

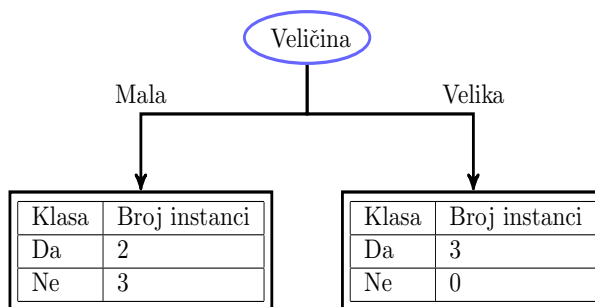
## Rešenje zadatka sa drvetom odlučivanja - vežbe 4

- Na osnovu datih podataka o životinjama iz trening skupa proceniti da li je životinja osobinama (*Velika*, *Biljke*, *Da*) opasna ili ne korišćenjem stabla odlučivanja dubine 2 uz korišćenje Ginijevog indeksa.

Velicina	Ishrana	Otrovnost	Opasna
Velika	Meso	Ne	Da
Mala	Meso	Ne	Ne
Mala	Biljke	Ne	Ne
Velika	Meso	Da	Da
Mala	Meso	Da	Da
Mala	Biljke	Ne	Ne
Mala	Biljke	Da	Da
Velika	Biljke	Ne	Da

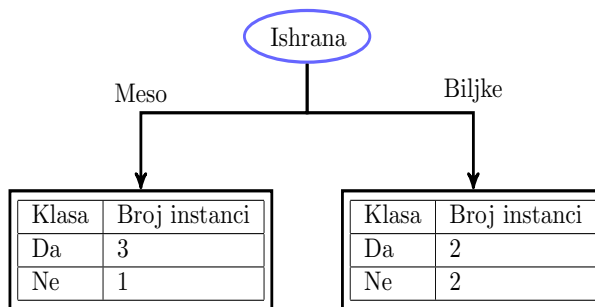
- Tražimo atribut koji je najbolji za podelu u korenu drveta odlučivanja.

– Podela po atributu **Velicina**



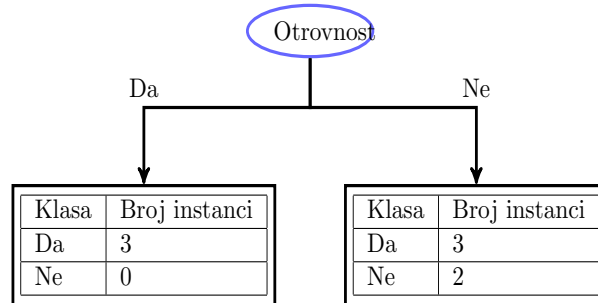
$$Gini_{split} = \frac{3}{8} \times (1 - (\frac{3}{3})^2 - (\frac{0}{3})^2) + \frac{5}{8} \times (1 - (\frac{2}{5})^2 - (\frac{3}{5})^2) = \frac{3}{10} = 0,3$$

– Podela po atributu **Ishrana**



$$Gini_{split} = \frac{4}{8} \times (1 - (\frac{1}{4})^2 - (\frac{3}{4})^2) + \frac{4}{8} \times \frac{1}{2} = \frac{7}{16} = 0,4375$$

– Podela po atributu **Otrovnost**



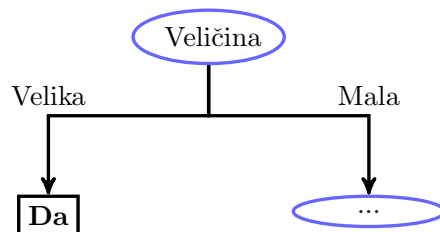
$$Gini_{split} = \frac{3}{8} \times (1 - (\frac{3}{3})^2 - (\frac{0}{3})^2) + \frac{5}{8} \times (1 - (\frac{3}{5})^2 - (\frac{2}{5})^2) = \frac{3}{10} = 0,3$$

Biramo atribut sa najmanjom vrednošću za  $Gini_{split}$ . Kako atributi **Velčina** i **Otrovnost** imaju istu vrednost za  $Gini_{split}$ , biramo jedan od njih, npr. **Velčina**.

Kako je čvor sa instancama koje imaju osobinu  $Velčina = Velika$  čist, tj. sve instance pripadaju istoj klasi (**Da**), dodeljujemo mu oznaku klase **Da**.

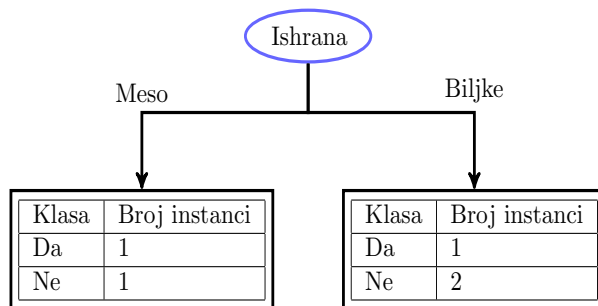
Kako čvor sa instancama koje imaju osobinu  $Velčina = Mala$  nije čist, tj. sve instance ne pripadaju istoj klasi, tražimo najbolji atribut za dalju podelu.

Nakon izbora atributa u korenu drveta, drvo odlučivanja izgleda:



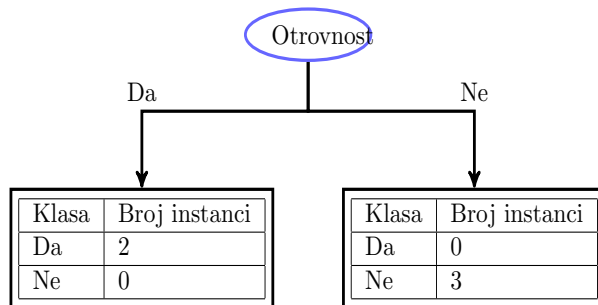
- Tražimo atribut za podelu instanci koje zadovoljavaju uslov  $Velčina = Mala$ .

– Podela po atributu **Ishrana**



$$Gini_{split} = \frac{2}{5} \times \frac{1}{2} + \frac{3}{5} \times (1 - (\frac{1}{3})^2 - (\frac{2}{3})^2) = \frac{7}{15}$$

– Podela po atributu **Otrovnost**



Kako su čvorovi koje smo dobili podelom čisti (sadrže instance samo jedne klase),  $Gini_{split}$  je 0.

Atribut *Otrovnost* je najbolji za dalju podelu. Pošto su čvorovi dobijeni podelom po izabranom atributu čisti (a i u zadatku se traži da napravimo drvo odlučivanja dubine 2) stajemo sa podelom.

Drvo odlučivanja nakon drugog koraka izgleda:

