Projekto ataskaitos turinys

# Bendri nurodymai

Ataskaita turi būti parengta tik vienoje kalboje – lietuvių arba anglų. Dokumentas turi turėti puslapių numerius ir tinkamai[[1]](#footnote-1) sumaketuotas.

Ataskaitoje turi būti:

1. Titulinis lapas pagal Universiteto reikalavimus
2. Turinys
3. Santrauka

Joje nurodyti duomenų rinkinio pavadinimą, adresą; trumpai aprašyti problemą, išvardinti panaudotus mašininio mokymo metodus ir jų gautus tikslumus.

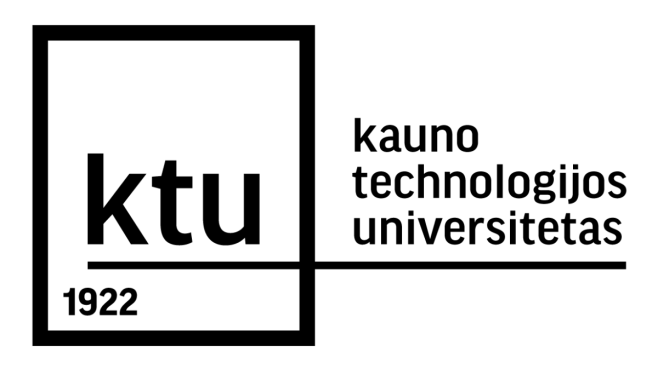
1. Atlikėjų sąrašas lentelės pavydale atskiruose stulpeliuose nurodant
   1. vardą, pavaradę ir užsiėmimo dieną ir laiką
   2. atsakomybes projekte ir
   3. parengtus skyrius ataskaitoje

Ataskaitos skyrių pavadinimai turi atitikti reikalavimų bloku formuluotes bei nurodant juose realizacijos sudėtingumą, pvz.:

* **Duomenų surinkimas** (iš sukurto programinio produkto) , arba
* **Duomenų surinkimas** (iš atvirų šaltininių panaudojant programinį kodą), arba
* **Duomenų surinkimas** (iš tekstinio dokumento)

Reikalavimai kiekvieno skyriaus turiniui pateikiamas žemiau. Skyrius pateikamas ataskaitoje tik jeigu komanda realizavo atitinkamą reikalavimų bloką. Kiekviename skyriuje be eksperimentų rezultatų turi būti pateikta realizacijos detalizacija (kodo fragmentas ir/arba naudojamos bibliotekos nurodymas ir išaiškinimas) kuri atitinka pasirinktą pasirinktą sudėtingumą.

Ataskaitos gale pateikti naudotos literatūros bei šaltitnių sąrašą.



**KAUNO TECHNOLOGIJOS UNIVERSITETAS INFORMATIKOS FAKULTETAS**

[**P176B101 Intelektikos pagrindai 2019**](https://moodle.ktu.edu/course/view.php?id=2081)

**technologijos projekto ataskaita**

Data:

2019-05-19

Studentai: Tomas Jasulaitis IFF-6/4

Lukas Biliutavičius IFF-6/4

Tadas Vasiliauskas IFF-6/4

Contents

[Bendri nurodymai 1](#_Toc9367448)

[1. Santrauka 3](#_Toc9367449)

[2. **Duomenų surinkimas** (iš atvirų šaltininių panaudojant programinį kodą) 4](#_Toc9367450)

[3. **Duomenų paruošimas ir valymas** (rankiniu būdu ir programiškai) 6](#_Toc9367451)

[4. **Dimensijų sumažinimas** (<pasirinktas sudėtingumas>) 6](#_Toc9367452)

[5. **Įžanginiai eksperimentai, patikrinantys ar dimensijų sumažinimą verta naudoti** (<pasirinktas sudėtingumas>) 6](#_Toc9367453)

[**6.** ***1*-ojo mašininio mokymosi metodo su mokytoju panaudojimas (programinė realizacija iškviečiant atitinkamą biblioteką)** 7](#_Toc9367454)

[**7.** ***2*-tojo mašininio mokymosi metodo su mokytoju panaudojimas (programinė realizacija iškviečiant atitinkamą biblioteką)** 8](#_Toc9367455)

[**8.** **Mašininio mokymosi metodų su mokytoju rezultato parinkimas balsavimo principu (pilnai suprogramuota)** 10](#_Toc9367456)

[**9.** **Balsavimo principu gautų rezultatų pritaikymas(pilnai suprogramuota)** 10](#_Toc9367457)

[10. **Mašininio mokymosi be mokytojo metodo panaudojimas** (<pasirinktas sudėtingumas>) 10](#_Toc9367458)

[11. **Mokymosi be mokytojo metodo gautų rezultatų pritaikymas (programoje) / pakomentavimas ataskaitoje** (<pasirinktas sudėtingumas>) 10](#_Toc9367459)

[**12.** **Kryžminė patikra** 10](#_Toc9367460)

[**13.** **Literatūra** 10](#_Toc9367461)

# Santrauka

Duomenų rinkinio pavadinimas – pokerio rankos duomenų rinkinys (Poker Hand Data Set)

<https://archive.ics.uci.edu/ml/datasets/Poker%2BHand?fbclid=IwAR2QtK-sO6WSzK2J4ZpL0g8B9SJIU3uPtoYn6kNs2_oFx4o8uQ6Pn_-eztw>

Problema – nustatyti pokerio kortų kombinaciją iš duotų penkių kortų.

Projekte buvo panaudoti RandomTree, DecisionTree mašininio mokymo metodai, jų tikslumai: 0.956 ir 0.905 atitinkamai.

|  |  |  |  |
| --- | --- | --- | --- |
| **Atlikėjas** | **Užsiemimo diena ir laikas** | **Atsakomybė projekte** | **Parengti ataskaitos skyriai** |
| Tomas Jasulaitis | Antradienis 13:30 | 1. Mašininio mokymo metodas su mokytoju (Decision Tree) | Santrauka, duomenų surinkimas, mašininio mokymo su mokytoju metodas. |
| Lukas Biliutavičius | Trečiadienis 11:00 | 1. Mašininio mokymo metodas su mokytoju (Random Forest)  2. Duomenų surinkimas  3. Duomenų paruošimas | 1. Duomenų paruošimas ir valymas 2. 2-tojo mašininio mokymosi metodo su mokytoju panaudojimas 3. [Balsavimo principu gautų rezultatų pritaikymas](#_Toc9367457) 4. Balsavimo principu gautų rezultatų pritaikymas |
| Tadas Vasiliauskas | ?? | 3. Mašininio mokymo metodas su mokytoju |  |

# **Duomenų surinkimas** (iš atvirų šaltininių panaudojant programinį kodą)

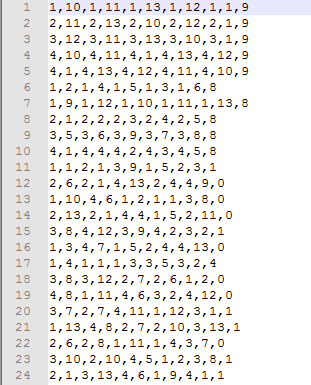
Duomenų rinkinys sudarytas is 11 skaitinių reikšmių. Pirmos 10 reikšmių nusako kortų naudojamų pokerio žaidime rūšį ir skaitinę reikšmę. 11 reikšmė nusako klasę, kuri priskiriama rankos stiprumui (0-9).

Duomenų rinkinyje yra 1025010 įrašų, jo adresas : <https://archive.ics.uci.edu/ml/datasets/Poker%2BHand?fbclid=IwAR2QtK-sO6WSzK2J4ZpL0g8B9SJIU3uPtoYn6kNs2_oFx4o8uQ6Pn_-eztw>

**Atributų informacija:**

1) S1 "Suit of card #1"   
Ordinal (1-4) representing {Hearts, Spades, Diamonds, Clubs}   
  
2) C1 "Rank of card #1"   
Numerical (1-13) representing (Ace, 2, 3, ... , Queen, King)   
  
3) S2 "Suit of card #2"   
Ordinal (1-4) representing {Hearts, Spades, Diamonds, Clubs}   
  
4) C2 "Rank of card #2"   
Numerical (1-13) representing (Ace, 2, 3, ... , Queen, King)   
  
5) S3 "Suit of card #3"   
Ordinal (1-4) representing {Hearts, Spades, Diamonds, Clubs}   
  
6) C3 "Rank of card #3"   
Numerical (1-13) representing (Ace, 2, 3, ... , Queen, King)   
  
7) S4 "Suit of card #4"   
Ordinal (1-4) representing {Hearts, Spades, Diamonds, Clubs}   
  
8) C4 "Rank of card #4"   
Numerical (1-13) representing (Ace, 2, 3, ... , Queen, King)   
  
9) S5 "Suit of card #5"   
Ordinal (1-4) representing {Hearts, Spades, Diamonds, Clubs}   
  
10) C5 "Rank of card 5"   
Numerical (1-13) representing (Ace, 2, 3, ... , Queen, King)   
  
11) CLASS "Poker Hand"   
Ordinal (0-9)   
  
0: Nothing in hand; not a recognized poker hand   
1: One pair; one pair of equal ranks within five cards   
2: Two pairs; two pairs of equal ranks within five cards   
3: Three of a kind; three equal ranks within five cards   
4: Straight; five cards, sequentially ranked with no gaps   
5: Flush; five cards with the same suit   
6: Full house; pair + different rank three of a kind   
7: Four of a kind; four equal ranks within five cards   
8: Straight flush; straight + flush   
9: Royal flush; {Ace, King, Queen, Jack, Ten} + flush

Pavyzdinis duomenų įrašo pavyzdys.



# **Duomenų paruošimas ir valymas** (rankiniu būdu ir programiškai)

Duomenų formatas pateiktas duomenų failuose:

**1,10,1,11,1,13,1,12,1,1,9**

Čia pirmi 10 skaičių – pokerio kortų aprašas (angl. suit, rank). Vienuoliktas skaičius – klasė(stipriausia kortų kombinacija).

Paruošiant duomenis, iš CSV formato reikia perdaryti į dvimačius masyvus.

Pirmasis masyvas [Nx10], kur N – eilučių kiekis duomenų rinkinyję. 10 stulpelių atitinkamai skirta kortos identifikavimui.

Antrasis masyvas [Nx1], paskutinis duomenų eilutės skaitmuo, reiškiantis turimų penkių kortų kombinaciją.

# **Dimensijų sumažinimas** (<pasirinktas sudėtingumas>)

Pateikti dimensijų mažinimo metodo pavadinimą, reikalingas formules ir trumpa dimensijų sumažinimo proceso santrauką bei tarpinių skaičiavimų rezultatus (pvz. kovariacijų matricos fragmentą bei nurodant slenkstį).

Pateikti atributus (dimensijas) iki sumažinimo ir po jos.

# **Įžanginiai eksperimentai, patikrinantys ar dimensijų sumažinimą verta naudoti** (<pasirinktas sudėtingumas>)

Padaryti eksperimentus (kryžminės patikros naudoti nebūtina) su pasirinktais mokymosi metodais bei duomenų rinkiniu su pradinių dimensijų skaičiumi. Eksperimentams atlikti gali būti naudojama paprasčiaiusia platforma (pvz. Matlab). Taip atlikti identiškus eksperimentus su duomenų rinkiniu kuriam sumažinote dimensijų skaičių.

Suvestinėje Lentelėje pateikti

|  |  |  |  |
| --- | --- | --- | --- |
|  | Maš.Mok.Metodas\_1 | ... | Maš.Mok.Metodas\_M |
| Pradinis duomenų rinkinys | <tikslumas> |  | <tikslumas> |
| Sumažintas duomenų rinkinys  (<atr\_1>, <atr\_N> -- atributai, kurių atsisakėte) | <tikslumas> |  | <tikslumas> |

Išvada dėl dimensijų sumažinimo pritaikymo tikslingumo remiantis pateikta lentele.

# ***1*-ojo mašininio mokymosi metodo su mokytoju panaudojimas (programinė realizacija iškviečiant atitinkamą biblioteką)**

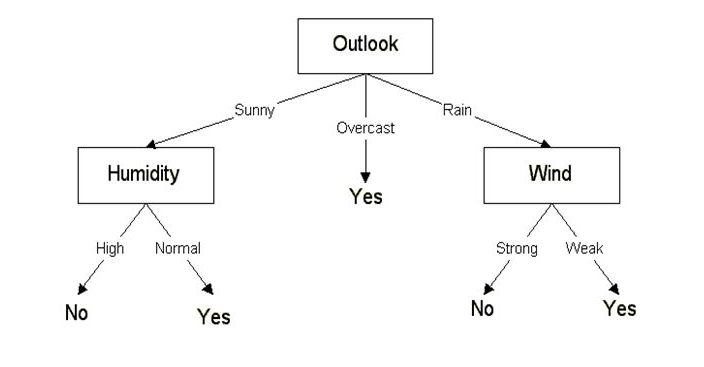
Metodo pavadinimas - Decision Tree

Decision Tree metodas priskiriamas prie mašininio mokymosi su mokytojų algoritmų šeimos. Šiuo algoritmu galima spręsti klasifikacijos ir regresijos problemas. Decision Tree sukuriamas treniravimo modelis, kuris gali būti naudojamas nuspėti klasę ar reikšmę pasirinktų kintamųjų naudojantis mokomojo nusprendimo taisykles (learning decision rules), kurias numanė iš treniravimo duomenų.

Decision Tree metodas yra nesunkiai perprantamas, duomenų paruošimo etapas yra ne toks svarbus nei kitiems modeliavimo metodams.

Šaltinis: <http://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/> <https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb>

Iliustracinis pavyzdys:



Nuoroda: <https://www.xoriant.com/blog/product-engineering/decision-trees-machine-learning-algorithm.html>

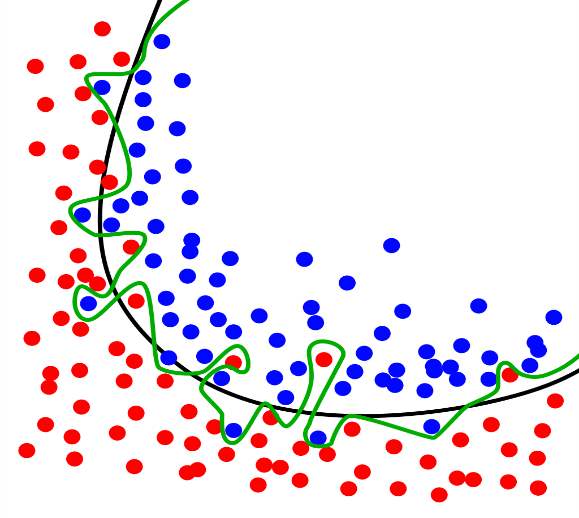
|  |  |
| --- | --- |
| Eksperimento numeris | Metodo tikslumo įvertis |
| 1 | 0,90405 |
| 2 | 0,90368 |
| 3 | 0,9058 |
| 4 | 0,90578 |
| 5 | 0,90579 |
| 6 | 0,90614 |
| 7 | 0,90393 |
| 8 | 0,90481 |
| 9 | 0,90639 |
| 10 | 0,90406 |

Kryžminės patikros eksperimentų metodo tikslumo įverčių reikšmės pasiskirsčiusios labai arti viena kitos, tačiau nėra labai aukštos, tik 90.6% tikslumas.

# ***2*-tojo mašininio mokymosi metodo su mokytoju panaudojimas (programinė realizacija iškviečiant atitinkamą biblioteką)**

Metodo pavadinimas - Random Forest

„Random Forest“ taip pat žinomas kaip „decision forest“ yra populiarus metodas naudojamas kurti modelius klasifikavimui ir regresijai. Šis metodas sukuria daug „decision tree“ kurių pagalba gaunamas geriausias atsakymas. „Random forest“ metodas išsprendžia „decision tree“ metodo persimokymo problemą.



Ši problema atsiranda kai mokymo duomenys yra labai arti vienas kito ir metodas klasifikuodamas gauna didesnę paklaidą. Šioje nuotraukoje juoda kreivė vaizduoja „random forest“ klasifikavima, o žalia – „decision tree“.



Sukūrus mišką, pagal paduodamus duomenis kiekvienas medis balsuoja už atsakymą. Didžiausiai balsų surinkęs atsakymas yra parenkamas kaip klasifikavimo klasė.

Šaltinis: <https://www.datascience.com/resources/notebooks/random-forest-intro>,

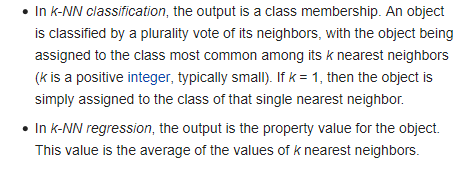
<https://en.wikipedia.org/wiki/Random_forest>

|  |  |
| --- | --- |
| Eksperimento numeris | Metodo tikslumo įvertis |
| 1 | 0,95422 |
| 2 | 0,95584 |
| 3 | 0,95683 |
| 4 | 0,95557 |
| 5 | 0,95694 |
| 6 | 0,95517 |
| 7 | 0,95561 |
| 8 | 0,95503 |
| 9 | 0,95565 |
| 10 | 0,95468 |

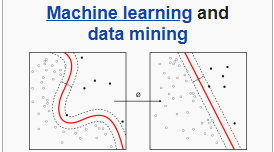
Kryžminės patikros eksperimentų metodo tikslumo įverčių reikšmės pasiskirsčiusios labai arti viena kitos. Didžiausias klasifikavimo tikslumas – 5 iteracija, **95,7%.**

# ***2*-tojo mašininio mokymosi metodo su mokytoju panaudojimas (programinė realizacija iškviečiant atitinkamą biblioteką)**

K-artimiausių kaimynų metodas yra ne parametrinis metodas, kuris naudojamas klasifikavimui ir regresijai. Tiek tam tiek tam metodui, įvedimo duomenys yra k artimiausi apmokymo pavyzdžiai. Išvados priklauso nuo to ar buvo naudojama regresijai ar klasifikavimui.



Šis metodas yra tingus metodas ir yra vienas iš paprasčiausių mašinų mokymo algortimų. Tiek klasifikacijai tiek regresijai galima priskirti svorinius koficientus, tai artimesni kaimynai prisidės prie vidurkio, o ne tolimesni.



Šaltinis: <http://accord-framework.net/docs/html/T_Accord_MachineLearning_KNearestNeighbors_1.htm>.

# **Mašininio mokymosi metodų su mokytoju rezultato parinkimas balsavimo principu (pilnai suprogramuota)**

|  |  |
| --- | --- |
| Metodas | Tikslumas |
| Random Forest | 0.957 |
| Decision Tree | 0.906 |
| Tado | 0.0 |

Visiems metodams atlikus kryžminę patikrą ir gražinus vidutinį tikslumą, programinis kodas parenką metodą su aukščiausiu klasifikavimo rodikliu.

# **Balsavimo principu gautų rezultatų pritaikymas(pilnai suprogramuota)**

Programiam kodui parinkus geriausią mašininio mokymosi metodo su mokytoju rezultatą, laukiama įvesties rinkinio.

Rinkinį sudaro 10 skaičių, pvz: 1 1 2 2 3 3 4 4 4 5

Pirmas skaičius: 1 – Širdys, 2 – Vynai, 3 – Būgnai, 4 – Giliai.

Antras skaičius: 1 – Tūzas, 2 – 2, 3 – 3, 4 – 4, 5 – 5, 6 – 6, 7 – 7, 8 – 8, 9 – 9, 10 – 10, 11 – Valetai, 12 – Dama, 13 – Karalius.

Atitinkamai pirmas ir antras skaičiai vaizduoja pirmąją kortą, trečias ir ketvirtas – antrąją kortą ir t.t.

Įvedus šiuos duomenis, geriausias iš trijų klasifikavimo metodų parodys kokią aukščiausia pokerio kombinacija yra galima iš duotų kortų.

# **Mašininio mokymosi be mokytojo metodo panaudojimas** (<pasirinktas sudėtingumas>)

Pateikti apmokymo rezultatus bei juos pakomentuoti.

# **Mokymosi be mokytojo metodo gautų rezultatų pritaikymas (programoje) / pakomentavimas ataskaitoje** (<pasirinktas sudėtingumas>)

Aprašyti pritaikymo būdą (pvz. programoje įvedame grybų atributų rinkinį ir programa pateikia rekomendaciją - valgyti grybą arba ne).

# **Kryžminė patikra**

Atskirai šis skyrius nepateikiamas, nes yra sudėtinė dalis skyriaus ***i*-tojo mašininio mokymosi metodo su mokytoju panaudojimas**

# **Literatūra**

Privaloma nurodyti naudotus resursus, įskaitant mokymo šaltinius (*tutorials*), kurie buvo naudoti projektui atlikti.

<http://crsouza.com/2012/01/04/decision-trees-in-c/>

<http://accord-framework.net/docs/html/T_Accord_MachineLearning_DecisionTrees_DecisionTree.htm>

<https://github.com/mdabros/SharpLearning>

<https://csharp.hotexamples.com/examples/Accord.MachineLearning.DecisionTrees.Learning/C45Learning/-/php-c45learning-class-examples.html>

<https://towardsdatascience.com/supervised-machine-learning-classification-5e685fe18a6d>

<http://accord-framework.net/docs/html/T_Accord_MachineLearning_KNearestNeighbors_1.htm>

1. <http://ifko.ktu.lt/~agnius/t120b111/Praktikos_dokumentai/Informatikos%20inzinerijos%20reikalavimai%20bakalauro%20baigiamajam%20darbui%20v0.1.doc> Priedai 3 ir 5 [↑](#footnote-ref-1)