**KAUNO TECHNOLOGIJOS UNIVERSITETAS INFORMATIKOS FAKULTETAS**

**Intelektikos pagrindai (P176B101)**

***Antro laboratorinio darbo ataskaita***

Atliko:

IFF-1/1 gr. Studentas

Vytenis Kriščiūnas

Priėmė:

lekt. Nečiūnas Audrius

lekt. Budnikas Germanas

**KAUNAS 2024**

**TURINYS**

1. Pasirinktas duomenų rinkinys 3

2. Pasirinkti atributai 3

3. Turimo duomenų rinkinio suskaidymas 3

4. Įvestys ir išvestys 4

5. Sprendimų medžio sudarymas 4

6. Grafinis sprendinių medžio atvaizdavimas 4

7. Sprendinių medžio testavimas 6

8. Sprendimų medžio gylio keitimas 6

9. Atsitiktinio miško sudarymas 7

10. Atsitiktinį mišką sudarančių medžių kiekio keitimas 7

11. Sprendinių medžio ir atsitiktinio miško palyginimas 7

# Pasirinktas duomenų rinkinys

Naudojau tą patį duomenų rinkinį, kaip ir pirmojo laboratorinio darbo metu su normalizuotais duomenimis.

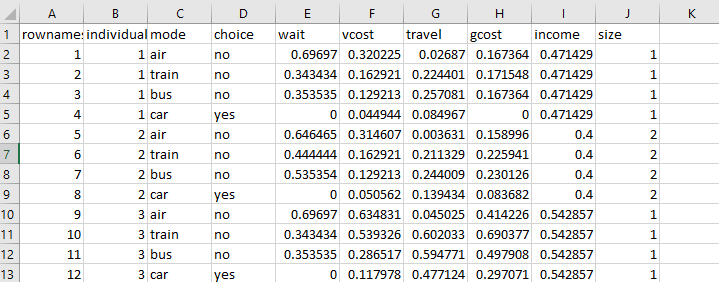
# Pasirinkti atributai

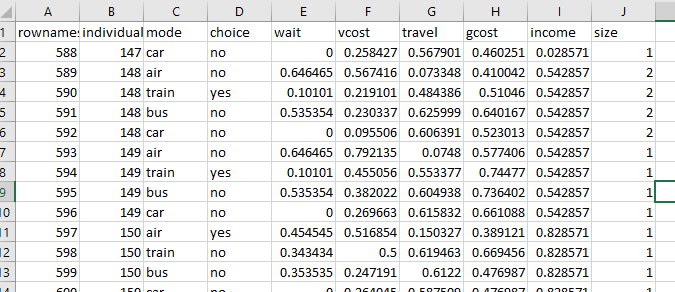
Įvesties atributai: „wait“ ir „travel“. Jie šie atributai padeda rasti geriausius prognozuojamus rezultatus.

Išvesties prognozuojamas atributas: „mode“, kurio kardinalumas lygus 4.

# Turimo duomenų rinkinio suskaidymas

Duomenų rinkinį suskaidžiau į 70-30% dalis, atitinkamai: apmokymo imtis ir testavimo imtis.





# Įvestys ir išvestys

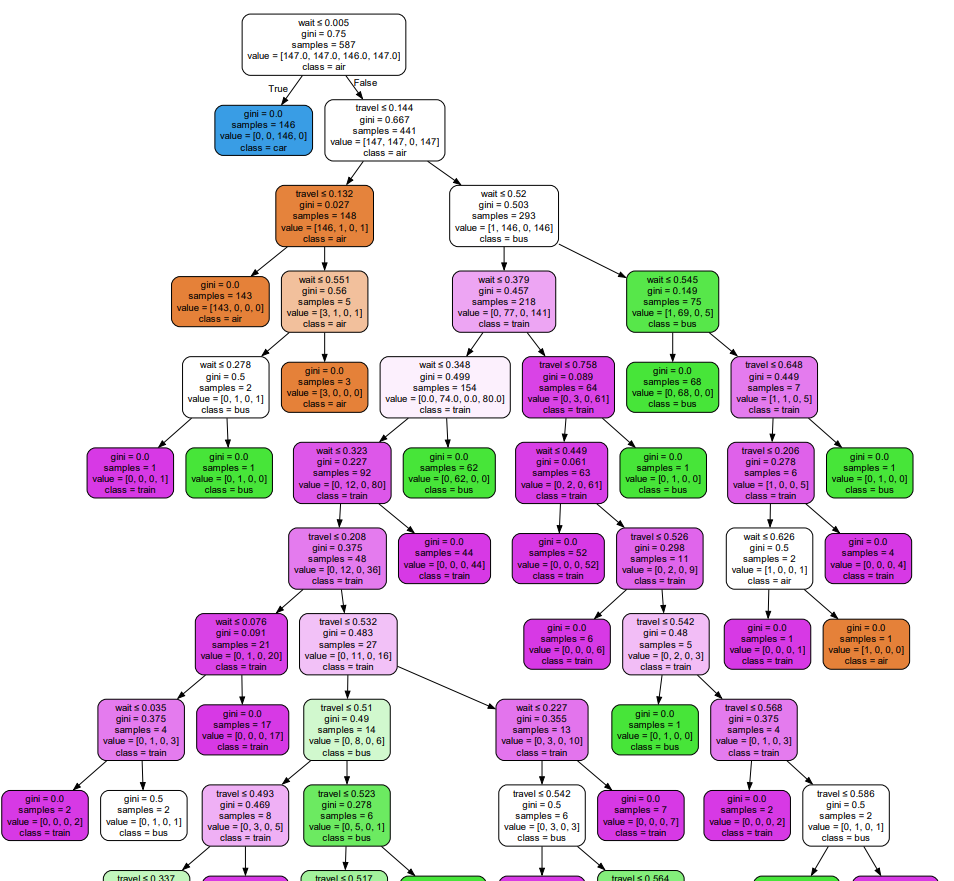
Atributo „mode“ galima variantai: *car, air, bus* ir *train.* Juos bandau nuspėti naudodamas tolydinius duomenis iš „wait“ ir „travel“ stulpelių.

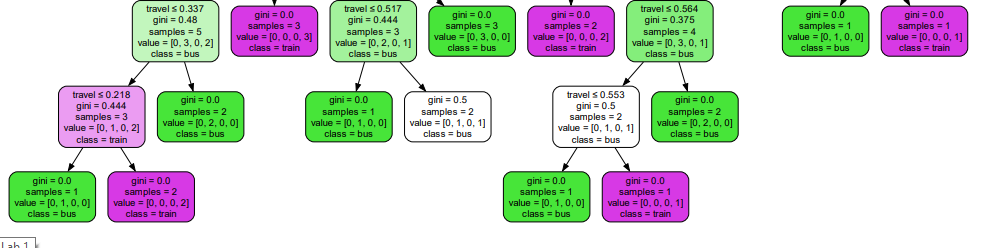
# Sprendimų medžio sudarymas

Naudojau CART algoritmą, nes ID3 algoritmas gali dirbti tik su kategoriniais duomenimis, o man reikia ir tolydinių duomenimų. Šis algoritmas naudoja Gini indeksą, kuris nurodo, kaip dažnai atsitiktinai pasirinkti elementai būtų neteisingai išdėstyti sprendimų medyje.

# Grafinis sprendinių medžio atvaizdavimas

Naudojau *graphviz* biblioteką sprendimų medžio formavimui.





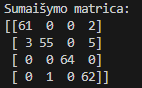
# Sprendinių medžio testavimas

Prognozavimo tikslumui nustatyti naudojau tikslumo metiką:

Prognozavimo tikslumas:

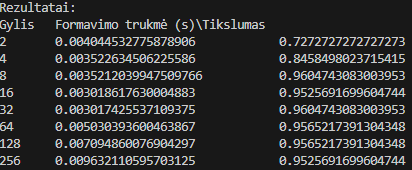


Sumaišymo matrica (stulpelis nurodo tikrąją atributo reikšmę, o eilutė spėjamas reikšmes):



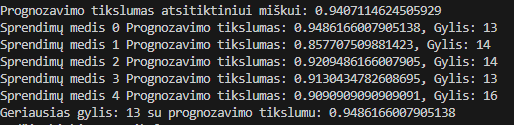
# Sprendimų medžio gylio keitimas

Galima pastebėti, kad gyliui augant auga ir jo formavimo trukmė. Tikslumas pagerėja pasiekus pakanamai dedidelį gylį ir praktiškai nusistovi, taigi labai didelis medžio gylis nesuteikia jokio pranašumo prognozavimo tikslumo prasme.



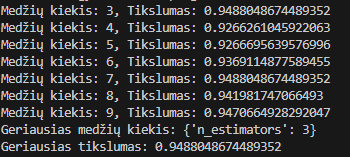
# Atsitiktinio miško sudarymas

Galima pastebėti, kad atsitiktinio miško prognozės tikslumas yra šiek tiek mažesnis nei vieno iš sprendimų medžio, taip yra todėl, kad medžių kiekis bei mažesnio tikslumo medžiai gali nulemti neteisingų rezultatų priėmimą (vote). Vis dėlto gautas tikslumas panaudojus atisitiktinį mišką dažnai yra didesnis už pavienius medžius.



# Atsitiktinį mišką sudarančių medžių kiekio keitimas

Mažiausiai medžių turintis atsitiktinis miškas – 3, gavo geriausią prognozės rezultatą. Galima daryti išvadą, kad atsitiktinai suformuotų medžių kiekis ir jų rūšis lemia atsitiktinio miško prognozės tikslumą.



# Sprendinių medžio ir atsitiktinio miško palyginimas

Sprendinių medžio tikslumas šiuo atvėju gavosi geresnis nei atsitiktinio miško, tai galėjo nulemti medžių skirtumai ir atsitiktiniame miške naudotų sprendimų medžių duomenų pasikartojimas – kuo yra daugiau vienodų duomenų tuo sunkiau priimti korektiškus sprendimus.

Suformavus naują atsitiktinį mišką yra gaunamas geresnis tikslumas už pavienį sprendimų medį:



