KAUNO TECHNOLOGIJOS UNIVERSITETAS

Intelektikos pagrindai (P176B101)

Laboratorinis darbas Nr. 2

Atliko:

IFAi-0 gr. studentas

Martynas Tvaska

2022 m. lapkričio 4 d.

Priėmė:

Arnas Nakrošis

Agnė Paulauskaitė-Tarasevičienė

KAUNAS 2022

TURINYS

1. Duomenų rinkinys 3

2. Sprendimų medis 4

2.1. Pasiruošimas darbui 4

2.2. Medis (1) 5

2.3. Medis (2) 6

2.4. Medis (3) 7

2.1. Medis (4) 8

3. Atsitiktinis miškas 9

3.1. Pasiruošimas darbui 9

3.2. Atsitiktinis miškas (1) 9

3.3. Atsitiktinis miškas (2) 10

3.4. Atsitiktinis miškas (3) 11

4. Palyginimas 12

5. Išvados 12

# Duomenų rinkinys

Laboratoriniai darbui pasirinktas duomenų rinkinys duoda 200 eilučių apie pacientus ir kokie vaistai jiems tiko.

Rinkinio atributai:

* Age - paciento amzius
* Sex - paciento lytis
* BP - spaudimas
* Cholesterol - cholesterolio lygis
* Na\_to\_K - kalis - natris
* Drug - Vaistas kuris padėjo

Duomenų pavyzdys:



Nuoroda į duomenų rinkinį:

* <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-ML0101EN-SkillsNetwork/labs/Module%203/data/drug200.csv>
* <https://www.kaggle.com/datasets/pablomgomez21/drugs-a-b-c-x-y-for-decision-trees>

# Sprendimų medis

## Pasiruošimas darbui

Darbui atlikti naudojama „sklearn“ biblioteka.

Sklearn bibliotekos funkcija „**DecisionTreeClassifier**“ medžio sudarymui naudoja CART algoritmą, medžio dalinimui numatytai GINI, bet galima rinktis tarp: GINI, ENTROPY, LOG\_LOSS.

* **ID3** yra godus algoritmas. Jis kiekvienam mazgui renkasi kategorini atributą kuris suteiks daugiausia informacijos. Medžiai auginami iki maksimalaus jų dydžio. Po to medis genimas, kad pritaikyti nematytiems duomenims.
* **C4.5** yra ID3 patobulinimas, suteikia galimybę naudoti ne tik kategorinius atributus.
* **C5.0** yra tikslesnis, naudoja mažiau atminties.
* **CART** panašus į C4.5, išvestyje palaiko skaitines reikšmes.

Medžio sudarymui naudosiu **CART** algoritmą, medžio dalinimui **GINI**.

Duomenų rinkinį dalinu į dvi dalis, **70% mokymui 30% testavimui**.

Pasirenku **"Drug"** kaip prognozuojamą atributą. Jo **kardinalumas 5**

<https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>

## Medis (1)

Medžio gylis: 4

Lapų skaičius: 6

Tikslumas: 100%

A picture containing timeline

Description automatically generated

Lyginami tikri duomenys su atspėtais:

Graphical user interface, application

Description automatically generated with medium confidence

Sumaišymo matrica:

A picture containing graphical user interface

Description automatically generated

## Medis (2)

Medžio gylis: 3

Lapų skaičius: 5

Tikslumas: 90%

Diagram

Description automatically generatedTrukmė: 0.2s

Lyginami tikri duomenys su atspėtais:

A picture containing application

Description automatically generated

Sumaišymo matrica:

A picture containing square

Description automatically generated

## Medis (3)

Medžio gylis: 2

Lapų skaičius: 3

Tikslumas: 85%

Trukmė: 0.2s

Diagram

Description automatically generated

Lyginami tikri duomenys su atspėtais:

A picture containing table

Description automatically generated

Sumaišymo matrica:

Graphical user interface, application

Description automatically generated

## Medis (4)

Medžio gylis: 1

Lapų skaičius: 2

Tikslumas: 70%

Trukmė: 0.1s

Diagram

Description automatically generated

Lyginami tikri duomenys su atspėtais:

A picture containing text

Description automatically generated

Sumaišymo matrica:

A picture containing chart

Description automatically generated

# Atsitiktinis miškas

## Pasiruošimas darbui

Sklearn bibliotekos funkcijos „**RandomForestClassifier**“ keletas pagrindinių kintamųjų:

* **n\_estimators** – medžių skaičius
* **criterion** – algoritmas medžio dalinimo kokybės vertinimui (GINI, ENTROPY, LOG\_LOSS)
* **max\_depth** – maksimalus medžio gylis

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>

## Atsitiktinis miškas (1)

Medžių skaičius: 5

Tikslumas: 98%

Trukmė: 2.1s

Diagram

Description automatically generated

Lyginami tikri duomenys su atspėtais:

A picture containing application

Description automatically generated

Sumaišymo matrica:

A picture containing square

Description automatically generated

## Atsitiktinis miškas (2)

Medžių skaičius: 4

Tikslumas: 100%

Trukmė: 2.1s

Diagram

Description automatically generated

Lyginami tikri duomenys su atspėtais:

A picture containing application

Description automatically generated

## Atsitiktinis miškas (3)

Medžių skaičius: 3

Tikslumas: 100%

Trukmė: 1.7s

Diagram

Description automatically generated

Lyginami tikri duomenys su atspėtais:

A picture containing application

Description automatically generated

# Palyginimas

Geriausius rezultatus pateikė medis (1).

Tikslumas 100%. Trukmė: 0.2s

Atsitiktinis miškas (3), geriausias iš miškų:

Tikslumas 100%. Trukmė 1.7s

Mano atveju norint gauti 100% tikslumą, užtenka sudaryti medį kurio gylis 4, naudojant CART algoritmą.

# Išvados

Pagal gautuosius rezultatus matome, kad šitas duomenų rinkinys yra „paruoštas“ klasifikavimo uždaviniams.

Dažnu atveju gaunamas labai didelis tikslumas, ko realiame gyvenime negali būti.