

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**INFORMATIKOS FAKULTETAS**

**Intelektikos pagrindai (P176B101)**  
***Laboratorinių darbų ataskaita***

Atliko:

IFF-1/4 gr. studentas

Mildaras Karvelis

2024 m. kovo 27 d.

Priėmė:

lekt. Audrius Nečiūnas

**KAUNAS 2023**

## TURINYS

1.	Ivadas.....	3
2.	Duomenų rinkinys.....	4
3.	Sprendimų medis.....	5
4.	Paklaidos ir tikslumai .....	6
5.	Išvados .....	7

## 1. Įvadas

Sprendimų medžio algoritmas kuriamas panaudojant Python programavimo kalbą. Darbo tikslo atlikimo eiga yra:

1. Pasirinkite duomenų rinkinį kuriam sudarysite sprendimų medį.
2. Kaip sprendimų medžio išvestį pasirinkite prognozuojamą atributą (Patariama pasirinkti kategorinį kintamąjį, kurio kardinalumas yra nuo 4 iki 10).
3. Turimą duomenų rinkinį suskaidykite į apmokymo bei testavimo poaibius. Apmokymo aibė turi būti didesnė nei testavimo.
4. Suskaidykite duomenų poaibius į įvestis ir išvestis.
5. Naudojant apmokymo duomenų rinkinį, sudarykite sprendimų medį. Galime rinktis iš keleto algoritmų ID3, C4.5, CART ir pan. Nuo to priklauso kokius indeksus/metodus naudosite medžio dalijimui (pvz., Gini, Gain ir t.t.). Žinoti koks yra skirtumas tarp šių algoritmų ir dalijimo indeksų.
6. Grafiškai atvaizduokite gautą sprendimų medį. Atvaizdavimui galima naudoti slearn ir graphviz arba kitas bibliotekas. Jei sudarytas sprendimų medis yra labai didelis, kad pavyktų įskaitomai pateikti ataskaitoje - įkelkite failą atskirai ir ataskaitoje pateikite tik medžio fragmentą ir komentarus apie gautą struktūrą.
7. Ištestuokite sudaryta sprendimų medį naudojant testavimo duomenis ir apskaičiuokite prognozavimo tikslumą/paklaidą. Nurodykite kokią paklaidos metriką skaičiuojate (pvz., MAE, MSE ir t.t). Taip pat klasifikavimo uždaviniui pateikite susimaišymo (angl. confusion) matricą.
8. Keičiant maksimalų medžio gylį, eksperimentiniu būdu išmatuokite skirtingų gylių (3-4 variacijos) medžių formavimo trukmę bei gaunamą tikslumą, t.y. medžio auginimas stabdomas nuo tam tikro gylio. Rezultatus pateikite ataskaitoje.
9. Naudojant tą patį apmokymo ir testavimo duomenų imties pasiskirstymą kaip ir formuojant sprendimų medį, suformuokite atsitiktinį mišką kurį sudaro 5 medžiai. Ataskaitoje pateiktų jų skirtumus. Maksimalus medžio gylis - gylis užfiksuotas eksperimento metu, kuris pateikė geriausius rezultatus.
10. Keičiant mišką sudarančių medžių kiekį [3-9], nustatykite geriausius rezultatus pateikiantį atsitiktinį mišką.
11. Palyginkite pirminio sprendimų medžio ir atsitiktinio miško gautus rezultatus.

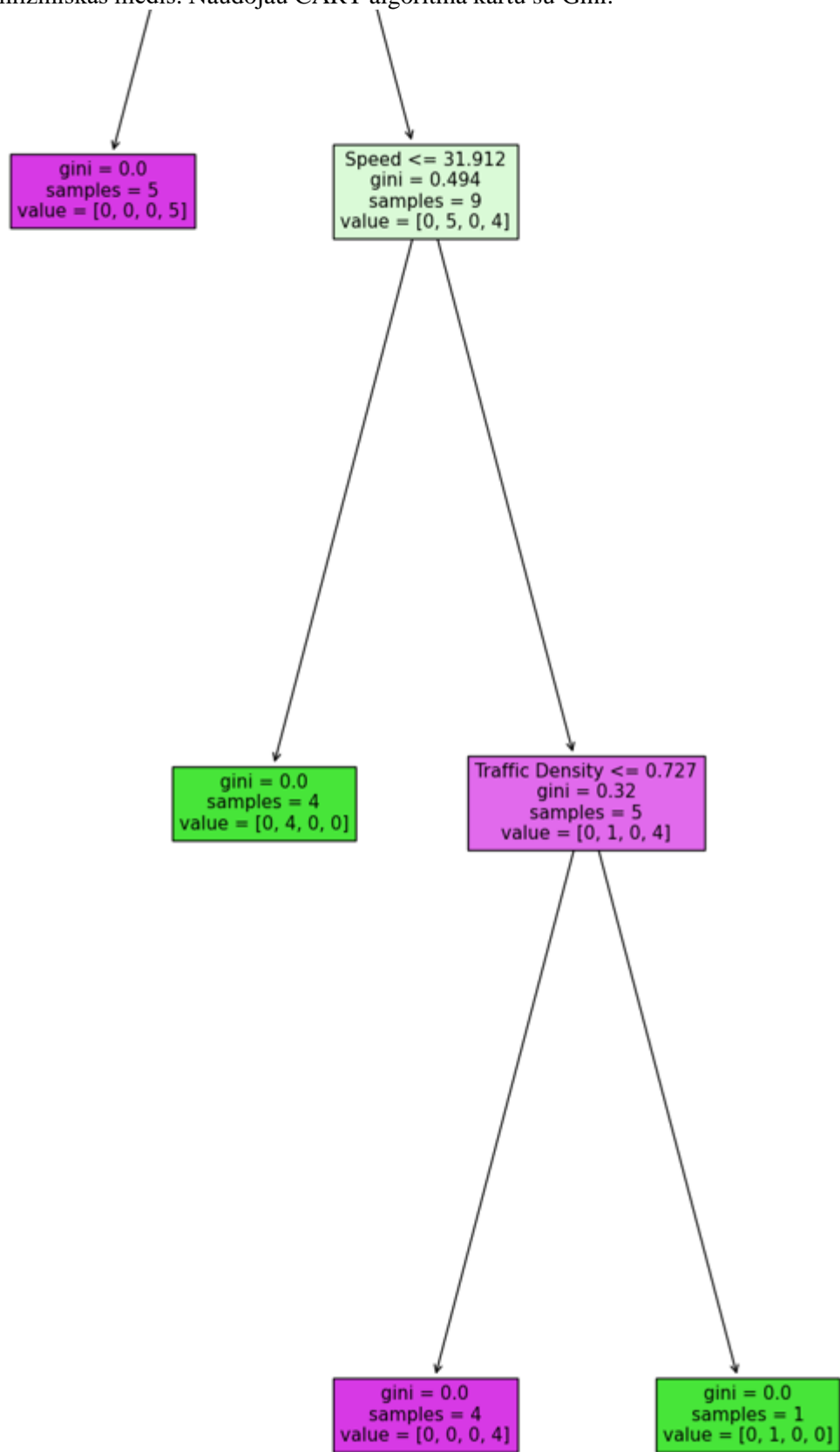
## 2. Duomenų rinkinys

Duomenų rinkinį sudaro:

- „City“ – miesto pavadinimas;
- „Vehicle Type“ – transporto priemonės tipas;
- „Weather“ – oro sąlygos;
- „Economic Condition“ – ekonominė padėtis;
- „Day of Week“ – savaitės diena;
- „Hour of Day“ – valanda (1-24 h.);
- „Speed“ – greitis, km/h;
- „Is Peak Hour “ – ar tai piko valanda? (True arba false, 0 arba 1);
- „Random Event Occured“ – ar kažkas įvyko? (True arba false, 0 arba 1);
- „Energy Consumption“ – energijos suvartojimas;
- „Traffic Density“ – eismo tankumas.

### 3. Sprendimų medis

Gavosi milžiniškas medis. Naudojau CART algoritma kartu su Gini.



1 pav. Medžio fragmentas

#### 4. Paklaidos ir tikslumai

Paskaičiavus su MAE - Mean Absolute Error, algoritmu gavome, jog medžio

- $MAE = 0.3466666666666667$

Paskaičiavus MAE su skirtingais medžio gyliais gavome:

- Gylis: 3, Mokymo laikas: 0.0036041736602783203, MAE : 0.7477486133611289
- Gylis: 4, Mokymo laikas: 0.0024623870849609375, MAE : 0.6454377331424548
- Gylis: 5, Mokymo laikas: 0.0027005672454833984, MAE : 0.5495349443735198
- Gylis: 6, Mokymo laikas: 0.0030205249786376953, MAE : 0.4911001697887927

Miškų tikslumai keičiant medžių kiekį miške:

- Medžių skaičius: 3, Tikslumas: 0.74
- Medžių skaičius: 4, Tikslumas: 0.76
- Medžių skaičius: 5, Tikslumas: 0.75
- Medžių skaičius: 6, Tikslumas: 0.75
- Medžių skaičius: 7, Tikslumas: 0.74
- Medžių skaičius: 8, Tikslumas: 0.73
- Medžių skaičius: 9, Tikslumas: 0.75

Geriausio miško sudėtis: 4 medžiai

Geriausias miško tikslumas: 0.76

## 5. Išvados

- Didelis kiekis duomenų gali eksponentiškai pakeisti algoritmo veikimo laiką;
- Didesnis medžių skaičius miške nevisada duoda geresnius tikslumo rezultatus;
- Didesnis medžio gylis duoda geresnius MAE rezultatus.