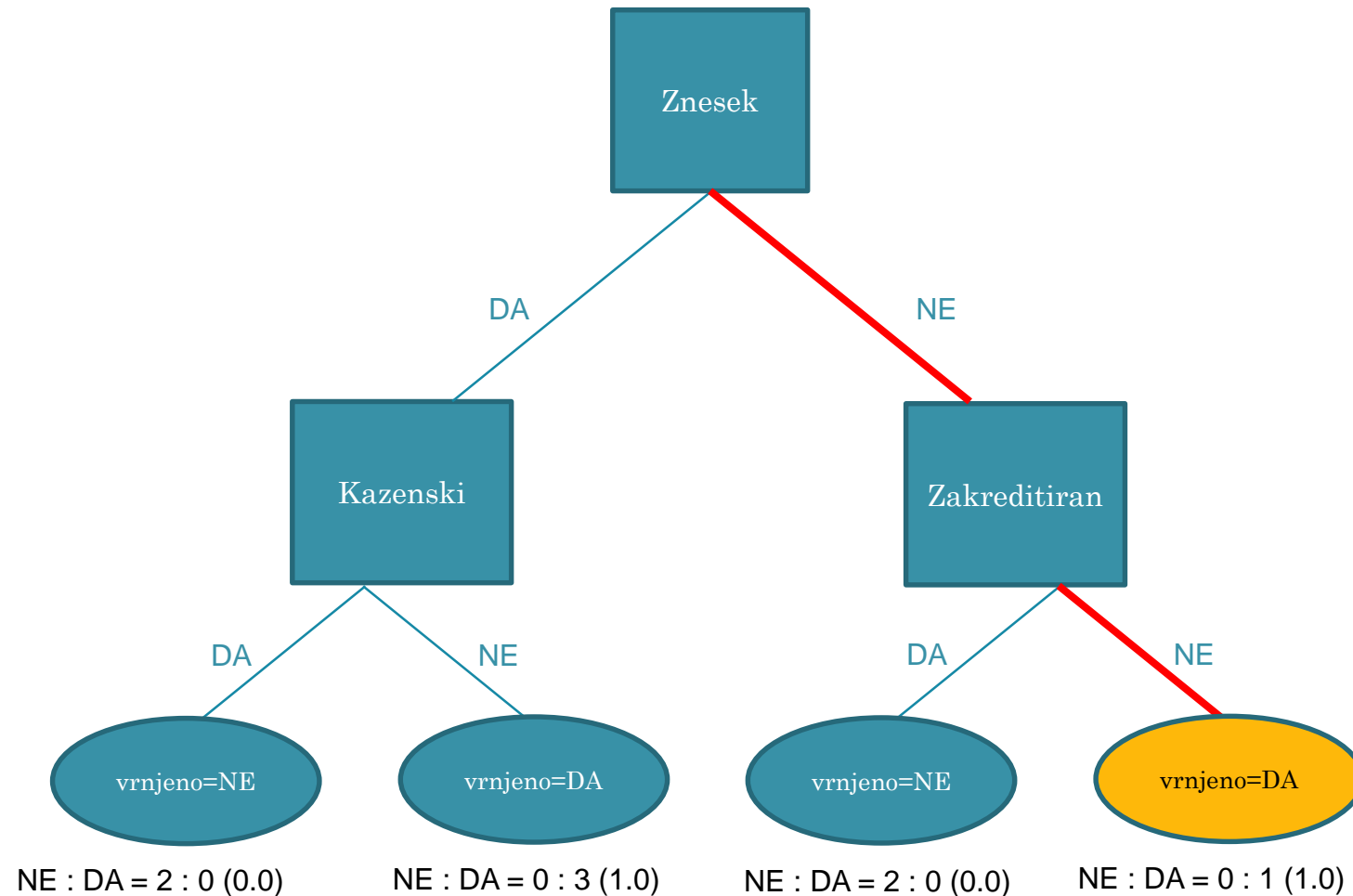
# NALOGA 4

Kakšna je napoved za direktorjevo ženo, ki ni v kazenskem postopku, nikoli ni imela kreditov in bi si sposodila 20 000 evrov?



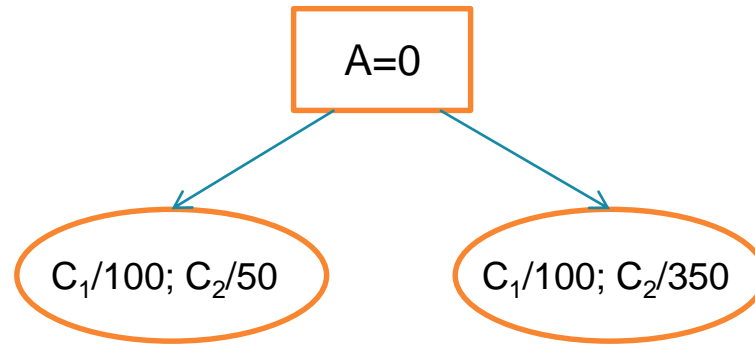
# NALOGA 4

Kakšna je napoved za direktorjevo ženo, ki ni v kazenskem postopku, nikoli ni imela kreditov in bi si sposodila 20 000 evrov?



# NALOGA 5

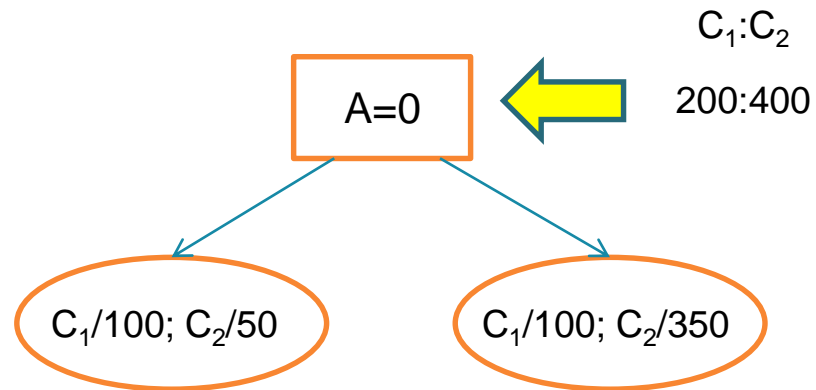
Spodaj je odločitveno drevo s frekvencami primerov v listih za razreda C1 in C2:



Izračunaj pričakovano klasifikacijsko točnost drevesa z  $m$ -oceno, če je:

- a)  $m = 0$ ,
- b)  $m = 1000$

# NALOGA 5



Število primerov:  $200 + 400 = 600$

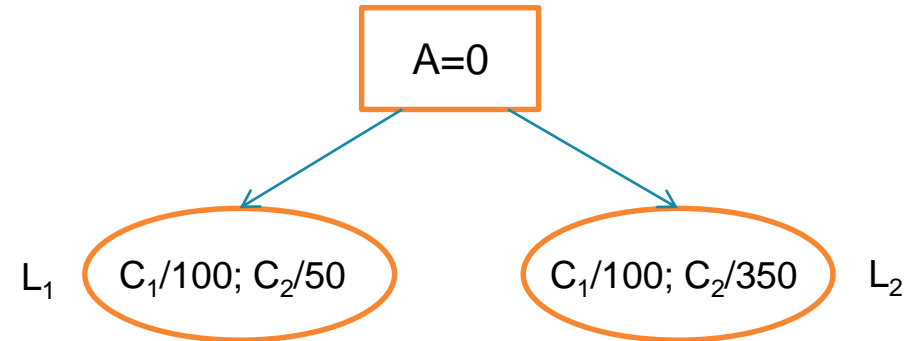
Apriorna verjetnost razredov:  $p_a(C_1) = \frac{200}{600} = 0.33$

$p_a(C_2) = \frac{400}{600} = 0.67$



# NALOGA 5

a)  $m = 0$



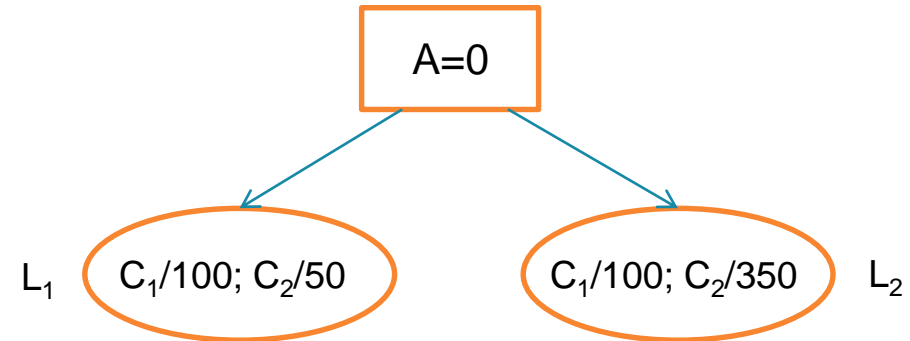
$$p(C_1|L_1) = \frac{r+m*pa}{n+m} =$$



Statična točnost lista = verjetnost klasifikacije v pravilni razred

# NALOGA 5

a)  $m = 0$



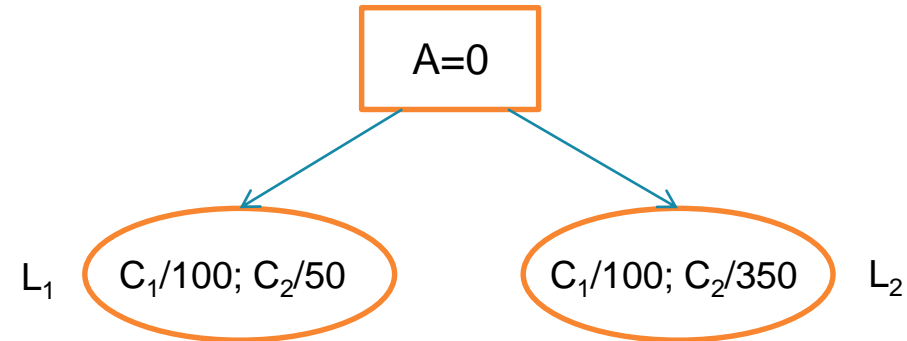
$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{\quad}{150+}$$



Statična točnost lista = verjetnost klasifikacije v pravilni razred

# NALOGA 5

a)  $m = 0$



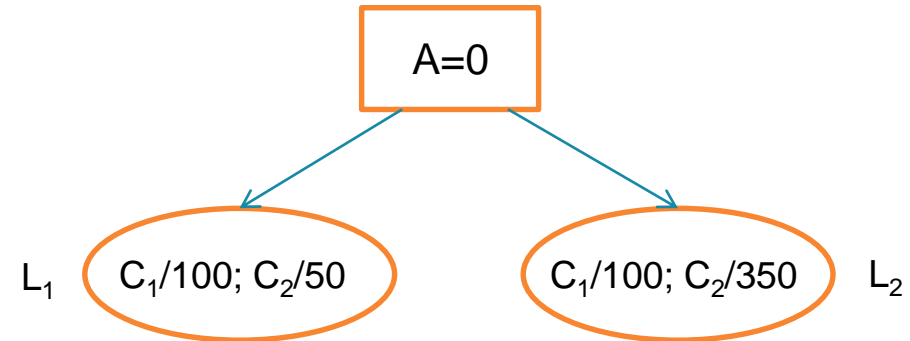
$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{\quad}{150+0}$$



Statična točnost lista = verjetnost klasifikacije v pravilni razred

# NALOGA 5

a)  $m = 0$



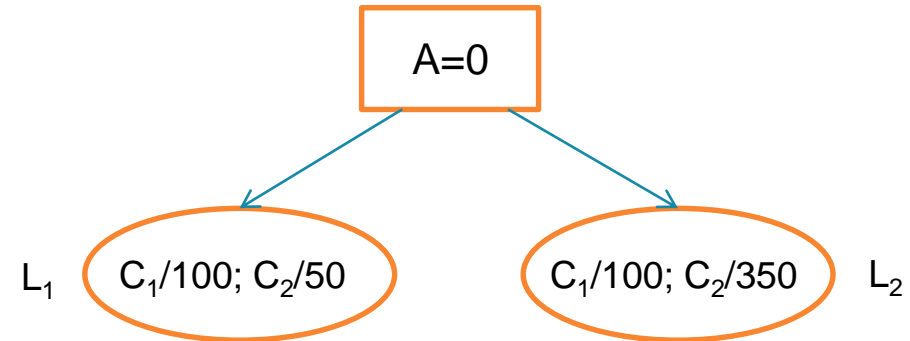
$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+}{150+0}$$



Statična točnost lista = verjetnost klasifikacije v pravilni razred

# NALOGA 5

a)  $m = 0$



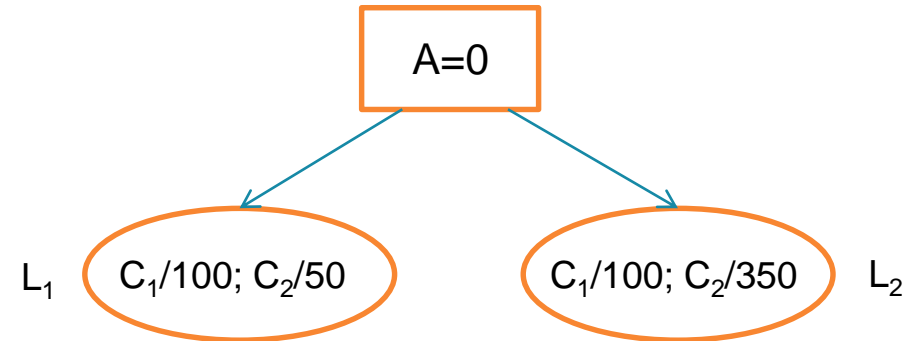
$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0}$$



Statična točnost lista = verjetnost klasifikacije v pravilni razred

# NALOGA 5

a)  $m = 0$



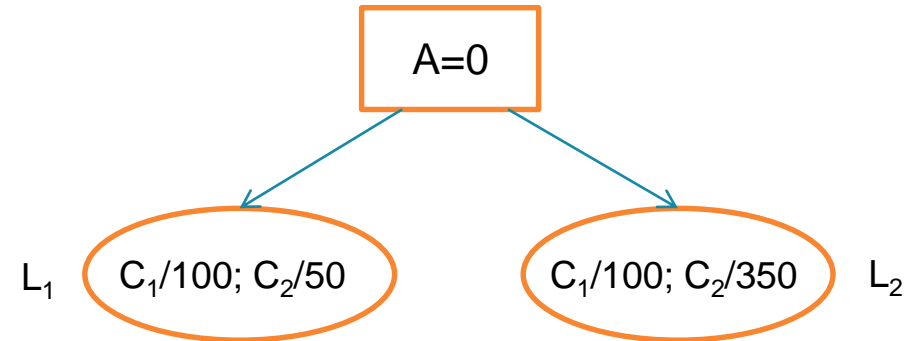
$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$



Statična točnost lista = verjetnost klasifikacije v pravilni razred

# NALOGA 5

a)  $m = 0$

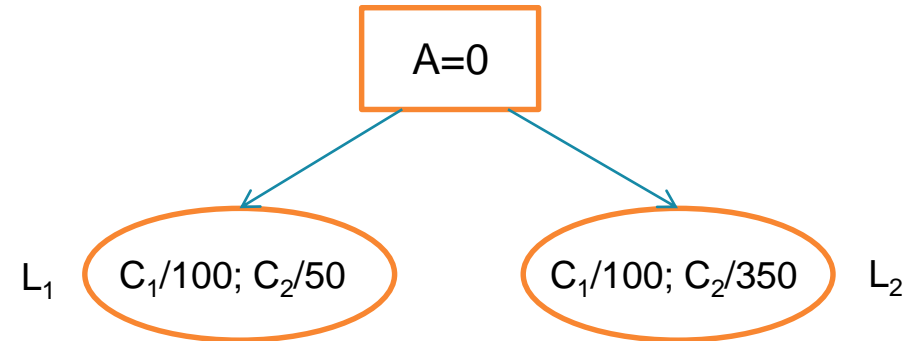


$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$

$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{450+}{450+}$$

# NALOGA 5

a)  $m = 0$



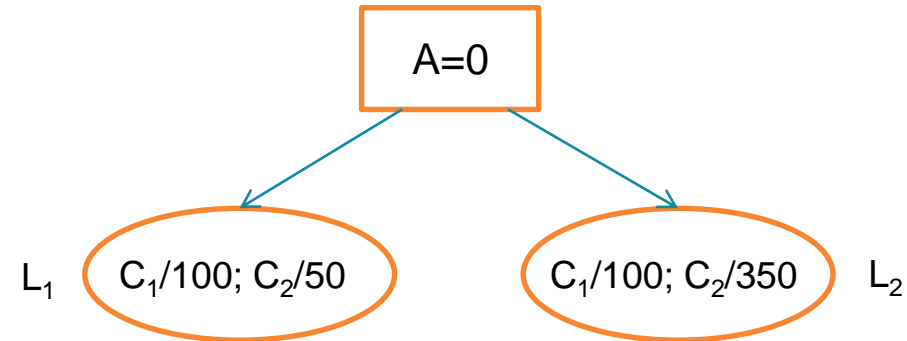
$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$

$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{450+0}{450+0}$$



# NALOGA 5

a)  $m = 0$

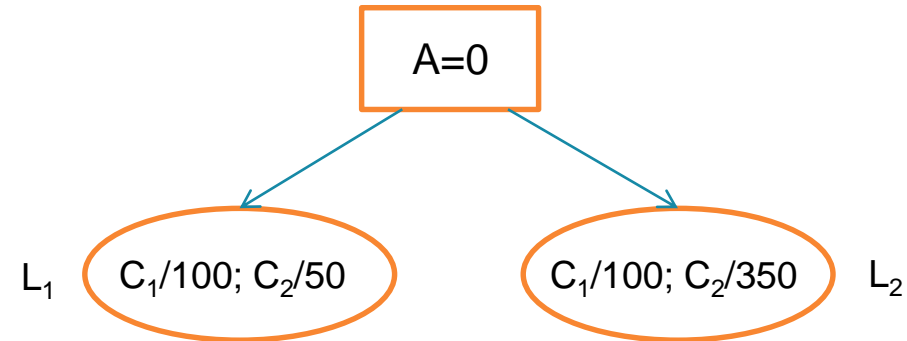


$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$

$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{350+}{450+0}$$

# NALOGA 5

a)  $m = 0$

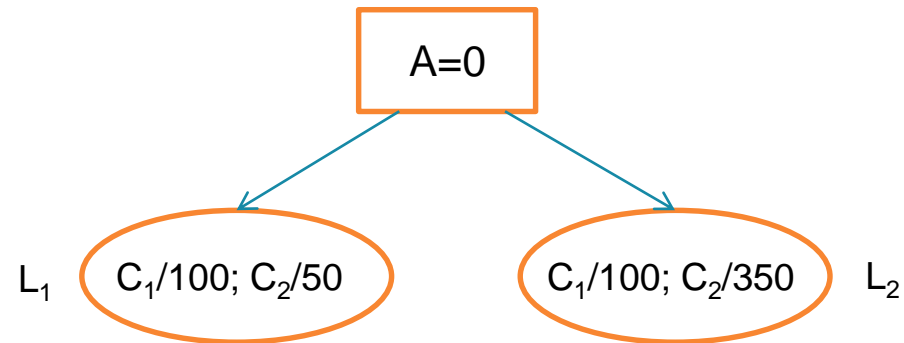


$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$

$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{350+0*0.67}{450+0}$$

# NALOGA 5

a)  $m = 0$

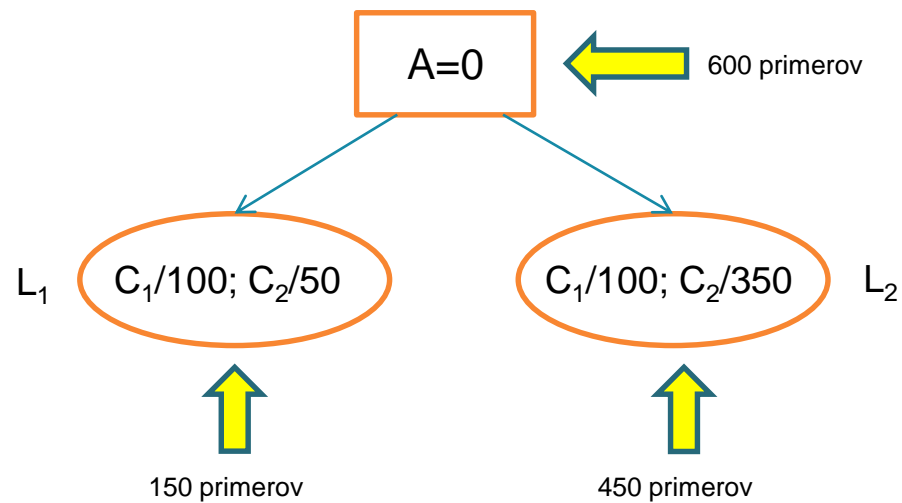


$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$

$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{350+0*0.67}{450+0} = 0.78$$

# NALOGA 5

a)  $m = 0$



$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$

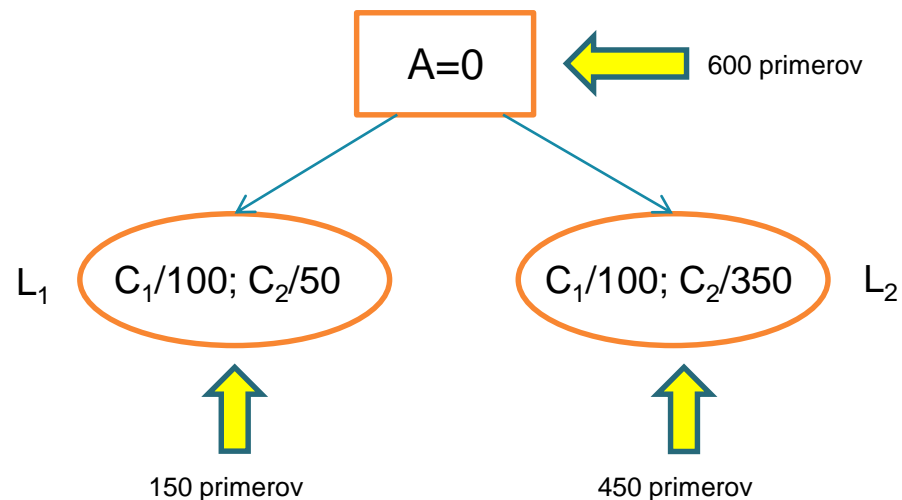
$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{350+0*0.67}{450+0} = 0.78$$

Pričakovana točnost odločitvenega drevesa:  $\frac{150}{600} * 0.67 +$

verjetnost, da pridemo v list  $L_1$ , pomnožena s točnostjo lista  $L_1$

# NALOGA 5

a)  $m = 0$



$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$

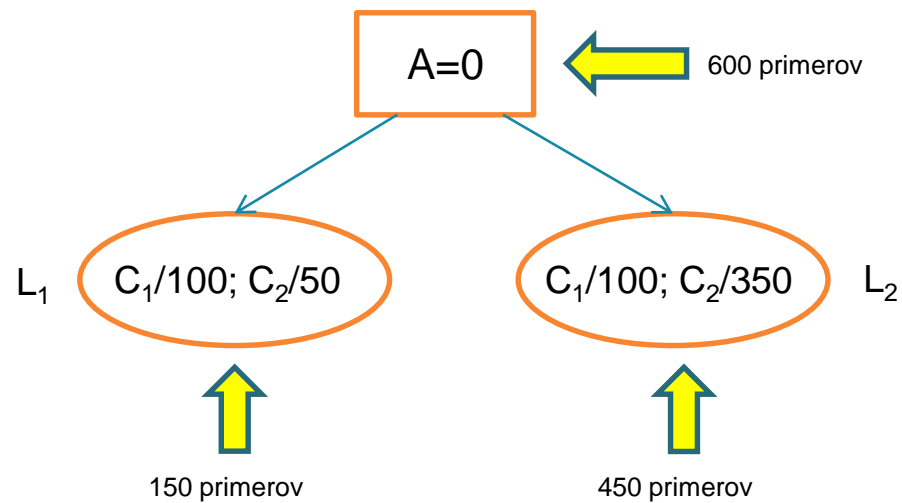
$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{350+0*0.67}{450+0} = 0.78$$

Pričakovana točnost odločitvenega drevesa:  $\frac{150}{600} * 0.67 + \frac{450}{600} * 0.78$

verjetnost, da pridemo v list  $L_2$ , pomnožena s točnostjo lista  $L_2$

# NALOGA 5

a)  $m = 0$



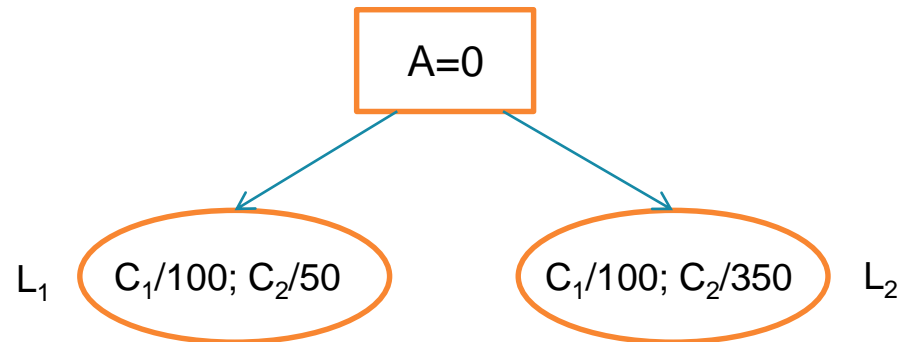
$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+0*0.33}{150+0} = 0.67$$

$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{350+0*0.67}{450+0} = 0.78$$

Pričakovana točnost odločitvenega drevesa:  $\frac{150}{600} * 0.67 + \frac{450}{600} * 0.78 = 0.75$

# NALOGA 5

a)  $m = 1000$



$$p(C_1|L_1) = \frac{r+m*pa}{n+m} = \frac{100+1000*0.33}{150+1000} = 0.37$$

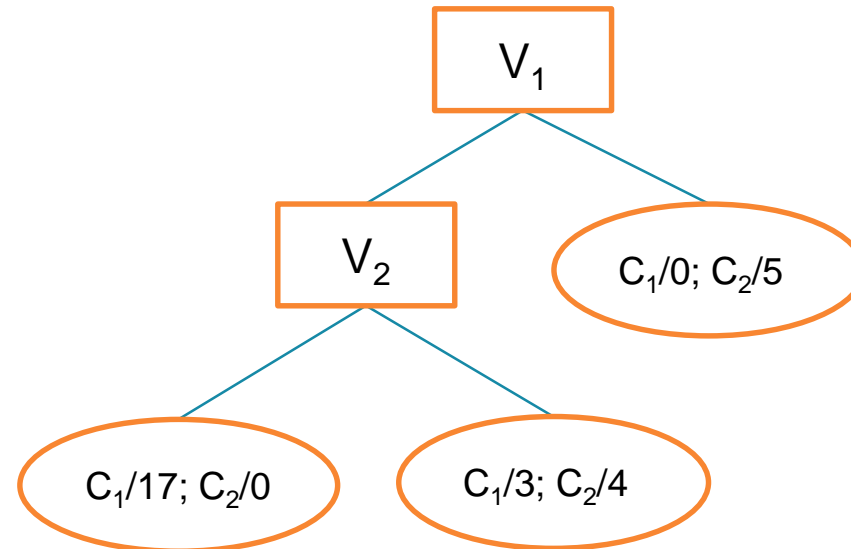
$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{350+1000*0.67}{450+1000} = 0.7$$

Pričakovana točnost odločitvenega drevesa:  $\frac{150}{600} * 0.37 + \frac{450}{600} * 0.7 = 0.62$

# NALOGA 6

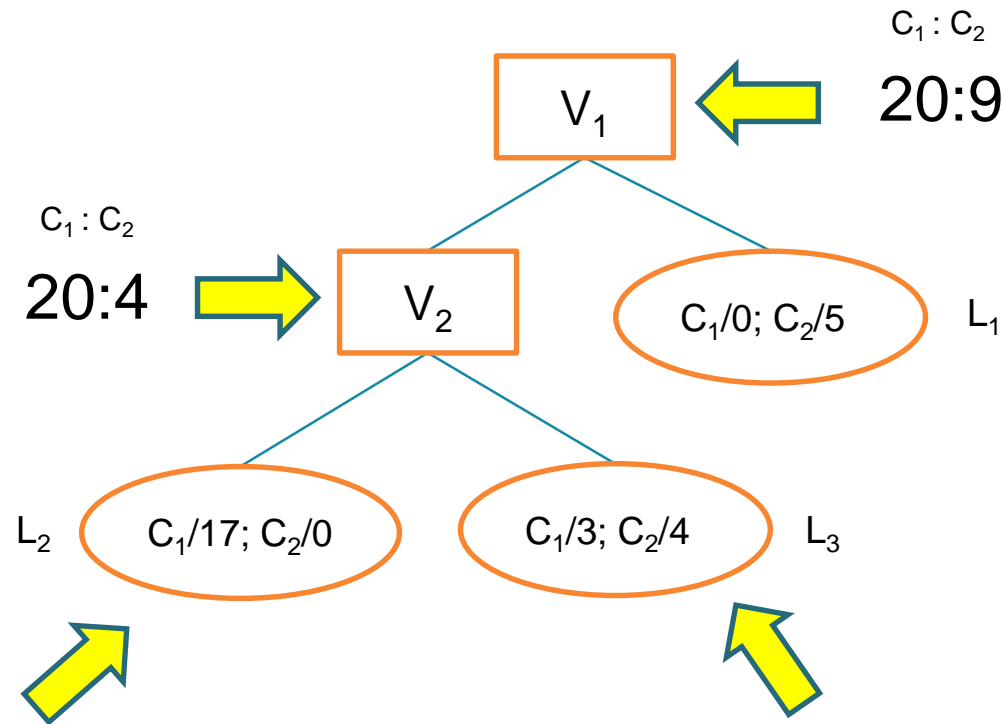
Podano je odločitveno drevo za klasifikacijo v razrede z naslednjimi apriornimi verjetnostmi razredov:  $p_a(C_1)=0.6$  in  $p_a(C_2)=0.4$ .

Obreži podano odločitveno drevo s postopkom minimizacije napake in vrednostjo  $m = 10$ .





# NALOGA 6



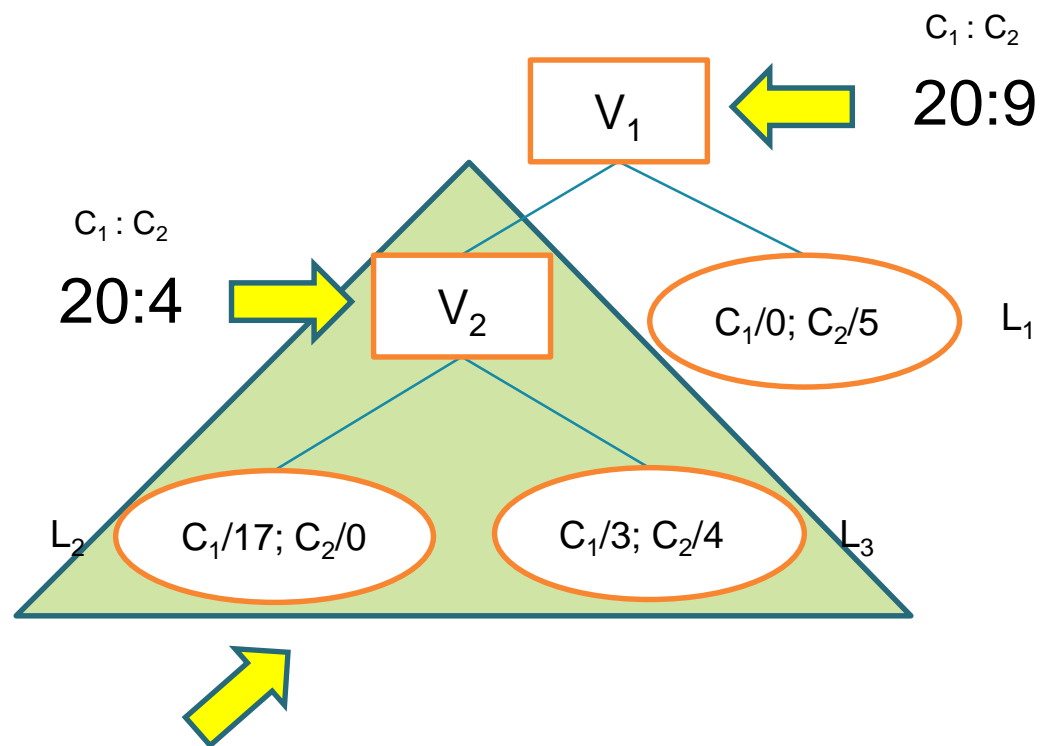
Statična točnost lista  $L_2$ :

$$p(C_1|L_2) = \frac{r+m*pa}{n+m} = \frac{17+10*0.6}{17+10} = 0.852$$

Statična točnost lista  $L_3$ :

$$p(C_2|L_3) = \frac{r+m*pa}{n+m} = \frac{4+10*0.4}{7+10} = 0.471$$

# NALOGA 6



Vzratna točnost v notranjem vozlišču  $V_2$ :

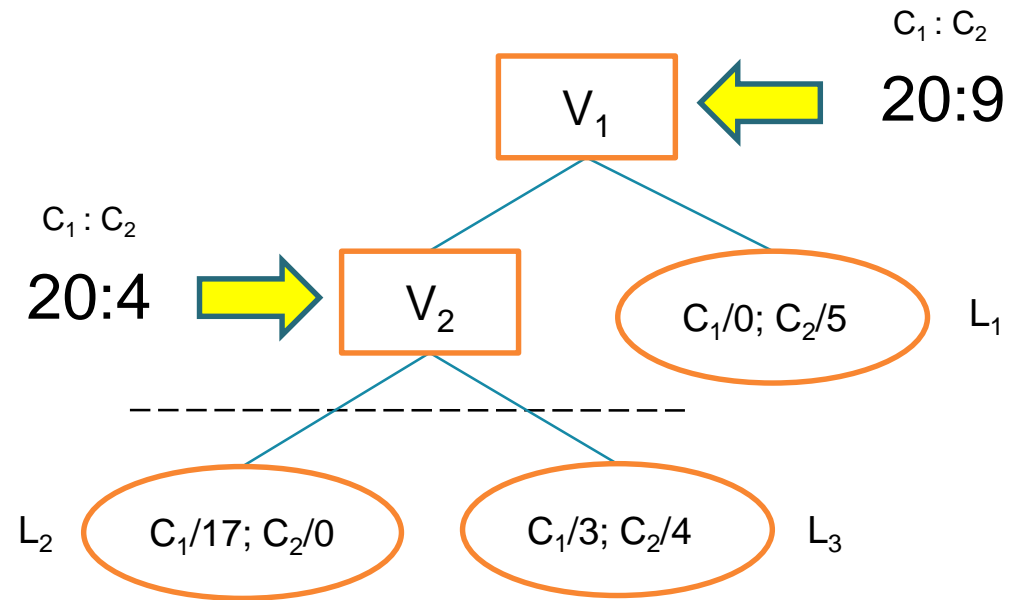
$$\frac{17}{24} * 0.852 + \frac{7}{24} * 0.471 = 0.741$$

Statična točnost notranjega vozlišča  $V_2$ :

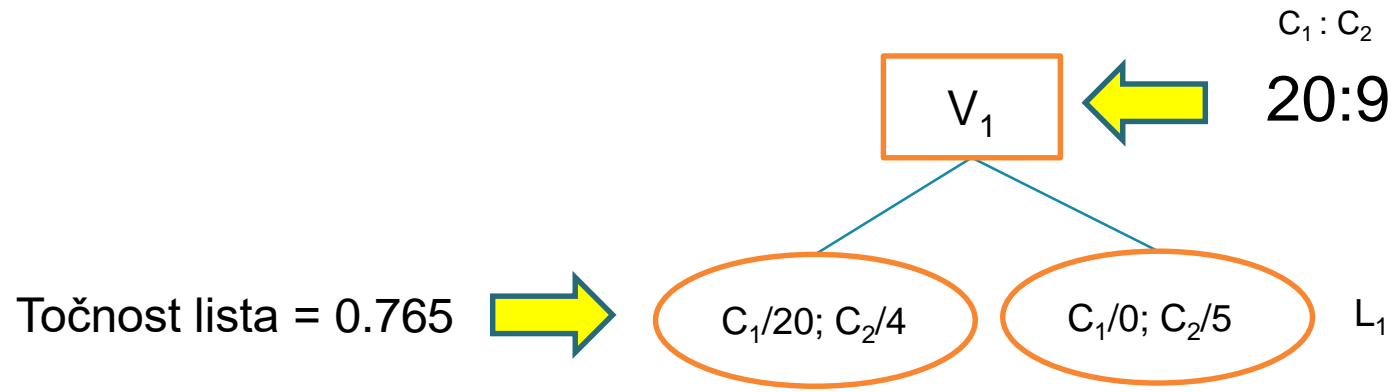
$$p(C_1|V_2) = \frac{r+m*pa}{n+m} = \frac{20+10*0.6}{24+10} = 0.765$$

**Ker je statična točnost notranjega vozlišča (0.765) večja od vzratne točnosti (0.741) → drevo **porežemo**.**

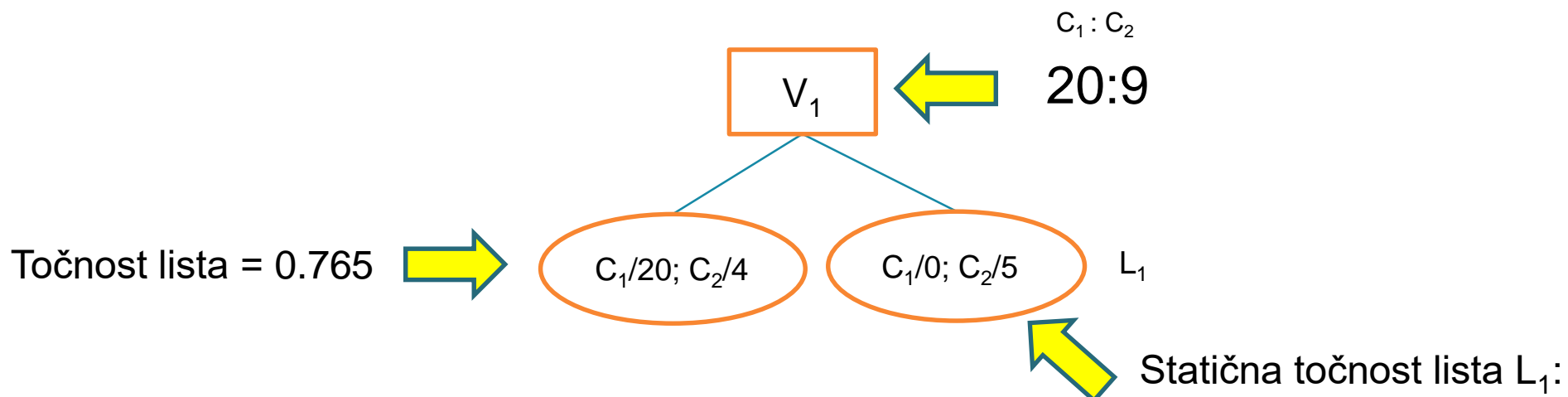
# NALOGA 6



# NALOGA 6



# NALOGA 6



Vzratna točnost korena:

$$\frac{24}{29} * 0.765 + \frac{5}{29} * 0.6 = 0.737$$

$$p(C_2|L_1) = \frac{r+m*pa}{n+m} = \frac{5+10*0.4}{5+10} = 0.6$$

Statična točnost korena:

$$p(C_1|V_1) = \frac{r+m*pa}{n+m} = \frac{20+10*0.6}{29+10} = 0.67$$

**Tokrat je vzratna točnost (0.737) večja od statične točnosti (0.67)**

→ drevesa **ne porežemo**.

# NALOGA 7

Klasifikator je na 4-razrednem problemu dosegel naslednje rezultate, predstavljene z matriko zmot. V vsaki celici je vneseno ustrezno število\_testnih\_primerov/cena\_napačne\_klasifikacije:

Pravi razred	Napovedani razred			
	$C_1$	$C_2$	$C_3$	$C_4$
$C_1$	12/0	0/1	4/2	4/2
$C_2$	5/1	12/0	2/1	5/3
$C_3$	5/10	0/3	20/0	15/2
$C_4$	8/2	0/1	2/4	46/0

Izračunaj:

- klasifikacijsko točnost klasifikatorja,
- pričakovano točnost večinskega klasifikatorja (predpostavi, da je verjetnostna distribucija po razredih v testni množici enaka distribuciji v učni množici),
- povprečno ceno napačne klasifikacije.

# NALOGA 7

Klasifikator je na 4-razrednem problemu dosegel naslednje rezultate, predstavljene z matriko zmot. V vsaki celici je vneseno ustrezno število\_testnih\_primerov/cena\_napačne\_klasifikacije:

Pravi razred	Napovedani razred			
	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>
C <sub>1</sub>	12/0	0/1	4/2	4/2
C <sub>2</sub>	5/1	12/0	2/1	5/3
C <sub>3</sub>	5/10	0/3	20/0	15/2
C <sub>4</sub>	8/2	0/1	2/4	46/0

Testna množica vsebuje:

$12 + 4 + 4 = 20$  primerov iz razreda C<sub>1</sub>

$5 + 12 + 2 + 5 = 24$  primerov iz razreda C<sub>2</sub>

$5 + 20 + 15 = 40$  primerov iz razreda C<sub>3</sub>

$8 + 2 + 46 = 56$  primerov iz razreda C<sub>4</sub>

Skupaj je to  $20 + 24 + 40 + 56 = 140$  primerov

a) klasifikacijska točnost klasifikatorja

$$(12 + 12 + 20 + 46) / 140 = 0.643$$

# NALOGA 7

Klasifikator je na 4-razrednem problemu dosegel naslednje rezultate, predstavljene z matriko zmot. V vsaki celici je vneseno ustrezno število\_testnih\_primerov/cena\_napačne\_klasifikacije:

Pravi razred	Napovedani razred			
	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>
C <sub>1</sub>	12/0	0/1	4/2	4/2
C <sub>2</sub>	5/1	12/0	2/1	5/3
C <sub>3</sub>	5/10	0/3	20/0	15/2
C <sub>4</sub>	8/2	0/1	2/4	46/0

Testna množica vsebuje:

$12 + 4 + 4 = 20$  primerov iz razreda C<sub>1</sub>

$5 + 12 + 2 + 5 = 24$  primerov iz razreda C<sub>2</sub>

$5 + 20 + 15 = 40$  primerov iz razreda C<sub>3</sub>

$8 + 2 + 46 = 56$  primerov iz razreda C<sub>4</sub>

Skupaj je to  $20 + 24 + 40 + 56 = 140$  primerov

b) pričakovana točnost večinskega klasifikatorja

večinski razred je C<sub>4</sub>, zato je točnost več. klas. =  $56 / 140 = 0.4$



# NALOGA 7

Klasifikator je na 4-razrednem problemu dosegel naslednje rezultate, predstavljene z matriko zmot. V vsaki celici je vneseno ustrezno število\_testnih\_primerov/cena\_napačne\_klasifikacije:

Pravi razred	Napovedani razred			
	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>
C <sub>1</sub>	12/0	0/1	4/2	4/2
C <sub>2</sub>	5/1	12/0	2/1	5/3
C <sub>3</sub>	5/10	0/3	20/0	15/2
C <sub>4</sub>	8/2	0/1	2/4	46/0

Testna množica vsebuje:

$12 + 4 + 4 = 20$  primerov iz razreda C<sub>1</sub>

$5 + 12 + 2 + 5 = 24$  primerov iz razreda C<sub>2</sub>

$5 + 20 + 15 = 40$  primerov iz razreda C<sub>3</sub>

$8 + 2 + 46 = 56$  primerov iz razreda C<sub>4</sub>

Skupaj je to  $20 + 24 + 40 + 56 = 140$  primerov

c) povprečna cena napačne klasifikacije

$$\begin{aligned} & ( \\ & 12 \cdot 0 + 0 \cdot 1 + 4 \cdot 2 + 4 \cdot 2 + \\ & 5 \cdot 1 + 12 \cdot 0 + 2 \cdot 1 + 5 \cdot 3 + \\ & 5 \cdot 10 + 0 \cdot 3 + 20 \cdot 0 + 15 \cdot 2 + \\ & 8 \cdot 2 + 0 \cdot 1 + 2 \cdot 4 + 46 \cdot 0 \\ & ) / 140 = 1.014 \end{aligned}$$