

Rīgas Tehniskā Universitāte

Datorzinātnes un informācijas tehnoloģijas fakultāte

2.Praktiskais darbs

“Mākslīgā intelekta pamati”

Vārds Uzvārds: Andrejs Gavriļuks

Stud.apl.nr.: 211RDB145

7. grupa

2022/2023 m.g.

## **Saturs**

[Saturs 2](#_Toc134755833)

[I daļa - Datu pirmapstrāde/izpēte 3](#_Toc134755834)

[Datu apraksts 4](#_Toc134755835)

[Formāta salīdzinājums 8](#_Toc134755836)

[Datu apstradāšana 9](#_Toc134755837)

[Secinājums I daļai 14](#_Toc134755838)

[II daļa – Nepārraudzītā mašīnmācīšanās 15](#_Toc134755839)

[Secinājums II daļai 18](#_Toc134755840)

[III daļa – Pārraudzītā mašīnmācīšanās 19](#_Toc134755841)

[Algoritmu apraksts 19](#_Toc134755842)

[Testēšana 22](#_Toc134755843)

[Secinājums III daļai 26](#_Toc134755844)

[Kopējais secinājums 27](#_Toc134755845)

# I daļa - Datu pirmapstrāde/izpēte

**Nosaukums**: Wine Quality Dataset

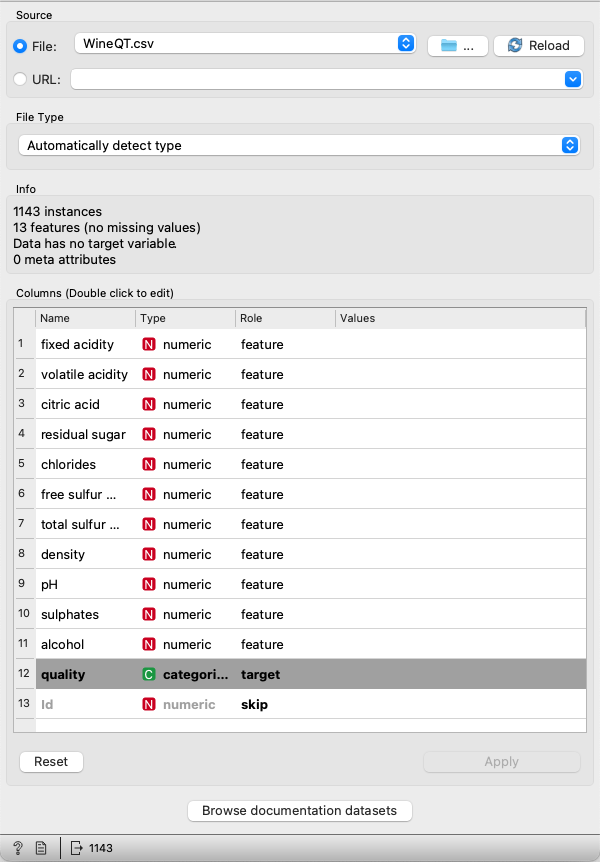
**Autors**:M YASSER H

Vīna un jo īpaši vīnogu vīna ķīmiskais sastāvs ir ārkārtīgi sarežģīts, jo tajā ir vairāk nekā 600 organisko un neorganisko vielu.

Vidēji sarkanvīns satur 86 % ūdens, 12 % etilspirta un aptuveni 1 % glicerīna. No organiskajām skābēm vīnā dominē pienskābe, citronskābe, ābolskābe, etiķskābe un dzintarskābe - aptuveni 0,4 %. Rūgtvielas un fenola savienojumi veido tikai 0,1 %, bet tieši šos savienojumus mēs aplūkosim, aplūkojot vīna krāsu un garšu.

Šajā datubāzē ir iekļauta informācija par kritērijiem, kas sastopami dažādos vīnos. Tie var ietekmēt vīna garšu un stilu. Piemēram, zems pH var piešķirt vīnam spilgtāku skābumu, sulfāti var uzlabot vīna stabilitāti, bet spirta līmenis var ietekmēt struktūru un alkohola rūgtumu.Ir jāizvēlas un jāapraksta datu kopa, pamatojoties uz informāciju, kas sniegta krātuvē, kurā datu kopa ir pieejama.

Šim praktiskajam darbam tiek izmantota datu bāze, kurā analizēta vīna kvalitāte. Šis datu kopums ir saistīts ar Portugāles vīna "Vinho Verde" sarkanajām šķirnēm. Datu kopa apraksta dažādu vīnā esošo ķīmisko vielu daudzumu un to ietekmi uz vīna kvalitāti. Šajā datubāzē ir 13 kritēriji un 1143 ieraksti.



1.att. Kategorijas

Attēlā 1.att. ir redzamas visas kategorijas, kurās iedalīta šī datubāze. Visi objekti ir iedalīti kategorijās pēc to nosaukuma, un katram objektam ir savas dažādas vērtības. Visas vērtības ir "numeric" formātā. Dažiem objektiem trūka vērtību, un tās bija jāaizstāj ar "0", lai galīgais rezultāts būtu precīzāks.

## **Datu apraksts**

Изображение выглядит как снимок экрана, График, текст, диаграмма

Автоматически созданное описание

2.att. Fixed Acidity

“Fixed acidity” - fiksēts skābums. Šis parametrs raksturo skābuma līmeni vīnā. To mēra citronskābes gramos uz litru vīna. Datubāzē fiksētā skābuma diapazons var būt no 4,6 gramiem līdz 15,9 gramiem litrā. Noteiktā skābuma līmenis vīnā var būt atkarīgs no šādiem ķīmiskajiem parametriem:

* kopējais skābums
* organisko skābju saturs
* vīna vecums
* Skābuma līmenis Ph

Изображение выглядит как текст, снимок экрана, График, диаграмма

Автоматически созданное описание

3.att. Volatile Acidity

“Volatile acidity” - gaistošais skābums. Šis parametrs raksturo skābuma līmeni, ko veido vīnā esošās gaistošās skābes. Šo parametru mēra gramos uz litru. Datu bāzē gaistošo skābju diapazons var būt no 0,12 gramiem līdz 1,58 gramiem litrā. Gaistošo skābju līmenis vīnā var būt atkarīgs no šādiem ķīmiskajiem parametriem:

* etiķskābes saturs
* mikrobioloģiskais faktors
* oksidēšanās process

Изображение выглядит как текст, График, снимок экрана, диаграмма

Автоматически созданное описание

4.att. Citric Acid

“Citric acid” - citronskābe. Šis parametrs raksturo citronskābes saturu vīnā. Šo parametru mēra gramos litrā. Datubāzē citronskābes diapazons var būt no 0 līdz 1 gramam.

Изображение выглядит как снимок экрана, текст, График, линия

Автоматически созданное описание

5.att. Residual Sugar

“Residual sugar” - atlikušo cukuru. Šis parametrs raksturo atlikušā cukura daudzumu, kas paliek vīnā pēc fermentācijas procesa pabeigšanas. Šo parametru mēra gramos litrā. Datubāzē atlikumcukura diapazons var būt no 0,9 līdz 15,5 gramiem. Atlikušā cukura līmenis vīnā var būt atkarīgs no šādiem ķīmiskajiem parametriem:

* fermentācijas ilgums
* stabilizācija
* skābums
* mijiedarbība ar citām sastāvdaļām

Изображение выглядит как текст, снимок экрана, График, линия

Автоматически созданное описание

6.att. Chloride

“Chlorides” – hlorīdi. Šis parametrs raksturo hlorīdu sāls saturu vīnā. Šo parametru parasti mēra gramos litrā. Datubāzē hlorīdu diapazons var būt no 0,01 līdz 0,61 gramam.

Изображение выглядит как текст, снимок экрана, График, линия

Автоматически созданное описание

7.att. Free Sulfur Dioxide

“Free sulfur dioxide” - brīvo sēra dioksīdu. Šis parametrs raksturo brīvā sēra dioksīda daudzumu vīnā. Šo parametru mēra miljondaļās (ppm). Datu bāzē brīvā sēra dioksīda diapazons var būt no 1 līdz 68 ppm. Brīvā sēra dioksīda daudzums vīnā var būt atkarīgs no šādiem ķīmiskajiem parametriem:

* Kopējais sēra dioksīda saturs
* pH līmenis
* reducējošo vielu līmenis

Изображение выглядит как текст, снимок экрана, График, Шрифт

Автоматически созданное описание

8.att. Total Sulfur Dioxide

“Total sulfur dioxide” - kopējais sēra dioksīds. Šis parametrs raksturo kopējo brīvā un saistītā sēra dioksīda saturu vīnā. Šo parametru parasti mēra miljondaļās (ppm). Datubāzē kopējā sēra dioksīda diapazons var būt no 6 ppm līdz 289 ppm. Kopējais sēra dioksīda līmenis vīnā var būt atkarīgs no šādiem ķīmiskajiem parametriem:

* brīvā sēra dioksīda saturs
* pH līmenis

Изображение выглядит как текст, диаграмма, График, снимок экрана

Автоматически созданное описание

9.att. Density

“Density” – blīvums. Šis parametrs raksturo vīna blīvumu, ko mēra g/cm³ vai kg/l. Vīna blīvums ir fizikāla īpašība, kas izsaka vīna masu tilpuma vienībā. Datu bāzē blīvuma diapazons var būt no 0,99 līdz 1 g/cm³.

Изображение выглядит как График, диаграмма, текст, линия

Автоматически созданное описание

10.att. pH

“pH”. Šis parametrs raksturo vīna skābuma vai sārmainības līmeni. Tā ir bezizmēra vērtība, ko mēra skalā no 0 līdz 14, kur 7 atbilst neitrālam pH. Datu bāzē pH diapazons var būt no 2,74 līdz 4,01.

Изображение выглядит как снимок экрана, текст, График, линия

Автоматически созданное описание

11.att. Sulphates

“Sulphates” – sulfāti. Šis parametrs raksturo sulfātu saturu vīnā. Sulfāti ir sērskābes (SO4²-) sāļi, un tie var atrasties vīnā kā dabiskas sastāvdaļas vai pievienoti kā pārtikas piedevas. Datu bāzē sulfātu diapazons var būt no 0,33 līdz 2. Sulfātu līmenis vīnā var būt atkarīgs no šādiem ķīmiskajiem parametriem:

* pH
* skābekļa daudzums

Изображение выглядит как График, текст, диаграмма, линия

Автоматически созданное описание

12.att. Alcohol

“Alcohol” – alkohols. Šis parametrs raksturo alkohola saturu vīnā. To mēra tilpuma procentos, un tas norāda etanola daudzumu vīnā. Datu bāzē alkohola saturs var būt no 8,4 % līdz 14,9 %.

Изображение выглядит как снимок экрана, текст, График, линия

Автоматически созданное описание

13.att. Quality

“Quality” – kvalitāte. Šis parametrs raksturo vīna kvalitātes vērtējumu, pamatojoties uz sensoriem datiem vai pieredzējušu degustatoru vai ekspertu piešķirtajiem punktiem. Tā ir skaitliska vērtība, parasti no 0 līdz 10, kur augstāka vērtība norāda uz augstākas kvalitātes vīnu. Datu bāzē vīna kvalitātes diapazons var būt no 3 līdz 8.

## **Formāta salīdzinājums**

Изображение выглядит как текст, электроника, снимок экрана, дисплей

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, Прямоугольник, дисплей

Автоматически созданное описание

14.att. Datu bāze CSV formata 15.att. Datu bāze Excel formata

Izmantojot rīku "impute" 16.att., iestatīju parametru tukšo šūnu aizpildīšanai datu bāzē uz "0", kas palīdzēja aizpildīt tukšās šūnas datu bāzē un radīt precīzākus galīgos rezultātus.

## **Datu apstradāšana**

Изображение выглядит как снимок экрана, текст

Автоматически созданное описание

16.att. Datu apstradāšana



17.att. Scatter Plot Alcohol and Sulphates

Šajā grafikā 17.att. tika izmantota funkcija "Find Informative Projections", kur X ass ir "Alcohol" un Y ass ir parametrs "Sulphates", šie atributi dod vislielāko atdalījumu un dod vislabāku vertējumu. Lai padarītu datus redzamākus, man vajadzēja ieslēgt opciju "Jitter numeric values", lai labāk viņus atdalīt.

Изображение выглядит как снимок экрана, программное обеспечение, Мультимедийное программное обеспечение, Графическое программное обеспечение

Автоматически созданное описание

18.att. Scatter Plot Alcohol and Volatile acidity

Šajā grafikā 18.att. tika izmantota funkcija "Find Informative Projections", kur X ass ir "Alcohol" un Y ass ir parametrs "Volatile acidity", šie atributi ir 2 vietā , kuri dod vislielāko atdalījumu un dod vislabāku vertējumu. Lai dati būtu labāk redzami, man vajadzēja ieslēgt opciju "Jitter numeric values", lai labāk viņus atdalīt.

Изображение выглядит как программное обеспечение, Мультимедийное программное обеспечение, Графическое программное обеспечение, Значок на компьютере

Автоматически созданное описание

19.att. Distribution Total Sulfur Dioxide

Šī histogramma 19.att. parāda, ka ir izvēlēts "Total sulfur dioxide", kas ir atkarīgs no vienas klases "Quality". Uz Y ass ir attēlota vērtības biežuma iestāšanās varbūtība. Šajā histogrammā ir sadalījums, un katram objektam ir atšķirīgs sadalījums šajā gadījumā viens objekts ir vadošais un ir 483, un tas svārstās no 0 līdz 500, kas ietekmē "Quality" klasi.

" Изображение выглядит как снимок экрана, программное обеспечение, Мультимедийное программное обеспечение, Графическое программное обеспечение

Автоматически созданное описание

20.att. Distribution Citric Acid

Šī histogramma 20.att. parāda, ka ir izvēlēts "Citric Acid", kas ir atkarīgs no vienas klases "Quality". Uz Y ass ir attēlota vērtības biežuma iestāšanās varbūtība. Šajā histogrammā ir sadalījums, un katram objektam ir atšķirīgs sadalījums šajā gadījumā viens objekts ir vadošais un ir 268, un tas svārstās no 0 līdz 300, kas ietekmē "Quality" klasi. Kā arī šeit ir redzams ka ir tikai 3 klases ir vadoši, bet viņus kopā ir 6.

Изображение выглядит как снимок экрана, текст, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

21.att. Bar Plot Residual Sugar

Šajā histogrammā 21.att. ir parādīti sadaļas "Residual sugar" parametri, "Quality" klase tiek izmantota, lai to sagrupētu, jo tas ir vienīgais parametrs, attiecībā uz kuru var iegūt rezultātus. Šajā piemērā redzams, ka augstākas kvalitātes vīnā “Residual sugar” saturs ir daudz lielāks - 7,20 g/l.

Изображение выглядит как снимок экрана, текст, программное обеспечение, Мультимедийное программное обеспечение

Автоматически созданное описание

22.att. Bar Plot Total Sulfur Dioxide

Šajā histogrammā 22.att. ir parādīti sadaļas "Total sulfur dioxide" parametri, "Quality" klase tiek izmantota, lai to sagrupētu, jo tas ir vienīgais parametrs, attiecībā uz kuru var iegūt rezultātus. Šajā piemērā redzams, ka augstākas kvalitātes vīnā “Total sulfur dioxide” saturs ir daudz lielāks - 143 ppm.

Изображение выглядит как снимок экрана, программное обеспечение, текст, Значок на компьютере

Автоматически созданное описание

23.att. Feature Statistics Median

Изображение выглядит как программное обеспечение, График, текст, линия

Автоматически созданное описание

24.att. Feature Statistics Dispersion

# Secinājums I daļai

Apkopojot praktiskā darba 1. daļu, var teikt, ka ir klases, kas dominē, piemēram, “Alcohol” un "Sulphates". Vizuālie dati liecina, ka šīs divas klases aizņem visvairāk vietas. Vismazāk vietas aizņem klase “Total sulfur dioxide”. Kopumā, aplūkojot rezultātus, ir skaidrs, ka datu objekti ir slikti atdalāmi. Ja aplūkojam vizuālo rezultātu "Scatter Plot" (izkliedes diagramma), redzams, ka dati nav nodalīti un dominē 2-3 datu objekti. Visas datu grupas ir ļoti tuvu viena otrai. Pamatojoties uz datu mediānu, varam teikt, ka “Density” (blīvums) ir augstāks, jo tas ir vistuvāk 1 un ir 0,99668. Dispersija ir mediānas novirze.

# II daļa – Nepārraudzītā mašīnmācīšanās

**Изображение выглядит как текст, снимок экрана, диаграмма, программное обеспечение

Автоматически созданное описание**

25.att. Hierarchical Clustering Avarage

Izmantojot parametru "Avarage", es redzu 4 klasterus. Klasterī 1 var redzēt 1 objektu. 2. klasterī var redzēt 3 objektus. Klasterī 3 var redzēt 12 objektus. Kopa 4 klasterus var redzēt 47 objektus, ja tiek palielināts logs "Hierarhiskā klasterizācija".

Изображение выглядит как текст, снимок экрана, программное обеспечение, дисплей

Автоматически созданное описание

k-Means parametri:

Fixed – ļauj iestatīt klasteriem fiksētus sākotnējos centroidus.

Silhouette score - tiek izmantots, lai novērtētu klasteru kvalitāti.

Preprocessing - attiecas uz sākotnējiem datu sagatavošanas posmiem pirms klasterizācijas algoritma piemērošanas.

Initialization - attiecas uz metodi, ko izmanto, lai sākotnēji inicializētu klasteru centroidus pirms algoritma atkārtošanas.

Re-runs - attiecas uz to, cik reižu algoritms tiek atkārtots ar dažādiem nejaušiem sākotnējiem centroidiem.

**Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описание**

26.att. Hierarchical Clustering Single

Izmantojot parametru "Single", es redzu 3 klasterus. Klasterī 1 var redzēt 1 objektu. Arī 2. klasterī redzams 1 objekts. Klasterī 3 var redzēt 1139 objektus, ja tiek palielināts logs " Hierarchical Clustering".

**Изображение выглядит как текст, снимок экрана, диаграмма, программное обеспечение

Автоматически созданное описание**

27.att. Hierarchical Clustering Complete

Izmantojot parametru "Complete", es redzu 4 klasterus. Klasterī 1 var redzēt 17 objektus. Kopā 2 var redzēt tos pašus 15 objektus. Klasterī Nr. 3 var redzēt 165 objektus, klasterī Nr. 4 var redzēt 946 objektus, ja tiek palielināts logs " Hierarchical Clustering".

Изображение выглядит как текст, снимок экрана, программное обеспечение, Значок на компьютере

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, дисплей, программное обеспечение

Автоматически созданное описание

28.att. Silhouette Plot 29.att. K-Means

Šajā grafikā redzams, ka klasteri ir sadalījušās pilnīgi dažādās pusēs. Klasteri, kas atrodas labajā pusē, ir pievienoti, bet tie, kas atrodas kreisajā pusē, nav pievienoti. Manā attēlā redzams, ka lielākā daļa klasteru ir savienojušies, bet, palielinot "Silhouette plot" logu, var redzēt, ka lielākā daļa klasteru nav savienojušies

# Secinājums II daļai

Rezultāti parāda, ka atkarībā no izvēlētā klasterizācijas parametra (Average, Single, Complete) un loga (Hierarhiskā klasterizācija), tiek iegūti atšķirīgi klasteru skaiti un to objektu sadalījums klasteros. Turklāt norādīts, ka klasteri ir sadalīti dažādās pusēs un ir veikti savienojumi starp dažiem klasteriem, bet citi klasteri paliek neatkarīgi. Tas varētu liecināt par dažādu objektu grupēšanu un to savstarpējo saistību vai izolāciju. Vairāki faktori, piemēram, izvēlētie klasterizācijas parametri un logi, ietekmē rezultātus, un šīs izmaiņas var novest pie atšķirīga klasteru skaits un to sastāvs. Turklāt, papildus informācija par "Silhouette plot" loga lietošanu norāda, ka daži klasteri nav savienoti. Atkarībā no parametra k-Means iestatījuma var arī teikt, ka dati grafikos ir labāk vai sliktāk atdalāmi vai tiem ir precīzāka frekvence.

# III daļa – Pārraudzītā mašīnmācīšanās

Изображение выглядит как диаграмма, текст, линия, круг

Автоматически созданное описание

30.att. Mašīnmācīšanas testēšana

## **Algoritmu apraksts**

**Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание**

31.att kNN algoritms

kNN algoritms - ir vienkāršs klasifikācijas un regresijas algoritms, ko izmanto, lai prognozētu vērtības, pamatojoties uz tuvākajiem kaimiņiem mācību datu kopā.

Parametri:

Number of neighbors – šis parametrs nosaka tuvāko kaimiņu skaitu, kas tiks izmantoti, lai klasificētu vai regresētu jauno novērojumu.

Metric – Šis parametrs nosaka attāluma metriku, ko izmanto, lai izmērītu tuvumu starp objektiem pazīmju telpā.

Weight - šis parametrs nosaka svēršanas shēmu, kas tiks izmantota, ņemot vērā kaimiņa tuvumu jauna novērojuma klasifikācijā vai regresijā.

Es izvēlējos kNN algoritmu, jo to ir viegli izpētīt, un mācību procesā mēs mācījāmies algoritmu, tāpēc man tas ir saprotams.

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

32.att. Gradient Boosting

Gradient Boosting algoritms – šis algoritms pieder pie ansambļa mašīnmācīšanās metožu saimes, kas apvieno vairākus vājus modeļus, lai izveidotu spēcīgāku prognozēšanas modeli.

Parametri:

Number of trees – šis parametrs nosaka bāzes modeļu (lēmumu koku) skaitu, kas tiks izveidoti uz gradientu balstītā palielināšanas procesā.

Learning rate - šis parametrs nosaka, cik lielā mērā katrs bāzes modelis (lēmumu koks) veicina galīgo prognozi. Tas nosaka, cik ātri vai lēni modelis pielāgojas iepriekšējo modeļu kļūdām gradientu šķirošanas laikā.

Limit depth of individual trees – šis parametrs nosaka katra atsevišķa gradientu šķirošanā izmantotā koka dziļuma robežu.

Do not split subsets smaller than – šis parametrs nosaka datu apakškopas minimālo lielumu, pie kura tiks veikta koka sadalīšana.

Es izmantoju šo algoritmu, jo tam ir augsta prognozēšanas precizitāte. Pateicoties lielam datu apjomam, šim algoritmam ir laba datu veida apstrāde, un to var dažādi pielāgot.

Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание

33.att. Neural Network

Neural Network algoritms - šis algoritms ir mašīnmācīšanās metode, kuras pamatā ir mākslīgie neironu tīkli. Neironu tīkli imitē smadzeņu neironu sistēmas struktūru un darbību un tiek izmantoti dažādiem uzdevumiem, tostarp klasifikācijai, regresijai un attēlu apstrādei.

Parametri:

Neurons in hidden layers – šis parametrs nosaka neironu skaitu katrā neironu tīkla slēptajā slānī. Slēptie slāņi ir starpposma slāņi starp neironu tīkla ieejas un izejas slāņiem, un tiem ir svarīga loma iezīmju iegūšanā un apstrādē no ieejas datiem.

Activation – šis parametrs attiecas uz neironu tīkla aktivizācijas funkciju izvēli. Aktivizācijas funkcijas nosaka, kā neironu tīklā tiek aktivizēti un pārraidīti signāli starp neironiem.

Solver – šis parametrs attiecas uz neironu tīkla apmācībai izmantotās optimizācijas metodes izvēli. Optimizācijas metode nosaka, kā modelis atjauninās svarus un pielāgos parametrus apmācības laikā.

Regulation – šis parametrs attiecas uz regulēšanas metodēm, ko izmanto, lai kontrolētu modeļa pārtrenēšanu. Regularizācija palīdz novērst neironu tīkla pārvērtēšanu un uzlabot tā vispārināšanas spēju.

Maximum number of iterations – šis parametrs nosaka maksimālo iterāciju (epohu) skaitu, ko modelis veiks apmācības laikā. Tas nosaka augšējo robežu mācību datu kopas caurlaižu skaitam, ko modelis veiks, mēģinot sasniegt konverģenci.

**Изображение выглядит как текст, снимок экрана, программное обеспечение, число

Автоматически созданное описание**

34.att. Data Sampler

Izmantojot šo parametru, es iestatīju 85 % apmācībai un 15 % testēšanai.

## **Testēšana**

**Pirmais tests**

Parametri:

Изображение выглядит как текст, программное обеспечение, снимок экрана

Автоматически созданное описание

35.att. Algoritmi 1 tests

Изображение выглядит как текст, снимок экрана, число, программное обеспечение

Автоматически созданное описание

36.att. Test and Score 1 tests

Изображение выглядит как снимок экрана, текст, программное обеспечение, число

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, число

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, число

Автоматически созданное описание

37.att. Confusion Matrix 1.tests

Šajā testā iegūtie rezultāti tiek atklāti, izmantojot Orange bazes pamatparametrus. Iegūtie matricas parametri ir aptuveni tuvu precīziem, proti, no 6 prognozētajiem variantiem 5 tika pareizi prognozēti, kas liecina par labu rezultātu precizitāti visos algoritmos. Algoritmos dotie parametri liecina, ka vērtības ir pēc iespējas tuvākas 1, kas ir labs rezultāts.

**Otrais tests**

Parametri:

Изображение выглядит как текст, программное обеспечение, диаграмма, Значок на компьютере

Автоматически созданное описание

38.att. Algoritmi 2 tests

Изображение выглядит как текст, программное обеспечение, Значок на компьютере, веб-страница

Автоматически созданное описание

39.att. Test and Score 2 tests

Изображение выглядит как снимок экрана, текст

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, число

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, число

Автоматически созданное описание

40.att. Confusion Matrix 2 tests

Šajā testā es saņēmu rezultātus, kas iegūti, pašam tos iestatot, un īpaši, lai redzētu, pie kādiem parametriem tiks konstatētas minimālās vērtības. Iegūtie matricas parametri liecina, ka rezultāti ir minimāli prognozēti un no 6 iegūtajiem rezultātiem tikai 2 vai 1 ir prognozēts. Ar dotajiem parametriem algoritmos redzams, ka vērtības ir minimālas un nav labs rādītājs analīzei.

**Trešais tests**

Parametri:

Изображение выглядит как текст, программное обеспечение, снимок экрана, диаграмма

Автоматически созданное описание

41.att. Algoritmi 3 tests

Изображение выглядит как текст, программное обеспечение, Значок на компьютере, веб-страница

Автоматически созданное описание

42.att. Test and Score 3 tests

Изображение выглядит как снимок экрана, текст, число

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, число, программное обеспечение

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, число

Автоматически созданное описание

43.att. Confusion Matrix 3 tests

Šajā testā rezultātus ieguvu, pats tos iestatot, lai varētu salīdzināt citus algoritmos pieejamos parametrus. Iegūtie matricas parametri arī parāda aptuveni vidējās vērtības matricās ar iestatītajiem parametriem, t. i., no 6 prognozēm es iegūstu 3-4 pareizas. Ar algoritmos dotajiem parametriem redzams, ka vērtības ir vairāk tuvas 1, un no tā ir skaidrs, ka tās ir relatīvi pareizas.

# Secinājums III daļai

Pamatojoties uz 3 testiem, var secināt, ka 3 testos noteiktie parametri ietekmē rezultātus. Izmantojot algoritmu “Gradient Boosting” ar pietiekamu parametru precizitāti, var iegūt precīzus rezultātus, kas būs tik tuvi 1 vai 100 %. Tomēr šis algoritms nevar būt 100 % precīzs rezultātu aprēķināšanā, tāpēc vislabāk ir izmantot algoritmu “Neural Network”.

# Kopējais secinājums

Rezumējot šo darbu, varu teikt, ka esmu iemācījies daudz jauna. Pateicoties darbam ar lietotni Orange, esmu uzzinājis vairāk par algoritmu veidiem un to darbību. Uzzināju arī noderīgu daļu par to, kā atšķiras rezultāti un no kā tie ir atkarīgi. Es arī uzzināju, kā klasteri un algoritmi ir saistīti. Izmantojot izvēlēto datubāzi, es sapratu, ka dati datubāzē ir nesabalansēti, tāpēc daži grafiki var nebūt precīzi redzami. Izmēģinot dažādus algoritmus, konstatēju, ka visefektīvākie ir "kNN", "Neural Network" un "Gradient Boosting", bet, ja es daudz atkārtoju ar citiem algoritmiem, varu iegūt aptuveni labu rezultātu.

# Avoti

# <https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/model/knn.html>

# <https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/model/logisticregression.html>

# <https://orange3.readthedocs.io/projects/orange-visual-programming/en/latest/widgets/model/neuralnetwork.html>

# <https://www.youtube.com/watch?v=dKURyzjh5Gc>