Stabla Odlučivanja

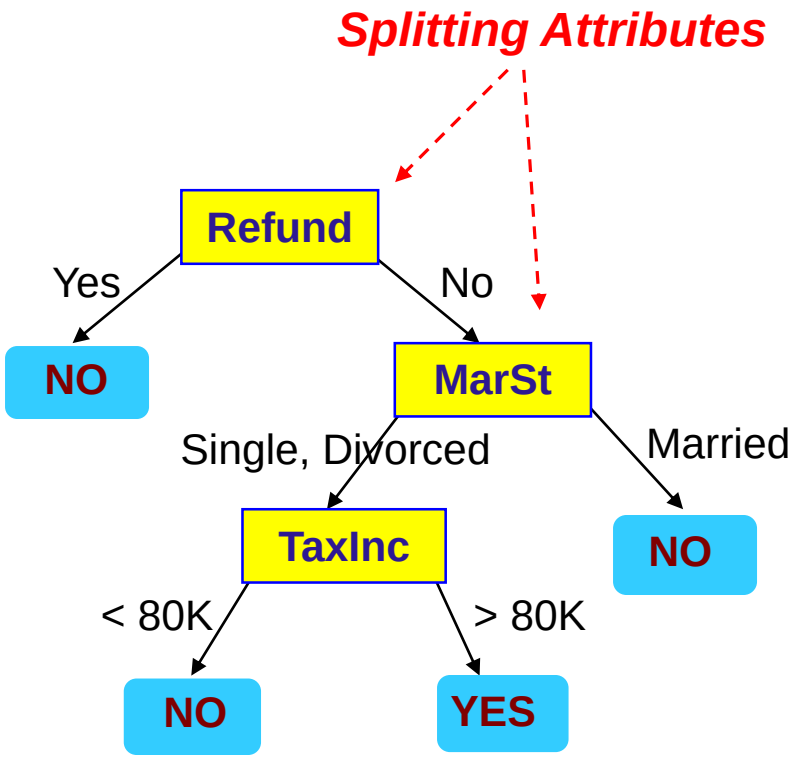
Milan M.Milosavljević

Primer stabla odlučivanja

http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/lguo/decisionTree.html

Tid	Refund (categorical)	Marital Status (categorical)	Taxable Income (continuous)	Cheat (class)
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

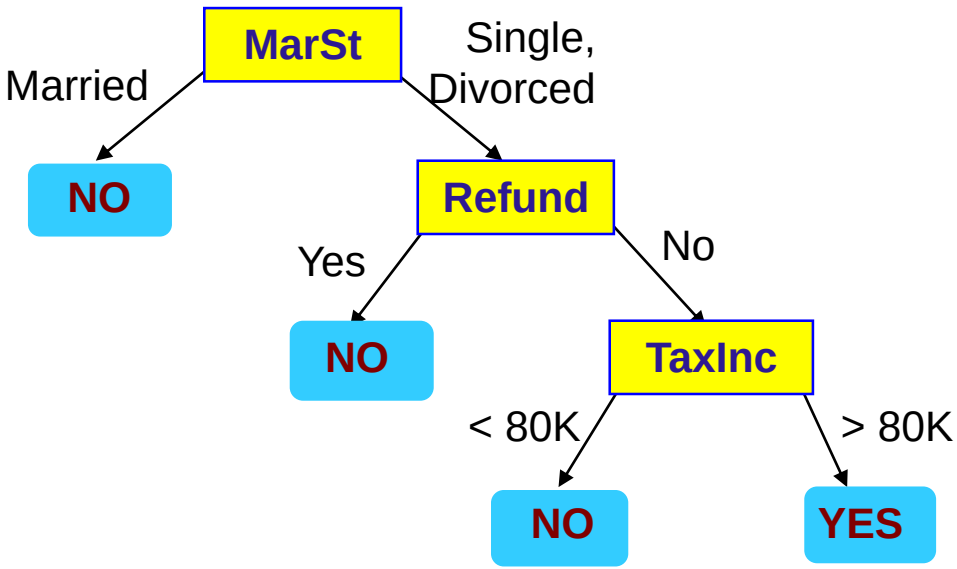
Obučavajući skup



Model: Stabla odlučivanja

Primer stabla odlučivanja

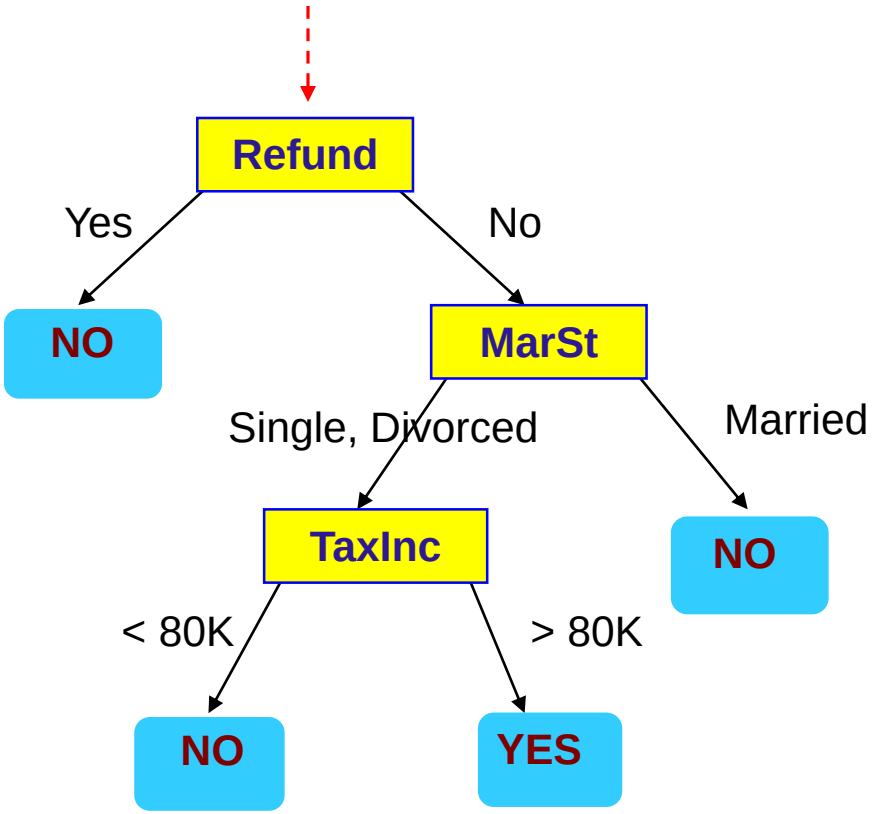
	categorical	categorical	continuous	class
Tid	**Refund**	**Marital Status**	**Taxable Income**	**Cheat**
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



Može biti više različitih stabala za zadati obučavajući skup

Primena modela na testne podatke

Startovati iz korena stabla.



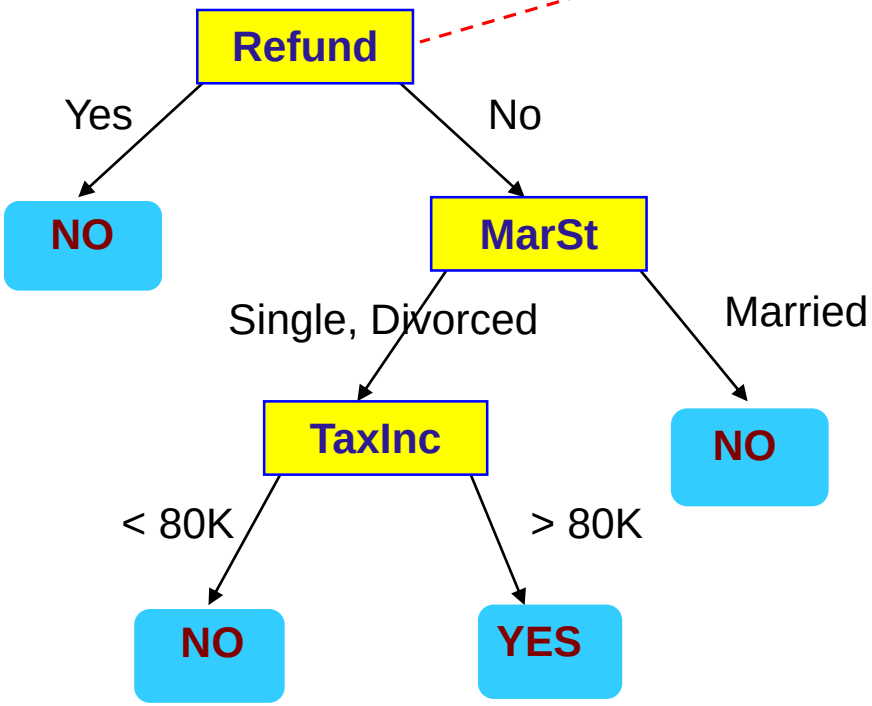
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

Test podaci

Primena modela na testne podatke

Test podaci

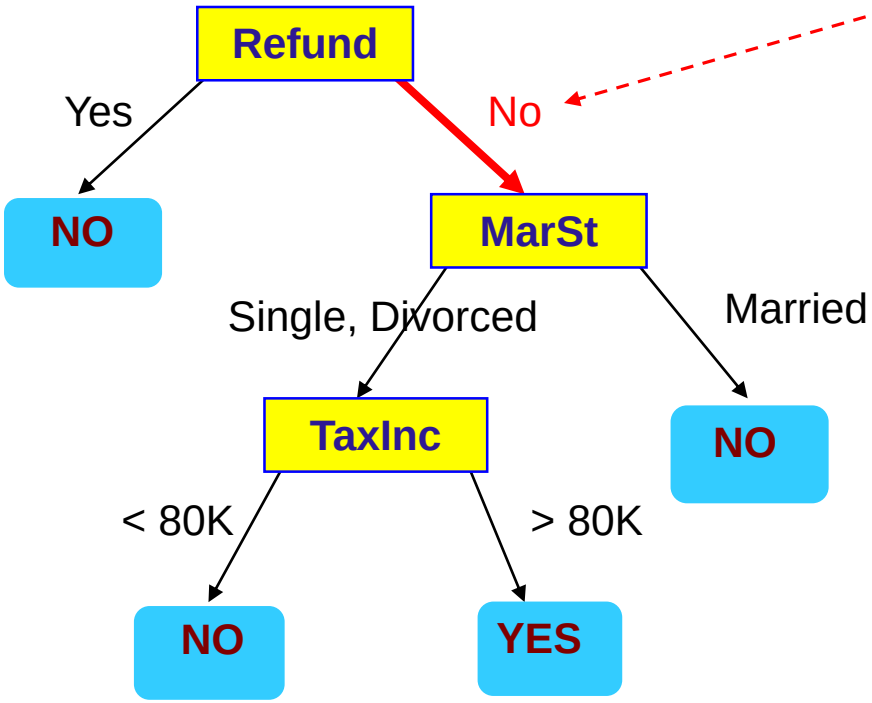
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Primena modela na testne podatke

Test podaci

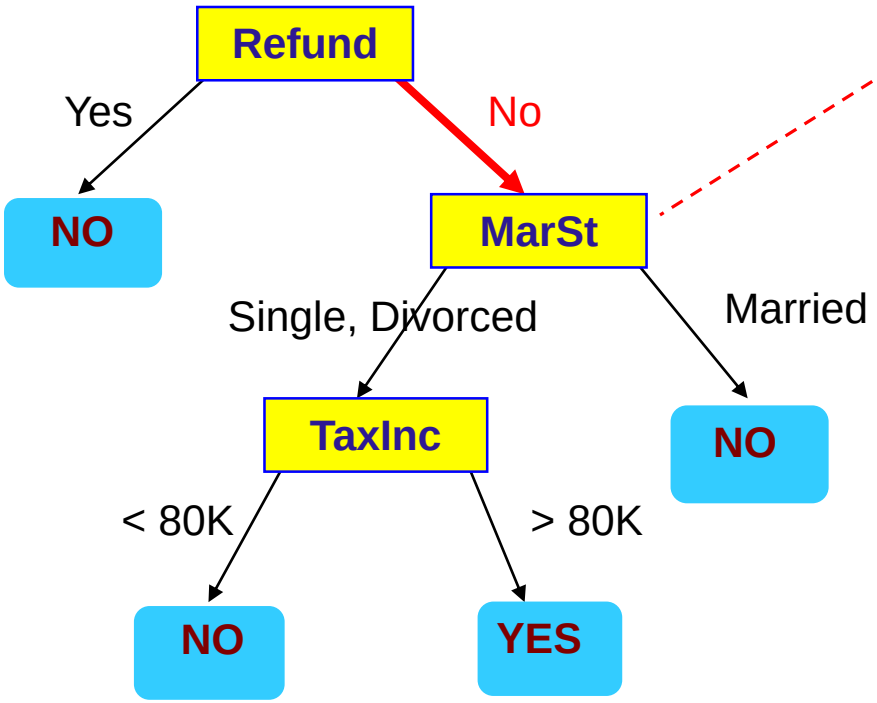
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Primena modela na testne podatke

Test podaci

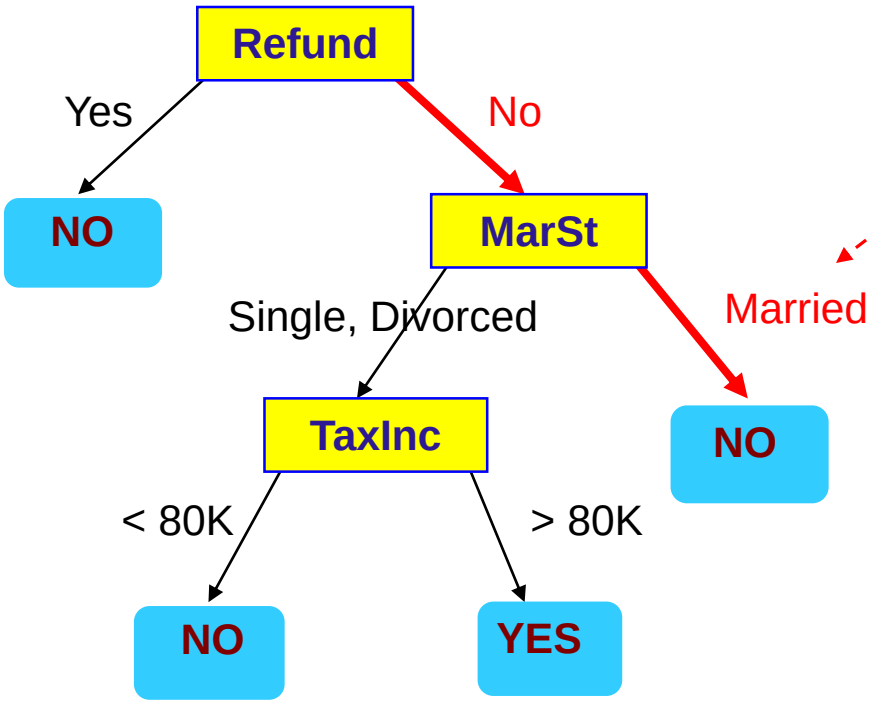
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Primena modela na testne podatke

Test podaci

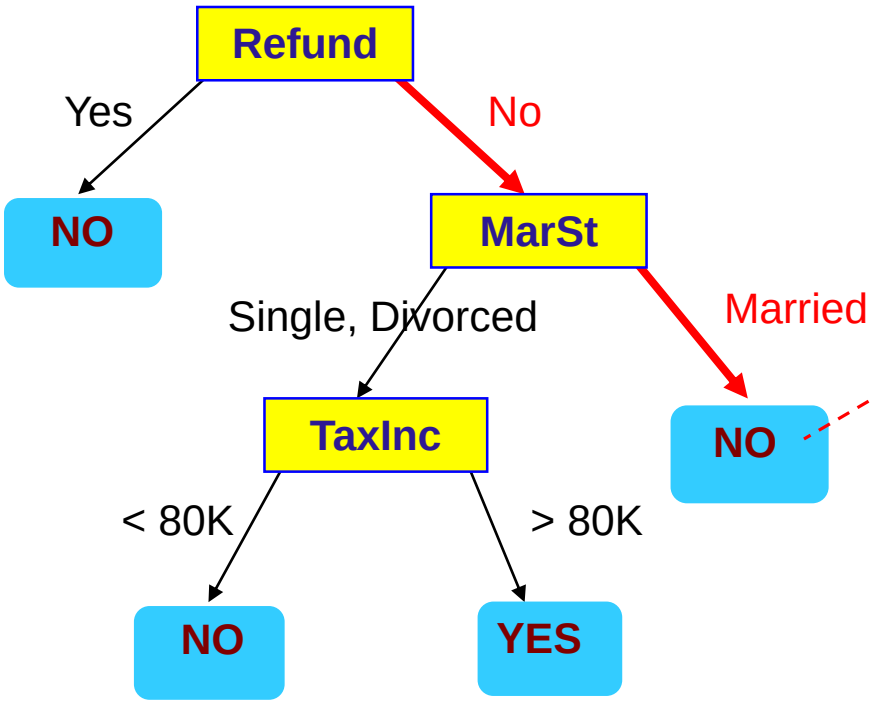
Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?



Primena modela na testne podatke

Test podaci

Refund	Marital Status	Taxable Income	Cheat
No	Married	80K	?

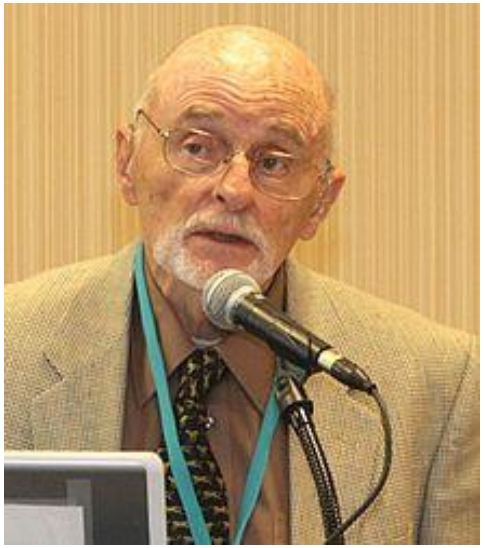


Dodeliti klasifikaciju za „Cheat“ - “No”

Indukcija po stablu odlučivanja

- Najpopularniji algoritmi algoritama:
  - Hantov algoritam
  - CART
  - ID3, C4.5
  - SLIQ,SPRINT

Hantov algoritam

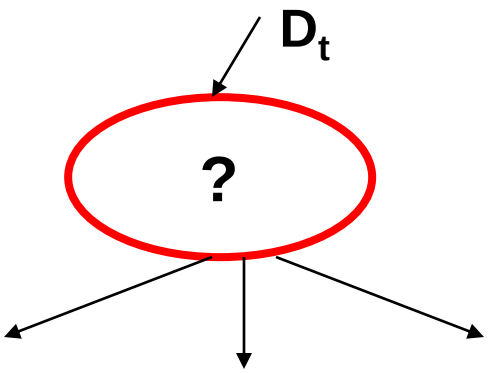


Earl B. Hunt
(1933 – 2016)

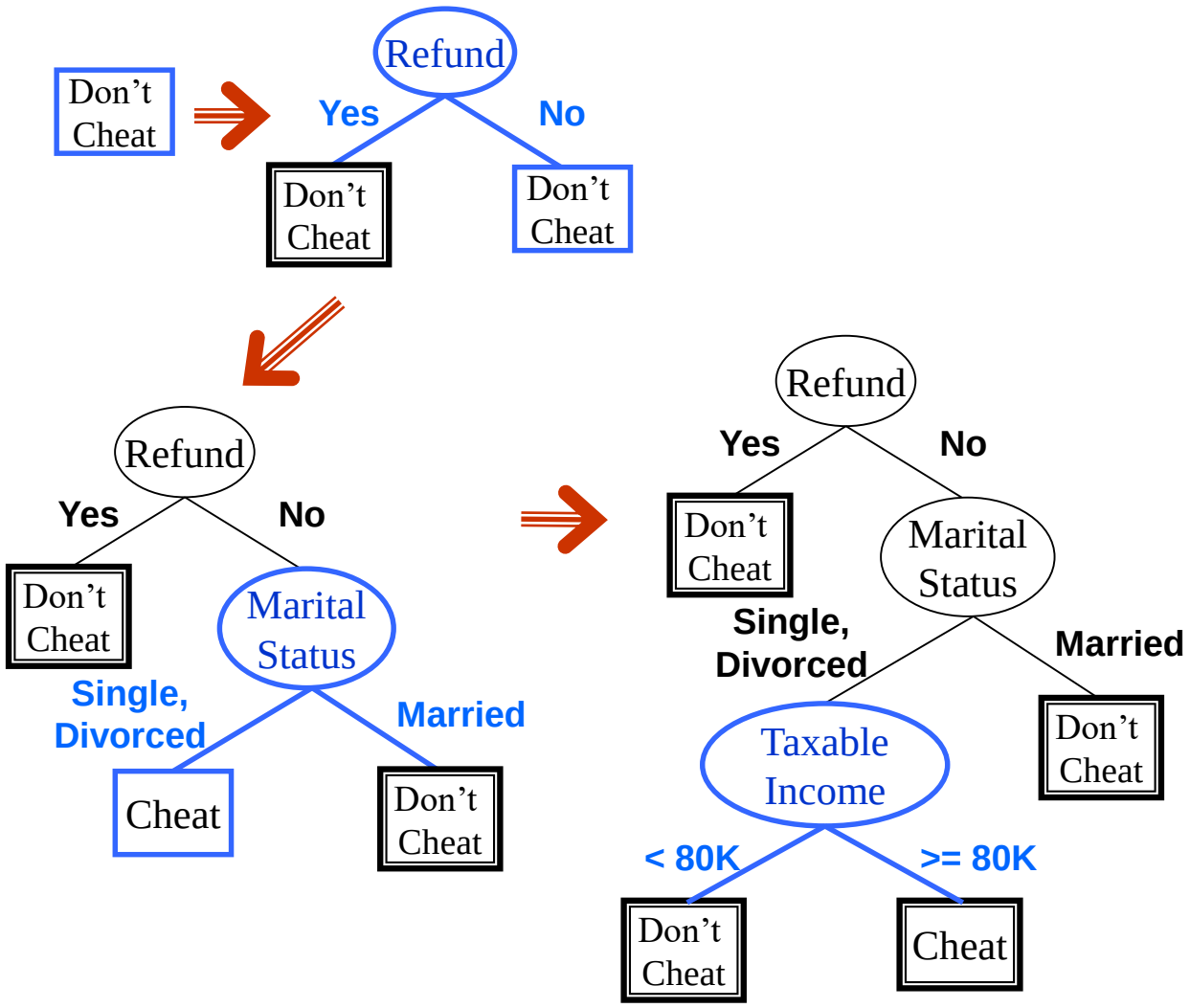
Opšta struktura Hantovog algoritma

- Neka je D_t skup slogova za trening koji se nalaze u čvoru t
- Opšta procedura:
  - Ako D_t sadrži slogove koji pripadaju istoj klasi y_t, tada je t list označen sa y_t
  - Ako je D_t prazan skup tada je t list označen sa predefinisanom klasom y_d
  - Ako D_t sadrži slogove koji se nalaze u više od jedne klase, tada se koristi test atribut radi podele podataka u manje podskupove. Ova procedura se rekurzivno primenjuje na svaki podskup.

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



Hantov algoritam



Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

Indukcija po stablu

- Strategija
  - Podeliti slogove prema testnom atributu koji optimizuje određeni kriterijum.

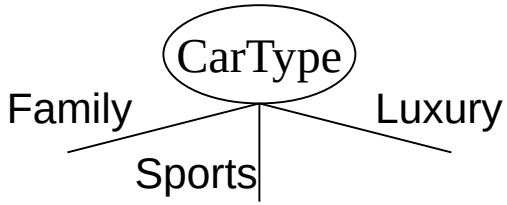
- Odluke koje treba doneti
  - Kako podeliti slogove
    - Kako navesti uslove testiranja za atribute?
    - Kako odrediti najbolju podelu?
  - Kada treba stati sa podelom

Kako navesti uslove testiranja za atribute?

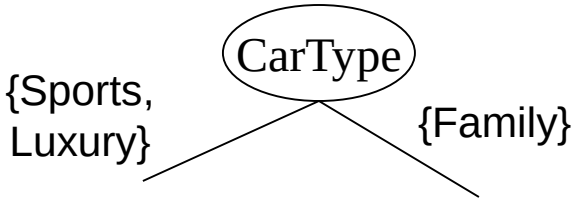
- Zavisnost od tipa atributa
  - Imenski (Nominalni)
  - Redni (Ordinarni)
  - Neprekidni
- Zavisnost od broja načina za deobu
  - Podela na 2 grane
  - Podela na više grana

Podela zasnovana na imenskim atributima

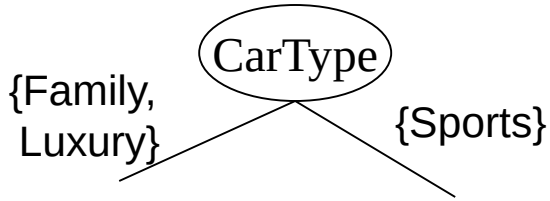
- Podela na više grana: koristi se toliko delova koliko ima različitih vrednosti



- Binarna podela: vrednosti se dele u dva podskupa. Treba naći optimalnu podelu.

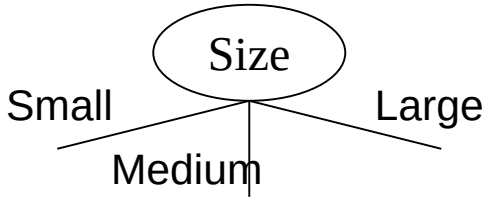


OR

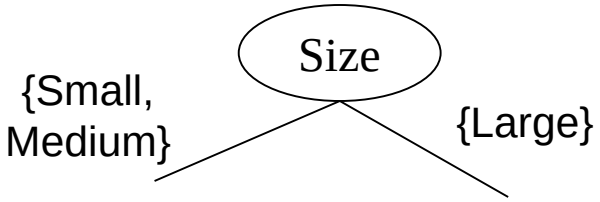


Podela zasnovana na rednim atributima

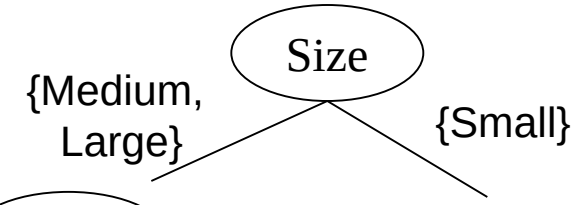
- Podela na više grana: koristi se toliko delova koliko ima različitih vrednosti.

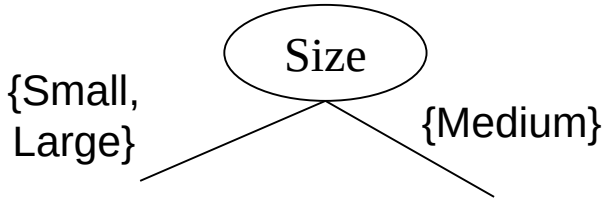


- Binarna podela: vrednosti se dele u dva podskupa. Treba naći optimalnu podelu.



OR



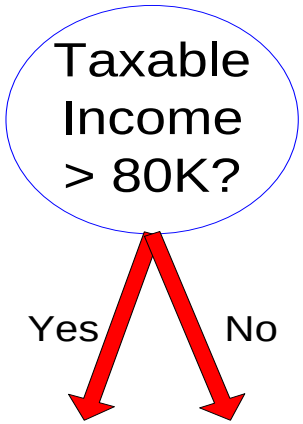


- Šta sa ovom podelom?

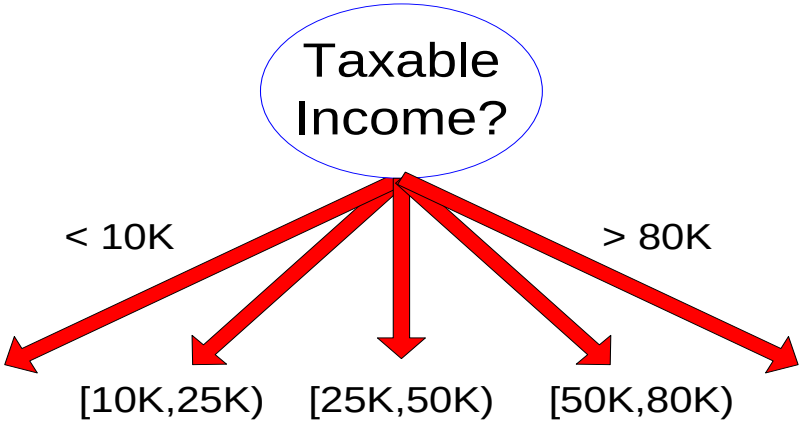
Podela zasnovana na neprekidnim atributima

- Različiti načini rada
  - Diskretizacijom se formiraju redni kategorički atributi
    - Statički – diskretizacija jednom na početku rada
    - Dinamički – opsezi mogu da se odrede podelom na jednake intervale, jednaku frekvenciju, precentile, klastere, ...
  - Binarna podela: $(A < v)$ ili $(A \geq v)$
    - razmatraju se sve moguće podele i pronalazi najbolja
    - računarski intenzivan posao

Podela zasnovana na neprekidnim atributima



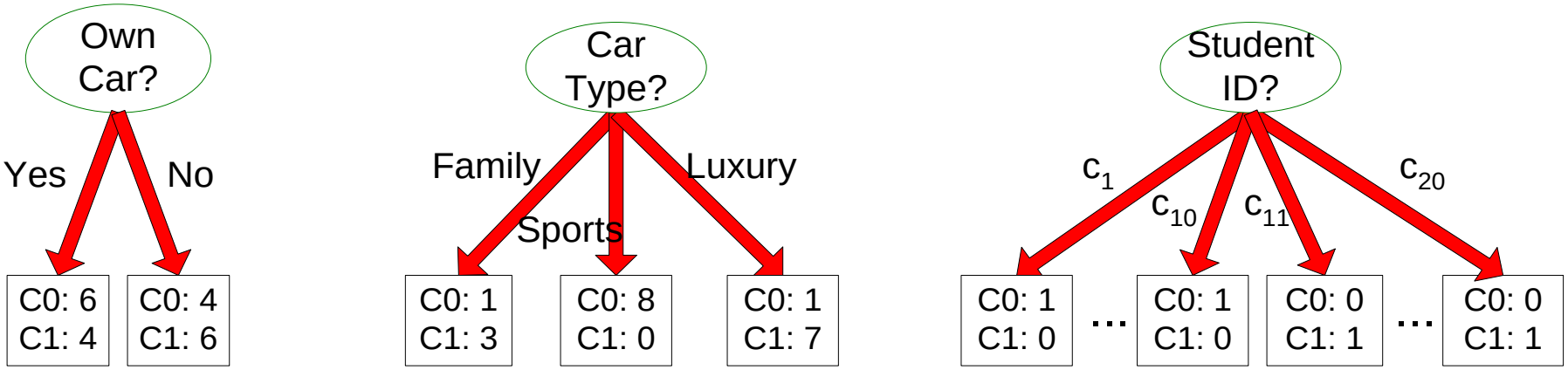
(i) Binary split



(ii) Multi-way split

Kako odrediti najbolju podelu

Pre podele : 10 slogova klase 0,
10 slogova klase 1



Koji testni uslov daje najbolje rezultate?

Kako odrediti najbolju podelu

- Greedy algoritam:
  - Prvenstvo imaju čvorovi sa homogenom distribucijom klasa
- Potrebno je naći meru nečistoće čvora:

C0: 5 C1: 5	C0: 9 C1: 1
Nehomogeno,	Homogeno,
Visok nivo nečistoće	Nizak nivo nečistoće

Mera nečistoće čvora

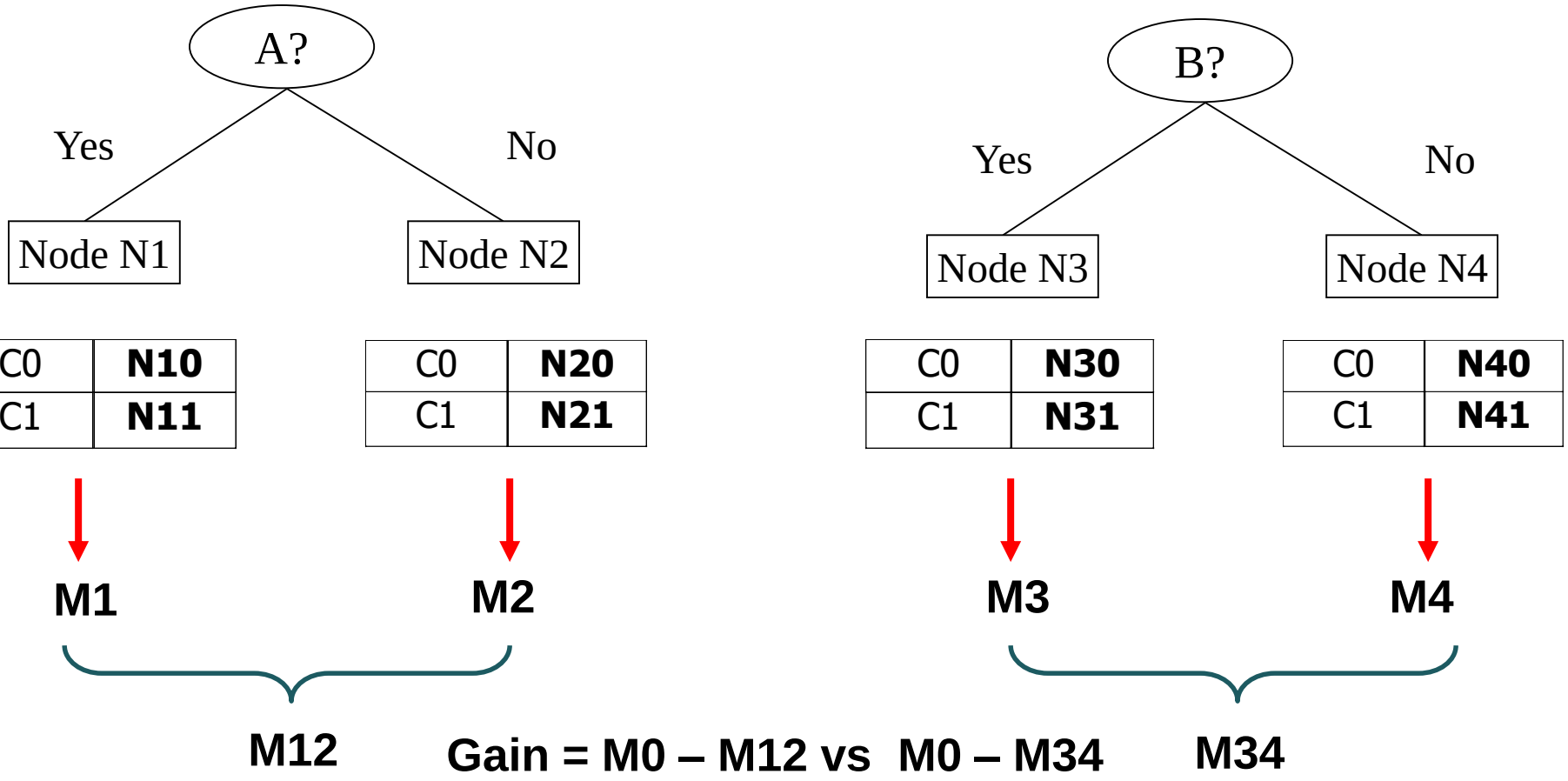
- Ginijev indeks (Gini)
- Entropija
- Greške u klasifikaciji

Kako naći nabolju podelu?

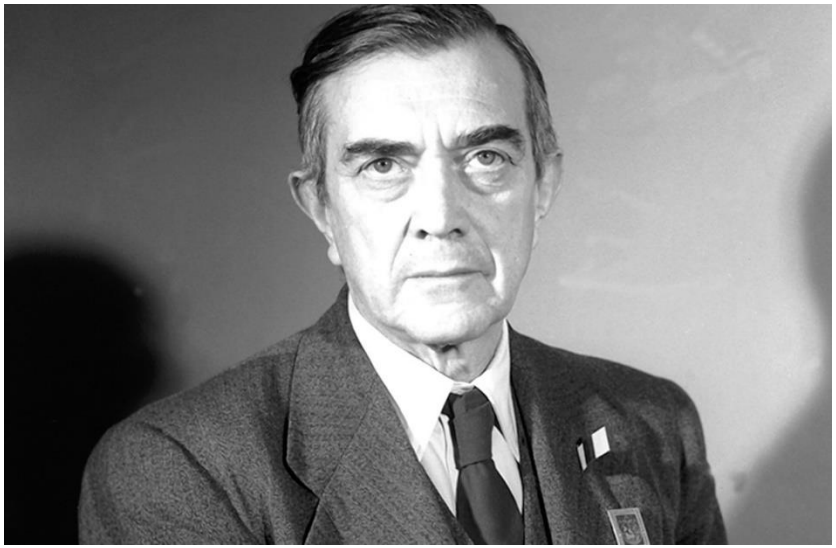
Pre podele:

C0	**N00**
C1	**N01**

→ **M0**



GINI index



Corrado Gini
(1884 – 1965)

Mera nečistoće : GINI

- Ginijev (Corrado Gini, italijanski statističar) indeks za dati čvor t :

$$GINI(t) = 1 - \sum_j [p(j \mid t)]^2$$

(Primedba: $p(j \mid t)$ je relativna frekvencija klase j u čvoru t).

  - Maksimum (1 - $1/n_c$) kada su slogovi ravnomerno distribuirani u svim klasama podrazumeva najmanje interesantne informacije
  - Minimum (0.0) kada svi slogovi pripadaju jednoj klasi, podrazumeva najinteresantije informacije

C1	**0**
C2	**6**
Gini=0.000	

C1	**1**
C2	**5**
Gini=0.278	

C1	**2**
C2	**4**
Gini=0.444	

C1	**3**
C2	**3**
Gini=0.500	

Primeri izračunavanja GINI

$$GINI(t) = 1 - \sum_j [p(j \mid t)]^2$$

C1	**0**
C2	**6**

P(C1) = 0/6 = 0 P(C2) = 6/6 = 1

Gini = 1 – P(C1)^2 – P(C2)^2 = 1 – 0 – 1 = 0

C1	**1**
C2	**5**

P(C1) = 1/6 P(C2) = 5/6

Gini = 1 – (1/6)^2 – (5/6)^2 = 0.278

C1	**2**
C2	**4**

P(C1) = 2/6 P(C2) = 4/6

Gini = 1 – (2/6)^2 – (4/6)^2 = 0.444

Podela zasnovana na GINI

- Koristi se u CART, SLIQ, SPRINT.
- Kada se čvor p deli u k delova (dete čvor) kvalitet se računa kao,

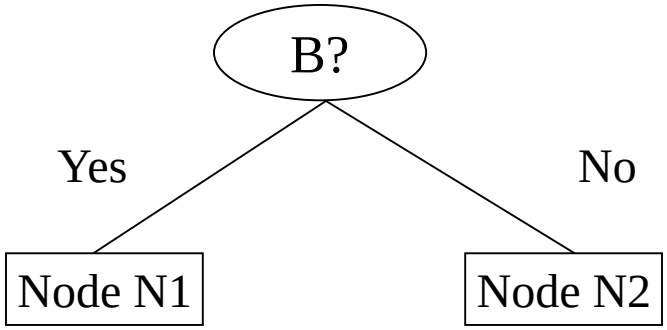
$$GINI_{split} = \sum_{i=1}^{k} \frac{n_i}{n} GINI(i)$$

gde je n_i = broj slogova u dete čvoru i,

n = broj slogova u čvoru p.

Binarni atributi: izračunavanje GINI indeksa

- Skup se deli u dve particije
- Efekti težina particija: poželjne su veće i čistije particije



	Parent
C1	**6**
C2	**6**
Gini = 0.500	

Gini(N1)
= 1 – (5/7)^2 – (2/7)^2
= 0.4082

Gini(N2)
= 1 – (1/5)^2 – (4/5)^2
= 0.320

	N1	N2
C1	**5**	**1**
C2	**2**	**4**
Gini=0.333		

Gini(Children)
= 7/12 * 0.4082 +
5/12 * 0.320
= 0.3715

Kategorički atributi: izračunavanje Gini indeksa

- Za svaku od različitih vrednosti izračuna se broj u svakoj klasi skupa podataka
- U donošenju odluka koristi se matrica brojanja

Multi-way split

	CarType		
	Family	Sports	Luxury
C1	1	2	1
C2	4	1	1
Gini	0.393		

Two-way split
(find best partition of values)

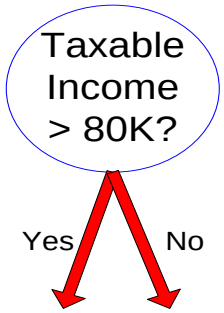
	CarType	
	{Sports, Luxury}	{Family}
C1	3	1
C2	2	4
Gini	0.400	

	CarType	
	{Sports}	{Family, Luxury}
C1	2	2
C2	1	5
Gini	0.419	

Kategorički atributi: izračunavanje Gini indeksa

- Koriste se binarne pitalice zasnovane na jednoj vrednosti
- Više izbora za vrednost po kojoj se deli
  - Broj mogućih vredosti za podelu = broju različitih vrednosti
- Svaka vrednost po kojoj se deli ima pridruženu matricu brojanja
  - U svakoj od particija se prebrojavaju klase, $A < v$ i $A \geq v$
- Jednostavan način za izbor najboljeg v
  - Za svako v, skenirati bazu podataka da bi se dobila matrica brojeva i izračunao Hinijev indeks
  - Zahteva ponavljanje posla i neefikasno je sa stanovišta izračunljivosti

Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes



Neprekidni atributi: izračunavanje Gini indeksa

- Za efikasno izračunavanje se za svaki
  - Sortira atribut po vrednostima
  - Dobijene vrednosti linearno skeniraju uz ažuriranje matrice brojanja i izračunavanje Ginijevog indeksa
  - Bira se pozicija za podelu sa najmanjim Ginijevim indeksom

Cheat	No		No		No		Yes		Yes		Yes		No		No		No		No	
Taxable Income																				
Sortirane vrednosti →	60		70		75		85		90		95		100		120		125		220	
Pozicije podele →	55		65		72		80		87		92		97		110		122		172		230	
	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>	<=	>
Yes	0	3	0	3	0	3	0	3	1	2	2	1	3	0	3	0	3	0	3	0	3	0
No	0	7	1	6	2	5	3	4	3	4	3	4	3	4	4	3	5	2	6	1	7	0
Gini	0.420		0.400		0.375		0.343		0.417		0.400		***0.300***		0.343		0.375		0.400		0.420	

Alternativni kriterijumi podele

- Entropija u datom čvoru t:

$$Entropy(t) = -\sum_j p(j \mid t) \log p(j \mid t)$$

(Primedba: *p(j | t)* je relativna frekvencija klase j u čvoru t).

  - Mera homogenosti čvora
    - Maksimum (log n_c) kada su slogovi ravnomerno distribuirani u svim klasama podrazumeva najmanje informacija
    - Minimum (0.0) kada svi slogovi pripadaju jednoj klasi, podrazumeva najviše informacij

  - Izračunavanja zasnovana na entropiji i Ginijevom indeksu su slična

Primer računanja entropije

$$Entropy(t) = -\sum_j p(j \mid t) \log_2 p(j \mid t)$$

C1	**0**
C2	**6**

P(C1) = 0/6 = 0 P(C2) = 6/6 = 1

Entropy = – 0 log 0 – 1 log 1 = – 0 – 0 = 0

C1	**1**
C2	**5**

P(C1) = 1/6 P(C2) = 5/6

Entropy = – (1/6) $\log_2$ (1/6) – (5/6) $\log_2$ (1/6) = 0.65

C1	**2**
C2	**4**

P(C1) = 2/6 P(C2) = 4/6

Entropy = – (2/6) $\log_2$ (2/6) – (4/6) $\log_2$ (4/6) = 0.92

Alternativni kriterijumi podele

- Sticanje informacija (eng. *information gain*):

$$GAIN_{split} = Entropy(p) - \left(\sum_{i=1}^{k} \frac{n_i}{n} Entropy(i) \right)$$

gde se roditelj čvor p se deli u k particija, a n_i je broj slogova u particiji i

  - Mera redukcije u entropiji se ostvaruje zbog podele. Podela se bira tako da se dobija najveća redukcija (maksimizira GAIN)
  - Ovaj način se koristi u ID3 i C4.5
  - Nedostaci: ima tendenciju da formira veliki broj malih ali čistih particija

Alternativni kriterijumi podele

- Gain odnos se koristi za određivanje valjanosti podele
- U C4.5 se kao kriterijum valjanosti koristi

$$GainRATIO_{split} = \frac{GAIN_{Split}}{SplitINFO}$$

$$SplitINFO = -\sum_{i=1}^{k} \frac{n_i}{n} \log \frac{n_i}{n}$$

gde se roditelj čvor p se deli u k particija, a n_i je broj slogova u particiji i

Kriterijumi deobe zasnovani na greškama pri klasifikaciji

- Greška klasifikacije u čvoru t :

$$Error(t) = 1 - \max_i P(i \mid t)$$

- Mera greške pri klasifikaciji
  - Maksimum (1 - 1/n_c) kada su slogovi ravnomerno distribuirani u svim klasama podrazumeva najmanje interesantne informacije
  - Minimum (0.0) kada svi slogovi pripadaju jednoj klasi, podrazumeva najinteresantije informacije

Primer greške pri izračunavanju

$$Error(t) = 1 - \max_i P(i \mid t)$$

C1	**0**
C2	**6**

P(C1) = 0/6 = 0 **P(C2) = 6/6 = 1**

Error = 1 – max (0, 1) = 1 – 1 = 0

C1	**1**
C2	**5**

P(C1) = 1/6 **P(C2) = 5/6**

Error = 1 – max (1/6, 5/6) = 1 – 5/6 = 1/6

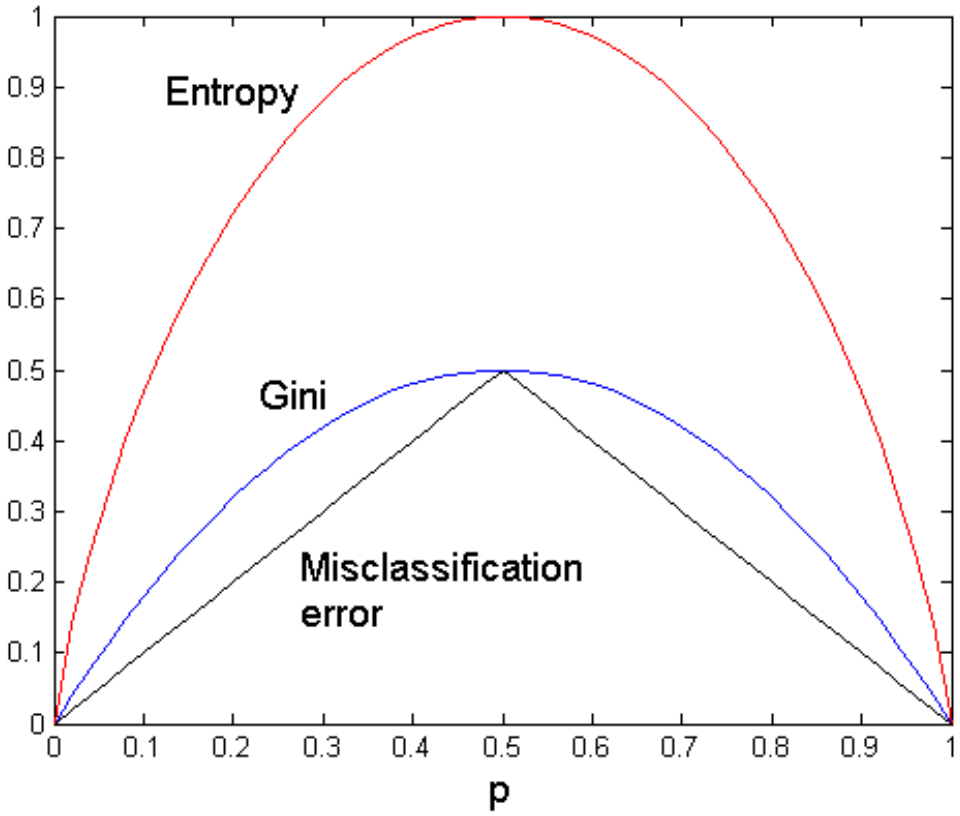
C1	**2**
C2	**4**

P(C1) = 2/6 **P(C2) = 4/6**

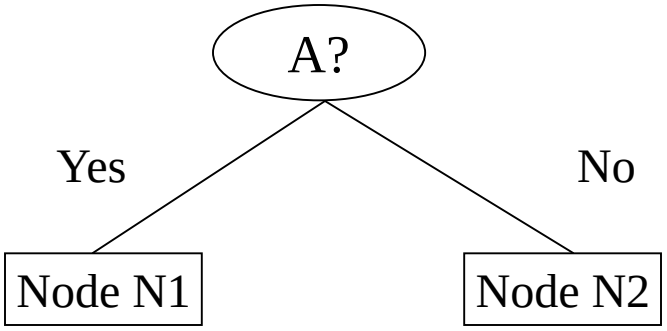
Error = 1 – max (2/6, 4/6) = 1 – 4/6 = 1/3

Poređenje među kriterijumima podele

Za problem 2-klase:



Greške u klasifikaciji / Gini



	Parent
C1	**7**
C2	**3**
Gini = 0.42	

Gini(N1)
= 1 – (3/3)² – (0/3)²
= 0

Gini(N2)
= 1 – (4/7)² – (3/7)²
= 0.489

	N1	N2
C1	**3**	**4**
C2	**0**	**3**
Gini=0.361		

Gini(Children)
= 3/10 * 0
+ 7/10 * 0.489
= 0.342

Gini daje poboljšanje !!

Kriterijum zaustavljanja indukcije po stablu

- Širenje se zaustavlja kada svi slogovi pripadaju istoj klasi
- Širenje se zaustavlja kada svi slogovi imaju iste vrednosti atributa

Klasifikacija zasnovana na stablima odlučivanja

- Prednosti:
  - Jednostavna konstrukcija
  - Brza klasifikacija neklasifikovanih instanci
  - Laka za interpertaciju za stabla male veličine
  - Preciznost je uporediva sa ostalim tehnikama klasifikacije za jednostavne tipove podataka

Praktični problemi pri klasifikaciji

- Obučavanje
- Nedostajuće vrednosti
- Cena klasifikacije

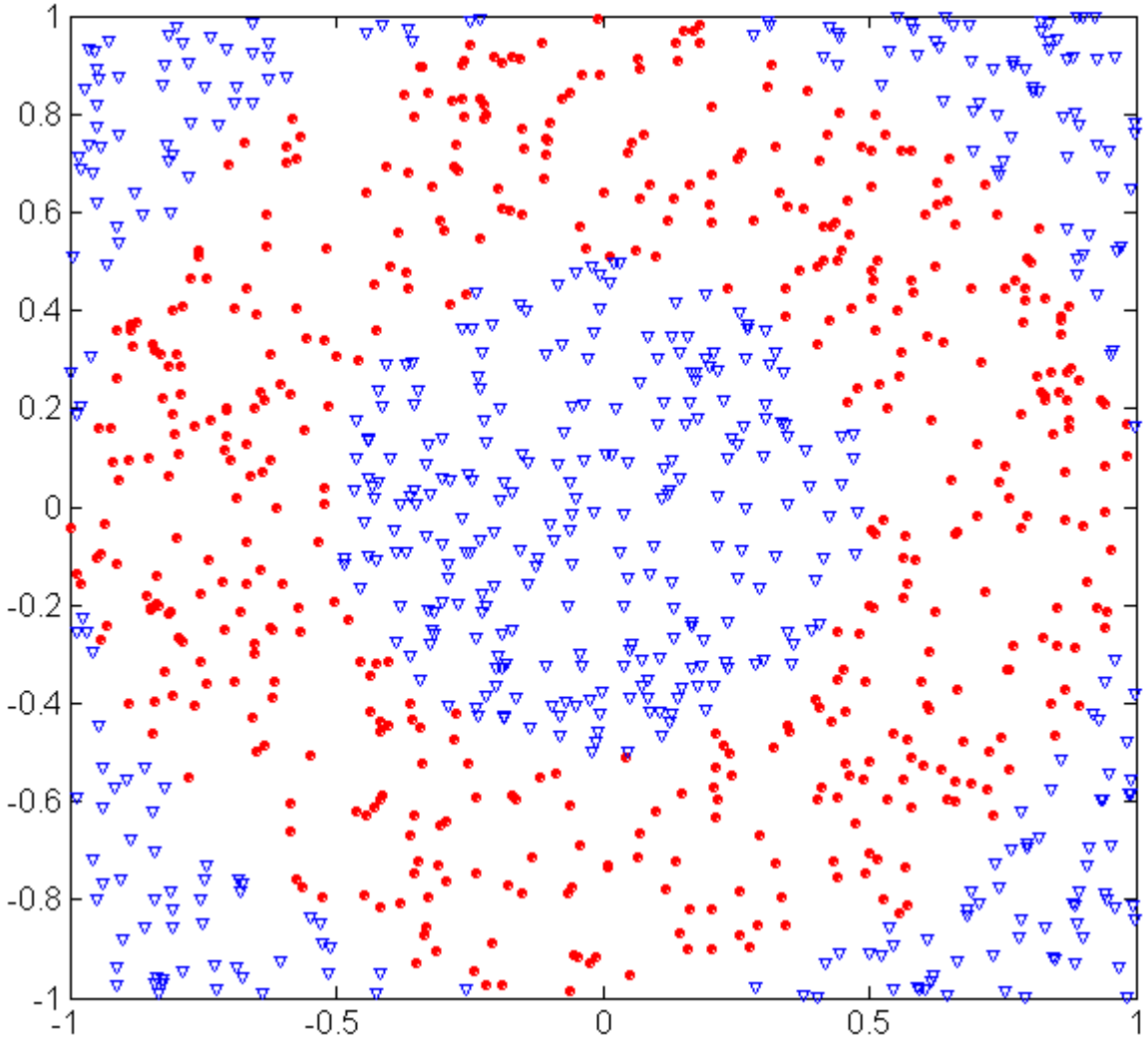
Fitovanje modela

- Greške u klasifikaciji
  - Greške pri treniranju (greške resubstitucije).
    - Broj grešaka u klasifikaciji za dati skup podataka za trening
  - Greške genetalizacije
    - Očekivana greška modela u odnosu na unapred nepoznate primere
- Dobar model mora korektno da klasifikuje i trening podatke i unapred nepoznate primere

Fitovanje modela

- Model koji isuviše dobro klasifikuje podatke za trening može da ima lošije karakteristike pri generalizaciji od modela koji ima veću grešku u procesu treninga – *previše prilagođen model* (eng. model overfitting)
  - U daljem tekstu preprilagođen model
- Ako je model isuviše jednostavan i greška pri treniranju i greška generalizacije mogu da budu jako velike – *premalo prilagođen model* (eng. model underfitting)
  - U daljem tekstu potprilagođen model

Preprilagođen i potprilagođen



500 kružnih i 500 trouglastih tačaka

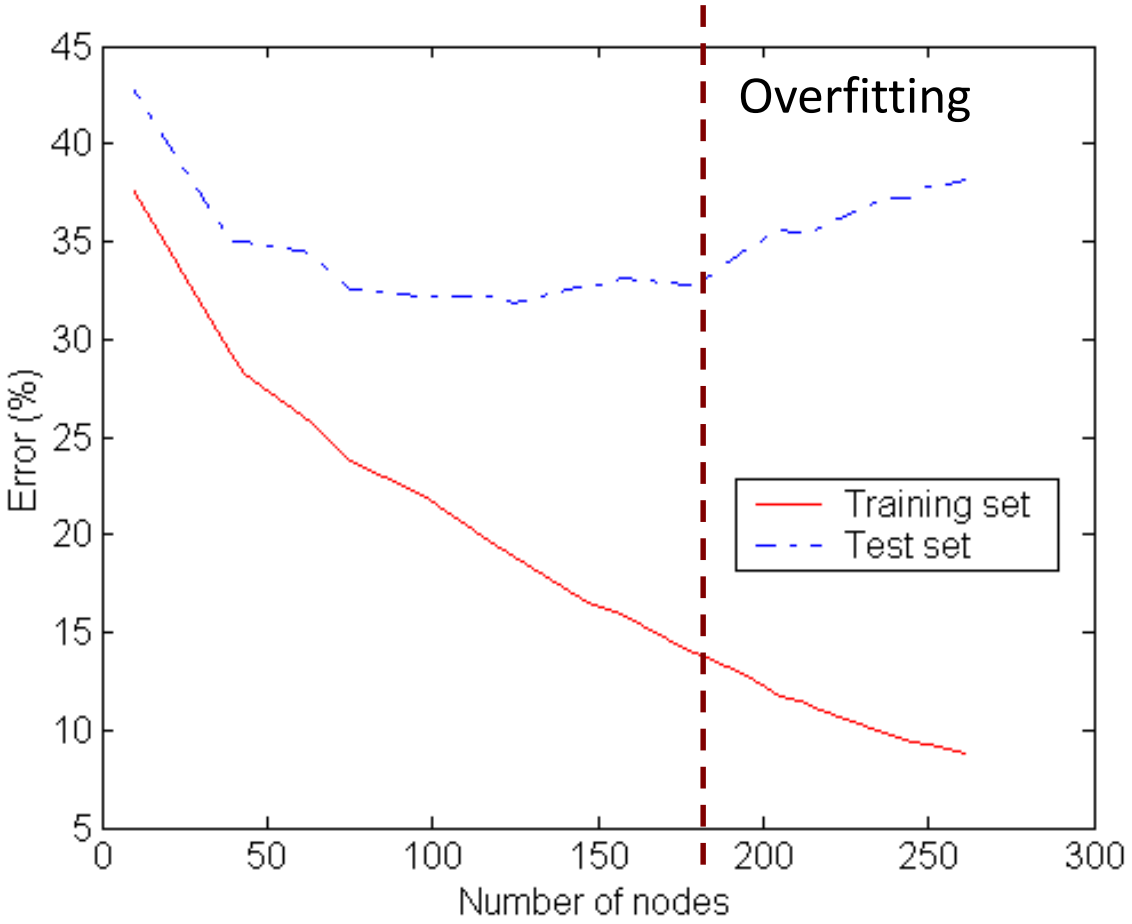
Kružne tačke: $0.5 \leq \text{sqrt}(x_1^2+x_2^2) \leq 1$

Trougaone tačke: $\text{sqrt}(x_1^2+x_2^2) > 0.5$ ili $\text{sqrt}(x_1^2+x_2^2) < 1$

30% tačaka se bira za trening, ostale za test

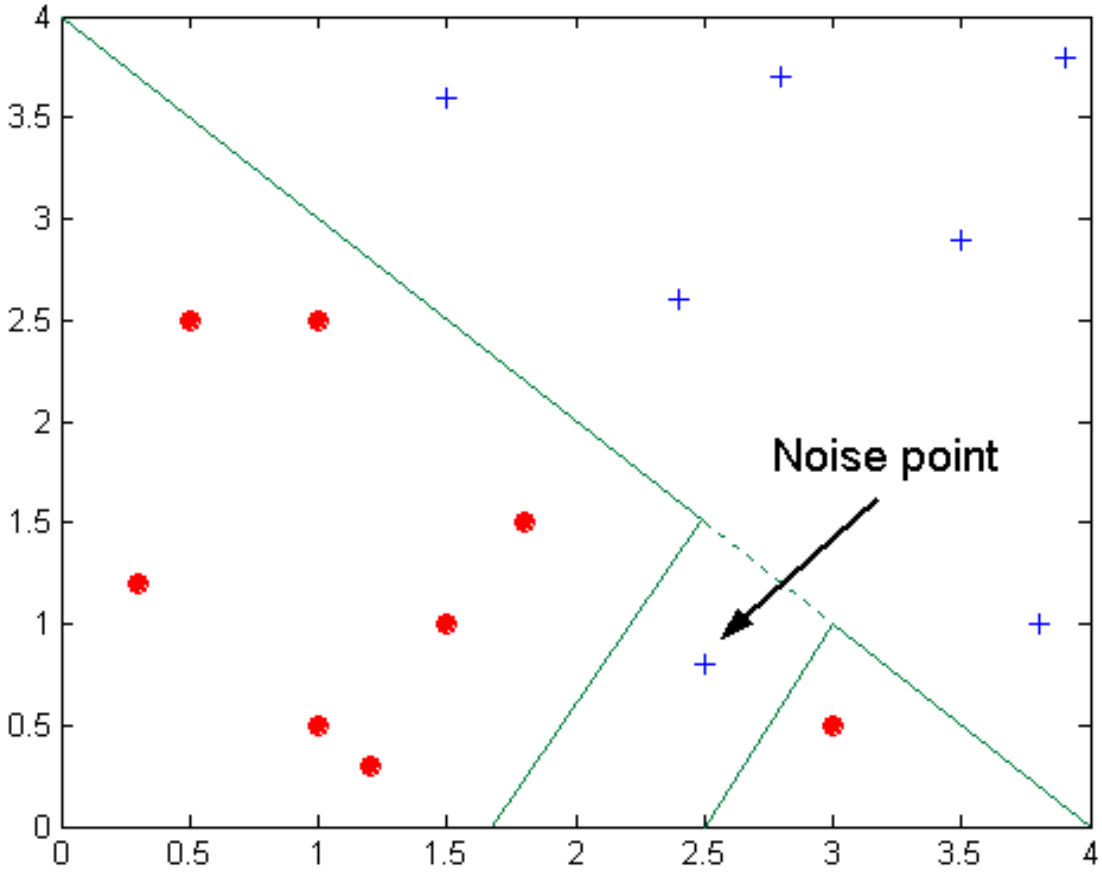
Na trening podatke se primenjuju stabla sa Ginijevim indeksom kao merom nečistoće

Preprilagođen i potprilagođen



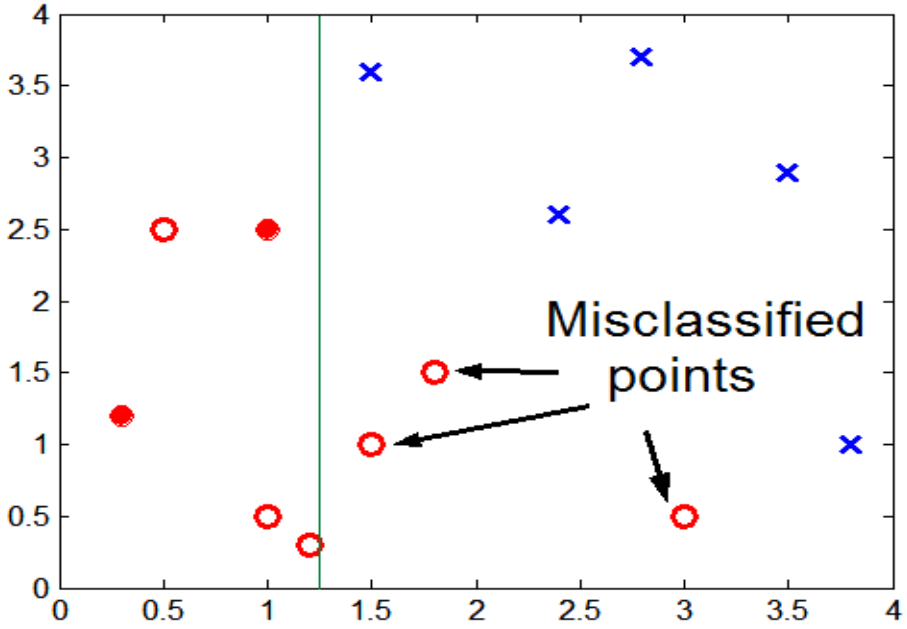
Potkresivanjem stabla na različitim nivoima dobijaju se različite veličine grešaka

Preprilagođenost zbog šuma



Granice podele se deformišu zbog postojanja šuma

Preprilagođenost zbog nepostojanja reprezentativnih primera



Modeli koji formiraju kriterijum klasifikacije na osnovu malog skupa za trening su podložni preprilagođenosti. Npr. nedostatak tačaka u donjoj polovini dijagrama onemogućava korektno predviđanje oznaka klasa u tom delu. U procesu klasifikacije se koriste ostali trening primeri koji su irelelvantni za klasifikaciju u tom delu.

Preprilagođenost: neki komentari

- Preprilagođenost se javlja kod stabla odlučivanja koja su kompleksnija nego što je potrebno
- U tom slučaju greške pri treniranju ne daju korektnu procenu načina ponašanja stabla u slučaju pojave prethodno nepoznatih podataka
- Zahteva nove načine procene greške

Procena greške u generalizaciji

- Neka je T stablo, t čvor, N broj listova u stablu T, e(t) broj pogrešno klasifikovanih slogova u t, e(T) ukupan broj grešaka u klasifikaciji po stablu T

- Greška resubstitucije: greška pri treniranju (Σ e(t))
- Greška pri generalizaciji: greška pri testiranju (Σ e'(t))
- Metode za procenu greške pri generalizaciji:
  - Optimistički pristup: e'(t) = e(t)

Procena greške u generalizaciji

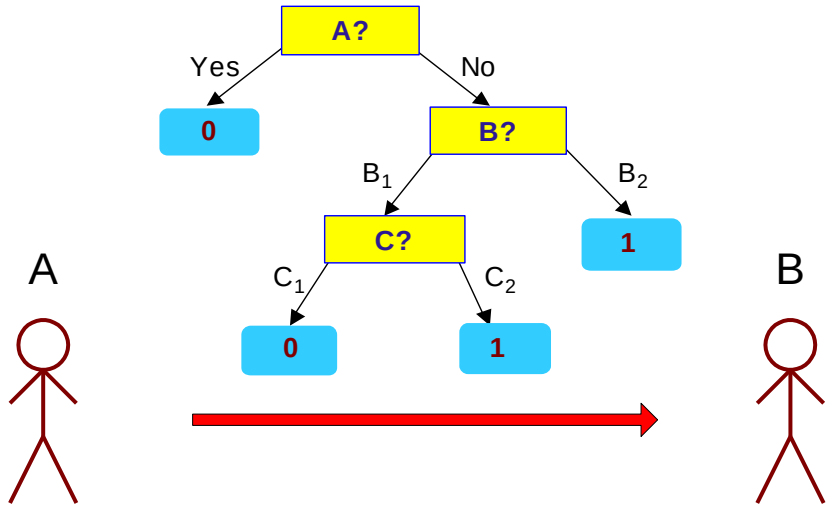
- Metode za procenu greške pri generalizaciji:
  - Optimistički pristup: $e'(t) = e(t)$
  - Pesimistički pristup:
    - Za svaki list: $e'(t) = (e(t)+0.5)$
    - Ukupan broj grešaka: $e'(T) = e(T) + N \times 0.5$
    - Za stabl sa 30 listova i 10 grešaka na treningu sa 1000 stavki:
      Greška treninranja = 10/1000 = 1%
      Greška uopštavanja = $(10 + 30 \times 0.5)/1000 = 2.5\%$

Princip škrtosti

- Naziva se još i Okamov (eng. Occam) rezač
- Od dva modela sa sličnom greškom generalizacije treba izabrati onaj koji je jednostavniji
- Kod složenijih modela veća je šansa za pogrešnim fitovanjem ukoliko u podacima postoje greške
- Pri proceni modela treba uključiti i njegovu složenost

Princip najmanje dužine opisa (MDL)

X	y
X_1	1
X_2	0
X_3	0
X_4	1
…	…
X_n	1



X	y
X_1	?
X_2	?
X_3	?
X_4	?
…	…
X_n	?

- Cena (model,podaci) = Cena(podaci|model) + Cena(model)
  - Cena je broj bitova potreban za kodiranje.
  - Traži se najmanje skup model.
- Cena(podaci|model) kodira pogrešno označene slogove pri klasifikaciji.
- Cena(model) koristi kodiranje modela (čvorovi + uslov podele)

Kontrola overfitinga

- Pre-potkresivanje (pravilo ranijeg zaustavljanja)
  - Algoritam se zaustavlja pre nego što stabl naraste do maksimalne veličine
  - tipični uslovi zaustavljanja za određeni čvor su:
    - Zaustavi se ako sve instance pripadaju istoj klasi
    - Zaustavi se ako su sve vrednosti atributa iste
  - Dodatna ograničenja:
    - Zaustavi se ako je broj instanci manji od neke unapred zadate granice
    - Zaustavi se ako je distribucija instanci nezavisna od raspoloživih osobina (npr. vidi se primenom χ^2 testa)
    - Zaustavi se ako širenje tekućeg čvora ne poboljšava meru čistoće (npr. Gini ili informaciono pojačanje (*information gain*)).

Kontrola overfitinga

- Potkresivanje po završetku
  - stablo odlučivanja raste do krajnjih granica
  - Iseku se čvorovi u stablu od dna ka vrhu
  - Ako se greška generalizacije poboljša posle otsecanja podstablo se zameni sa listom.
  - Labele kalse lista se određuju prema dominantnoj klasi instanci podstabla
  - Za potkresivanje po završetku se može koristiti MDL (*minimum description length*)

Primer potkresivanja po završetku

Class = Yes	20
Class = No	10
Error = 10/30	

Trening greška (pre deobe) = 10/30

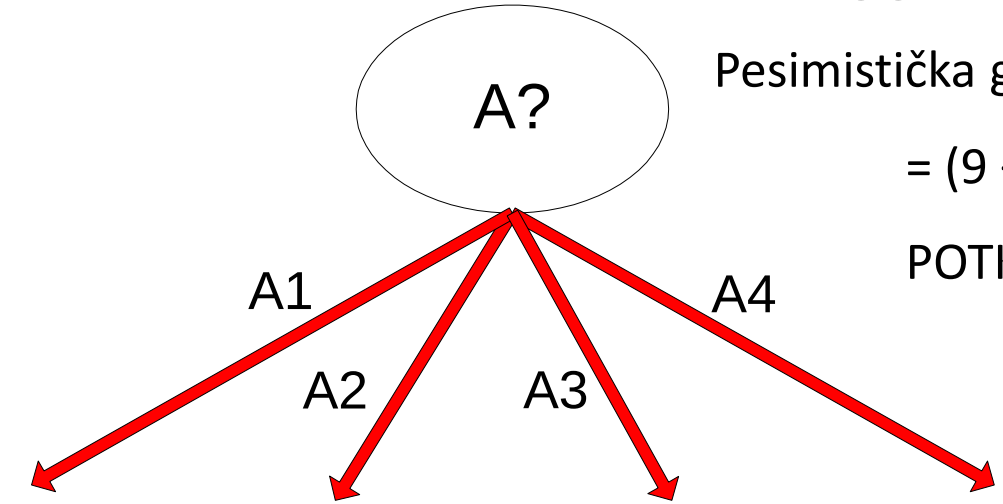
Pesimistička greška = (10 + 0.5)/30 = 10.5/30

Trening greška (posle deobe) = 9/30

Pesimistička greška (posle deobe)

= (9 + 4 × 0.5)/30 = 11/30

POTKRESATI stablo!



Class = Yes	8
Class = No	4

Class = Yes	3
Class = No	4

Class = Yes	4
Class = No	1

Class = Yes	5
Class = No	1

Rukovanje atributima sa nedostajućim vrednostima

- Nedostajuće vrednosti utiču na stablo odlučivanja na različite načine:
  - Kako računati meru nečistoće
  - Kako distribuirati instance sa nedostajućim vredostima na decu čvorove
  - Kako klasifikovati test instancu sa nedostajućom vrednošću

Izračunavanje mere nečistoće

Tid	Refund	Marital Status	Taxable Income	Class
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	?	Single	90K	Yes

Nedostajuća vredost

Pre podele:

Entropy(Parent)
= -0.3 log(0.3)-(0.7)log(0.7) = 0.8813

	Class = Yes	Class = No
Refund=Yes	**0**	**3**
Refund=No	**2**	**4**
Refund=?	**1**	**0**

Podela na REFUND:

Entropy(Refund=Yes) = 0

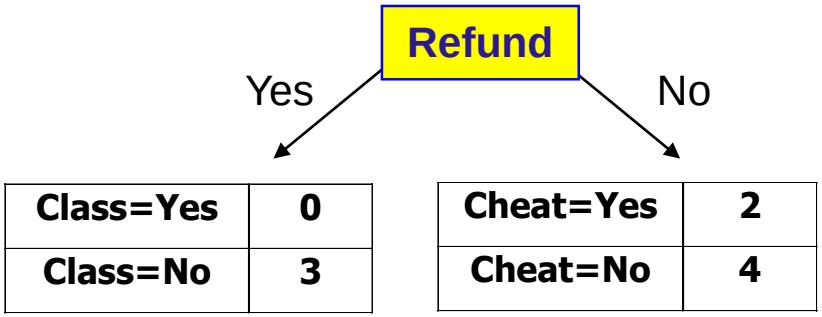
Entropy(Refund=No)
= -(2/6)log(2/6) – (4/6)log(4/6) = 0.9183

Entropy(Children)
= 0.3 (0) + 0.6 (0.9183) = 0.551

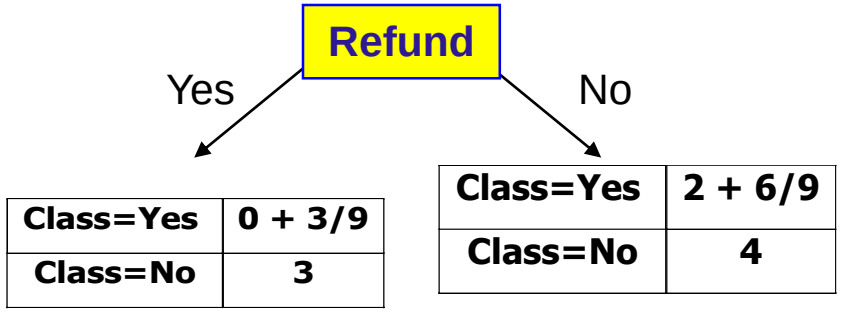
Gain = 0.9 × (0.8813 – 0.551) = 0.3303

Ditribucija instanci

Tid	Refund	Marital Status	Taxable Income	Class
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No



Tid	Refund	Marital Status	Taxable Income	Class
10	?	Single	90K	Yes



Verovatnoća da je Refund=Yes is 3/9

Verovatnoća da je Refund=No is 6/9

Dodeliti slog levom detetu sa težinom = 3/9 i desnom detetu sa težinom = 6/9