

Duomenų tyrybos konspektas

1.	Įvadas į duomenų tyrybą	4
	<i>Testas prieš pradedant mokytis skyrių „Duomenų tyryba“</i>	<i>4</i>
1.1.	Duomenys, informacija, žinios	4
	<i>Kas yra duomenys?</i>	<i>4</i>
	<i>Duomenų, informacijos ir žinių sąryšiai</i>	<i>4</i>
	<i>Pavyzdžiai, kuriuose duomenys susiejami su informacija ir žiniomis</i>	<i>7</i>
	<i>Skyrelio 1.1. medžiagai įtvirtinti siūlomas testas su pasirenkamaisiais atsakymais.....</i>	<i>8</i>
1.2.	Duomenų tyryba	8
	<i>Duomenų tyrybos privalumai</i>	<i>10</i>
	<i>Duomenų tyrybos trūkumai</i>	<i>10</i>
	<i>Duomenų tyrybos sąryšiai su kitais mokslais (sritimis)</i>	<i>10</i>
	<i>Duomenų tyrybos taikymai.....</i>	<i>11</i>
	<i>Duomenų tyrybos įgyvendinimo iššūkiai.....</i>	<i>12</i>
	<i>Skyrelio 1.2 medžiagai įtvirtinti siūlomos užduotys</i>	<i>13</i>
1.3.	Duomenų tipai.....	16
	<i>Struktūrizuoti duomenys.....</i>	<i>16</i>
	<i>Pusiaus struktūrizuoti duomenys</i>	<i>17</i>
	<i>Nestruktūrizuoti duomenys</i>	<i>18</i>
	<i>Struktūrizuotų, pusiau struktūrizuotų ir nestruktūrizuotų duomenų pavyzdžiai, savybės ir jų palyginimas</i>	<i>18</i>
	<i>Struktūrizuotų ir nestruktūrizuotų duomenų privalumai ir trūkumai.....</i>	<i>19</i>
	<i>Struktūrizuotų ir nestruktūrizuotų duomenų pagrindiniai skirtumai</i>	<i>20</i>
	<i>Įvairių tipų duomenų vaizdavimas</i>	<i>21</i>
	<i>Duomenų šaltiniai</i>	<i>22</i>
	<i>Veiksmai su duomenimis.....</i>	<i>23</i>
	<i>Skaitmenizacija ir skaitmeninė transformacija</i>	<i>24</i>
	<i>Skyrelio 1.3 medžiagai įtvirtinti siūlomos užduotys</i>	<i>24</i>
1.4.	Duomenų tyrybos modelis	25
	<i>Testas išnagrinėjus temą</i>	<i>26</i>
	Vertinimas	26
	1 skyriaus pateiktis.....	26
2.	Apie duomenis išsamiau.....	26
2.1.	Didieji duomenys (Big Data).....	26
	<i>Svarbiausios didžiųjų duomenų charakteristikos</i>	<i>26</i>
	<i>Didžiųjų duomenų sąsajos su šiuolaikinėmis technologijomis</i>	<i>29</i>
	<i>Didžiųjų duomenų paslaugų architektūra</i>	<i>31</i>

Didžiųjų duomenų keliama iššūkiai	34
Skyrelio 2.1. medžiagai įtvirtinti siūlomas testas su pasirenkamaisiais atsakymais (anglų k.) .	35
Šaltinis: https://study.com/academy/exam/topic/big-data-fundamentals.html	35
2.2. Atviri duomenys	36
Atvirų duomenų apibrėžimas ir savybės.....	36
Atvirų duomenų šaltiniai	37
Skyrelio 2.2. medžiagai įtvirtinti siūlomas kursas apie atvirus duomenis: https://data.europa.eu/elearning/en/#/id/co-01	37
2.3. Kintamieji ir matavimo skalės	37
Skyrelio 2.3. medžiagai įtvirtinti siūlomi klausimai	38
2.4. Vertinimas.....	38
3. Skaičiuoklės naudojimas duomenų tyryboje.....	39
3.1. Kaip greitai susirasti norimos funkcijos aprašymą ir panaudojimo pavyzdžių?.....	39
3.2. Kaip teisingai užrašyti funkcijos argumentus?	40
3.3. Funkcijos.....	40
Matematinės, trigonometrinės, statistinės funkcijos	40
Loginės funkcijos.....	42
Datos ir laiko funkcijos	43
Teksto funkcijos	43
Duomenų sujungimas panaudojant ampersando (&) simbolį	43
Funkcija VLOOKUP	43
3.4. Nuorodos į praktines užduotis	44
Užduotys funkcijų, kurios buvo ankstesnėje programoje, mokymuisi ir įgūdžių įtvirtinimui.....	44
Naujai įvestų funkcijų paaiškinimai, pavyzdžiai mokymuisi	44
Šaltiniai ir idėjos užduotims ir projektams.....	54
Chat GPT pasiūlymai (kalba netaisyta)	55

1. Įvadas į duomenų tyrybą

Testas prieš pradedant mokytis skyrių „Duomenų tyryba“

1. Kokį vaidmenį vaidina žodis „tyryba“ duomenų tyryboje?

- a) Tai reiškia, kad analizuojant duomenis bus taikomi moksliniai metodai.
- b) Tik specialistai, turintys atitinkamą išsilavinimą, tai supranta.
- c) Toks žodžių junginys gerai skamba.

2. Teiginys „Mokytis duomenų tyrybos yra naudinga tik programuotojams“ yra:

- a) Teisingas
- b) Klaidingas.

3. Ką turėtumėte atlikti, norėdami pademonstruoti, ar krepšinio komandos žaidėjai yra yra aukštesni, negu vidutinio ūgio žmonės?

- a) Surinkti duomenis.
- b) Turėti žinių apie statistiką ir tikimybes.
- c) Teisingi a) ir b) variantai.

1.1. Duomenys, informacija, žinios

Kas yra duomenys?

Kasdieniam gyvenime mus nuolat supa duomenys. Tekstas, kurį dabar skaitote, yra duomenys. Jūsų draugų telefonų numerių sąrašas išmaniajame telefone yra duomenys, taip pat ir laikrodžio rodomas dabartinis laikas. Duomenimis operuojame skaičiuodami turimus pinigus arba rašydami laiškus draugams.

Sukūrus kompiuterius duomenys tapo daug svarbesni. Pagrindinis kompiuterių vaidmuo – atlikti skaičiavimus, tačiau jiems reikia duomenų, kuriais jie galėtų operuoti. Todėl turime suprasti, kaip kompiuteriai saugo ir apdoroja duomenis.

Atsiradus internetui, kompiuterių, kaip duomenų tvarkymo įrenginių, vaidmuo padidėjo. Pagalvojus, dabar kompiuterius vis dažniau naudojame duomenims apdoroti ir bendrauti, o ne faktiniams skaičiavimams atlikti. Kai rašome elektroninį laišką draugui arba ieškome informacijos internete – iš esmės kuriame, saugome, perduodame ir tvarkome duomenis.

Duomenų, informacijos ir žinių sąryšiai

(<https://data.europa.eu/elearning/en/module1/#/id/co-01>)

Duomenys yra žaliava, iš kurios galima gauti informacijos ir žinių. Įsivaizduokite duomenis kaip vietas, vaizdus, aprašymus, atsiliepimus ir kainas, kurie sudaro informacijos, galinčios padėti planuoti, pavyzdžiui, atostogas, pagrindą.

Duomenys tampa **informacija**, kai jiems suteikiamas kontekstas. Remiantis ankstesniu pavyzdžiu, vietos, vaizdai, aprašymai ir kainos – visa tai gali padėti pateikti informaciją, susijusią su turistų lankomomis vietomis. Duomenų rinkimas ir pateikimas padeda formuoti informaciją.

Žinios yra tai, kas gaunama iš informacijos ir pritaikoma jūsų poreikiams. Žinių kaupimas – tai procesas, kurio metu informacija paverčiama pasirinkimais. Remdamiesi ankstesniu pavyzdžiu, žinodami, kad niekas iš jūsų šeimos narių nemėgsta pramogų parkų, galėsite nuspręsti, kurių vietų atostogų metu vengti ir kurios labiau tiktų jūsų šeimai.



DIKW piramidė

Šaltinis: Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*, 33(2), 163-180.



• Duomenys

- Neturi reikšmės ar vertės, nes jie neturi konteksto ir interpretacijos.
- Atskiri, objektyvūs faktai arba stebėjimai, kurie yra neorganizuoti ir neapdoroti, ir neperteikia jokios konkrečios prasmės
- Elementarus ir užfiksuotas daiktų, įvykių, veiklos ir sandorių aprašymas.
- Dažnai yra didesnių fizinių sistemų elementai (pvz. prietaisų skydeliai), kurie suteikia užuominų apie tai, kokius duomenis reikia pastebėti ir kaip juos skaityti.

Informacija

- Suformatuoti duomenys, kuriuos galima apibrėžti kaip tikrovės atvaizdavimą.
- Duomenys, kurie suteikia pridėtinės vertės dalyko supratimui.

- Duomenys, kurie buvo sutvarkyti taip, kad jie turi prasmę ir vertę gavėjui.
- Duomenys, apdoroti siekiant tam tikro tikslo.

Žinios

- Duomenų ir informacijos derinys, prie kurio pridedama ekspertų nuomonė, įgūdžiai, ir patirtis, kad būtų gautas vertingas turtas, kuriuo galima naudotis priimant sprendimus.
- Tai duomenys ir (arba) informacija, kurie buvo susisteminti ir apdoroti, siekiant perteikti supratimą, patirtį, sukauptą mokymąsi ir kompetenciją, taikomus dabartinei problemai ar veiklai.
- Žinios grindžiamos informacija, kuri išgaunama iš duomenų. Nors duomenys yra daiktų savybė, žinios yra žmonių savybė, kuri juos skatina veikti tam tikru būdu.

Procesai, kurių metu informacija paverčiama į žinias apibūdinami įvairiai:

- daugelio informacijos šaltinių sintezė per tam tikrą laiką;
- įsitikinimų struktūrizavimas;
- studijos ir patirtis;
- organizavimas ir apdorojimas, siekiant perteikti supratimą, patirtį, sukauptą mokymąsi ir patirtį.

Žinios yra:

- kontekstinės informacijos, vertybių, patirties ir taisyklių derinys;
- informacija, ekspertų nuomonė, įgūdžiai ir patirtis;
- informacija kartu su supratimu ir gebėjimais;
- suvokimas, įgūdžiai, mokymas(is), sveikas protas ir patirtis.

Išmintis (įžvalgos)

- Išmintis yra labai neapčiuopiama sąvoka. Ji labiau susijusi su žmogaus intuicija, supratimu, aiškinimu ir veiksmais nei su sistemomis.
- Išmintis laikoma sukauptomis žiniomis, kurios leidžia suprasti, kaip taikyti vienos srities sąvokas naujoms situacijoms ar problemoms.
- Išmintis yra aukščiausias abstrakcijos lygis, pasižymintis įžvalgumu ir gebėjimu matyti už horizonto.
- Tai gebėjimas kritiškai arba praktiškai elgtis bet kurioje situacijoje.

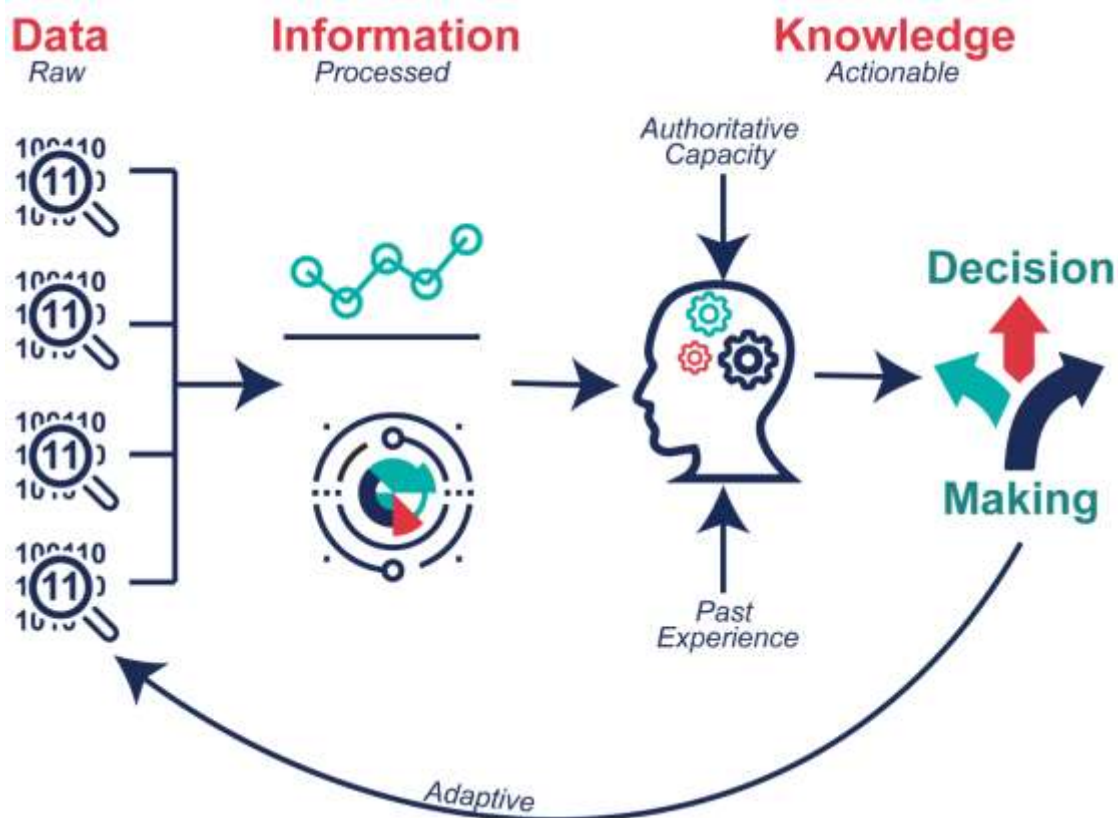
Ši palyginti nedidelė išminties sąvokos aptarimo apimtis rodo, kad platesnėje informacinių sistemų, žinių vadybos ir vadybos literatūroje diskusijoms apie išminties prigimtį ir jos ugdymo būdus skiriama nedaug dėmesio.

Šaltinis: <https://internetofwater.org/valuing-data/what-are-data-information-and-knowledge/>

Duomenys – tai neapdorotos vertės (skaitinės, tekstinės ir pan.), gautos taikant įvairius duomenų gavimo metodus.

Informacija sukurama, kai duomenys apdorojami, organizuojami arba struktūrizuojami, siekiant suteikti kontekstą ir prasmę. Informacija iš esmės yra apdoroti duomenys.

Žinios yra tai, ką mes žinome. Kiekvieno žmogaus žinios yra unikalios, tai yra sukaupta ankstesnė patirtis ir įžvalgos, kuriomis remdamiesi aiškiname informaciją ir suteikiame jai prasmę. Kad žinios virstų veiksmais, asmuo turi turėti įgaliojimus ir gebėjimus priimti ir įgyvendinti sprendimą. Žinios (ir įgaliojimai) reikalingos tam, kad būtų gauta informacija, kuri gali turėti poveikį.



Duomenų srautas į informaciją ir žinias nėra vienakryptis. Gautos žinios gali atskleisti surinktų duomenų perteklių ar trūkumą. Dėl to, siekiant geriau patenkinti naudotojų poreikius, gali tekti keisti surinktus duomenis arba tai, kaip šie duomenys paverčiami informacija.

Duomenų, informacijos ir žinių charakteristikos

Duomenys	Informacija	Žinios
Objektyvūs	Turi būti objektyvi	Subjektyvios
Neturi prasmės	Turi prasmę	Turi prasmę, skirtą specifiniam tikslui
Neapdoroti	Apdoroti	Apdoroti ir suprasti
Kiekybiškai įvertinami, gali būti duomenų perteklius	Kiekybiškai įvertinama, gali būti informacijos perteklius	Kiekybiškai neįvertinamos, negali būti pertekliaus

Pavyzdžiai, kuriuose duomenys susiejami su informacija ir žiniomis

- Duomenys: Oro temperatūra yra 25 laipsniai. Informacija: Šiandien yra karšta diena. Žinios: Per karštas oras gali sukelti dehidrataciją ir kitus sveikatos sutrikimus.
- Duomenys: 60% žmonių, gyvenančių miesto centre, naudojami viešuoju transportu. Informacija: Miesto centro gyventojai labiau linkę naudoti viešąjį transportą nei kitos vietos gyventojai. Žinios: Viešojo transporto sistema miesto centre yra gerai išvystyta, o tai rodo, kad galima sumažinti privačių transporto priemonių naudojimą ir šiltnamio efektą.

- Duomenys: Statistika rodo, kad 70% vaikų turi savo mobilųjį telefoną. Informacija: Daugelis vaikų turi mobilųjį telefoną. Žinios: Vaikų elgesys ir poreikiai pasikeitė, o tai reiškia, kad reikia atnaujinti mokyklų technologijų planus ir svarbu padidinti interneto saugumą, kad apsaugotume vaikus nuo įvairių pavojų.
- Duomenys: Didelis skaičius studentų pasirinko inžinerijos specialybes. Informacija: Inžinerijos studijos populiarėja tarp studentų. Žinios: Dėl didėjančio technologijų ir pramonės sektoriaus vystymosi, inžinerijos specialistams yra didelis poreikis ir jie turi galimybę pasiekti aukštus atlyginimus ir pasirinkti iš įvairių karjeros kelių.

Skyrelio 1.1. medžiagai įtvirtinti siūlomas testas su pasirenkamaisiais atsakymais

Testas (<https://www.propofs.com/quiz-school/quizshow.php?title=data-information-knowledge&q=1>)

1. _____ gali būti paveikslėliai, skaičiai, tekstas be konteksto
 A) Duomenys
 B) Informacija
 C) Žinios
 D) Išmintis (wisdom)
2. Kai duomenys tampa reikšmingais gavėjui, tai vadinama _____
 A) Duomenų rinkiniu
 B) Informacija
 C) Žiniomis
 D) Tinkamo atsakymo nėra
3. Kuris iš žemiau pateiktų teiginių yra teisingas?
 A) Informacija yra duomenų dalis
 B) Duomenys yra informacijos dalis
 C) Duomenys ir informacija yra priešingos prigimtys.
 D) Tinkamo atsakymo nėra.
4. Ar teisingas teiginys: „Kad duomenys būtų naudingi, jie turi virsti informacija“?
 A) Taip
 B) Ne
5. Kuriuos iš žemiau išvardintų pavyzdžių galima laikyti duomenimis?
 A) Datos
 B) Kainos
 C) Išlaidos
 D) Visi išvardinti pavyzdžiai yra duomenys
6. Kurie iš žemiau išvardintų pavyzdžių yra informacija?
 A) Ataskaita
 B) Receptas
 C) Asmens tapatybės kortelė
 D) Visi išvardinti pavyzdžiai yra informacija

1.2. Duomenų tyryba

<https://microsoft.github.io/Data-Science-For-Beginners/#/1-Introduction/01-defining-data-science/README>

Duomenų tyryba apibrėžiama kaip mokslo sritis, kuri naudoja mokslinius metodus žinioms ir išvalgoms iš struktūrizuotų ir nestruktūrizuotų duomenų išgauti, taip pat žinioms ir išvalgoms iš duomenų taikyti įvairiose gyvenimo srityse.

Šiame apibrėžime pabrėžiami šie svarbūs duomenų tyrybos aspektai:

Pagrindinis duomenų tyrybos tikslas yra išgauti žinias iš duomenų, kitaip tariant, suprasti duomenis, rasti paslėptus ryšius ir sukurti modelį.

Duomenų tyryba naudoja mokslinius metodus, pavyzdžiui, tikimybių ir statistikos. Tiesą sakant, kai pirmą kartą buvo pradėtas vartoti terminas duomenų tyryba, kai kurie žmonės teigė, kad duomenų tyryba tėra naujas madingas statistikos pavadinimas. Šiandien tapo akivaizdu, kad ši sritis yra daug platesnė.

Gautos žinios turėtų būti pritaikytos tam, kad būtų gauta tam tikrų praktiškai pritaikomų išvalgų, t. y. praktinių išvalgų, kurias galima pritaikyti įvairiose situacijose.

Turėtume gebėti dirbti ir su struktūrizuotais, ir su nestruktūrizuotais duomenimis. Prie skirtingų duomenų tipų aptarimo grįšime vėliau.

Taikymo sritis yra svarbi sąvoka, ir duomenų mokslininkams dažnai reikia turėti bent tam tikrų žinių apie probleminę sritį, pavyzdžiui: finansai, medicina, rinkodara ir pan.

Kitas svarbus duomenų tyrybos aspektas yra tai, kad ji tiria, kaip duomenis galima rinkti, saugoti ir valdyti naudojant kompiuterius. Statistika mums suteikia matematinius pagrindus, o duomenų tyryba taiko matematines sąvokas, kad iš tikrųjų iš duomenų būtų galima gauti išvalgų.

Jim Gray teigia, kad duomenų tyrybą galima laikyti atskira mokslo paradigma, kuri yra:

- **Empirinė**, kurioje daugiausia remiamasi stebėjimais ir eksperimentų rezultatais.
- **Teorinė**, kai naujos sąvokos atsiranda iš esamų mokslo žinių.
- **Kompiuterinė**, kai naujus principus atrandame remdamiesi tam tikrais kompiuteriniais eksperimentais.
- **Duomenimis grindžiama**, pagrįsta ryšių ir dėsningumų atradimu duomenyse.

<https://www.javatpoint.com/data-mining>

Informacijos gavybos procesas, kurio metu iš didžiulių duomenų rinkinių nustatomi dėsningumai, tendencijos ir naudingi duomenys, leidžiantys verslui priimti duomenimis pagrįstus sprendimus, vadinamas **duomenų tyryba**.

Kitaip tariant, galima sakyti, kad **duomenų tyryba** – tai paslėptų informacijos modelių tyrimo procesas, kurio metu įvairiais aspektais informacija klasifikuojama į naudingus duomenis, kurie renkami ir kaupiami tam tikrose srityse, pavyzdžiui, duomenų saugyklose, veiksminga analizė, duomenų gavybos algoritmas, padedantis priimti sprendimus ir kitus duomenų reikalavimus, kad galiausiai būtų galima sumažinti išlaidas ir gauti pajamų.

Duomenų tyryba – tai automatinė didelių informacijos sankaupų paieška siekiant rasti tendencijas ir dėsningumus, kurie pranoksta paprastas analizės procedūras. Duomenų tyryba naudoja sudėtingus matematinius algoritmus duomenų segmentams ir įvertina būsimų įvykių tikimybę. Duomenų tyryba taip pat vadinama duomenų žinių atradimu (angl. Knowledge Discovery of Data, KDD).

Duomenų tyryba – tai procesas, kurį organizacijos naudoja siekdamos iš didžiulių duomenų bazių išgauti konkrečius duomenis verslo problemoms spręsti. Jis pirmiausia neapdorotus duomenis paverčia naudinga informacija.

Duomenų tyryba yra panaši į duomenų mokslą, kurį vykdo asmuo, konkrečioje situacijoje, konkrečiam duomenų rinkiniui, turėdamas tam tikrą tikslą. Šis procesas apima įvairių rūšių paslaugas, pavyzdžiui, teksto tyrybą, žiniatinklio tyrybą, garso ir vaizdo įrašų tyrybą, vaizdinių

duomenų tyrybą ir socialinės žiniasklaidos tyrybą. Tyryba atliekama naudojant paprastą arba labai specifinę programinę įrangą.

Didžiausias iššūkis – išanalizuoti duomenis ir iš jų išgauti svarbią informaciją, kurią galima panaudoti problemai spręsti arba įmonės plėtrai. Yra daug galingų priemonių ir metodų, kuriais galima išgauti duomenis ir rasti iš jų geresnių įžvalgų.

Duomenų tyrybos privalumai

- Duomenų tyrybos metodas leidžia organizacijoms gauti žiniomis pagrįstus duomenis.
- Duomenų tyryba leidžia organizacijoms atlikti pelningus veiklos ir gamybos pakeitimus.
- Palyginti su kitomis statistinių duomenų taikymo sritimis, duomenų tyryba yra ekonomiškai efektyvi.
- Duomenų tyryba padeda organizacijos sprendimų priėmimo procese.
- Ji palengvina automatinį paslėptų dėsningumų atradimą, taip pat tendencijų ir elgsenos prognozavimą.
- Ją galima įdiegti tiek naujoje sistemoje, tiek esamose platformose.
- Tai greitas procesas, kuris naujiems naudotojams leidžia per trumpą laiką išanalizuoti didžiulius duomenų kiekius.

Duomenų tyrybos trūkumai

- Yra tikimybė, kad organizacijos gali parduoti naudingus klientų duomenis kitoms organizacijoms už pinigus. Kaip teigiama pranešime, "American Express" pardavė savo klientų kredito kortelių pirkinius kitoms organizacijoms.
- Daugeliu duomenų tyrybos analitinės programinės įrangos yra sudėtinga naudotis, o darbui su ja reikia išankstinio mokymo.
- Skirtingos duomenų tyrybos priemonės veikia skirtingai dėl jų konstrukcijoje naudojamų skirtingų algoritmų. Todėl tinkamų duomenų tyrybos priemonių pasirinkimas yra labai sudėtinga užduotis.
- Duomenų tyrybos metodai nėra tikslūs, todėl tam tikromis sąlygomis gali sukelti sunkių padarinių.

Duomenų tyrybos sąryšiai su kitais mokslais (sritimis)

<https://microsoft.github.io/Data-Science-For-Beginners/#/1-Introduction/01-defining-data-science/README>

Duomenų bazės

Svarbiausias aspektas yra **kaip saugoti** duomenis, t. y. kaip juos struktūrizuoti taip, kad būtų galima greičiau apdoroti. Yra įvairių tipų duomenų bazių, kuriose saugomi struktūrizuoti ir nestruktūrizuoti duomenys, kuriuos ir aptarsime savo kurse.

Didieji duomenys

Dažnai mums reikia saugoti ir apdoroti labai didelius duomenų kiekius, kurių struktūra yra palyginti paprasta. Yra specialių metodų ir įrankių, skirtų tiems duomenims saugoti paskirstytu būdu kompiuterių klasteryje ir juos efektyviai apdoroti.

Mašininis mokymasis

Vienas iš būdų suprasti duomenis – **sukurti modelį**, kuris galėtų numatyti norimą rezultatą. Modelių kūrimas iš duomenų vadinamas **mašininio mokymusi**. Mašininį mokymąsi nagrinėsime skyriuje „Dirbtinis intelektas“.

Dirbtinis intelektas

Mašininio mokymosi sritis, vadinama dirbtiniu intelektu (DI), taip pat remiasi duomenimis ir apima didelio sudėtingumo modelių, imituojančių žmogaus mąstymo procesus, kūrimą. Dirbtinio intelekto metodai dažnai leidžia nestruktūrizuotus duomenis (pvz., natūralią kalbą) paversti struktūrizuotomis įžvalgomis.

Vizualizavimas

Didžiuliai duomenų kiekiai žmogui yra nesuprantami, tačiau, sukūrę naudingas vizualizacijas, naudodami šiuos duomenis, galime juos geriau suprasti ir padaryti tam tikras išvadas. Taigi, svarbu žinoti daugybę informacijos vizualizavimo būdų. Duomenų vizualizavimui bus skirtas atskiras skyrius.

Duomenų tyrybos taikymai

<https://www.javatpoint.com/data-mining>

Duomenų tyrybą visų pirma naudoja organizacijos, turinčios intensyvius vartotojų poreikius – mažmeninės prekybos, komunikacijos, finansų, rinkodaros bendrovės, nustatyti kainą, vartotojų pageidavimus, produkto pozicionavimą ir poveikį pardavimams, klientų pasitenkinimą ir įmonės pelną. Duomenų tyryba leidžia mažmeninės prekybos įmonei naudoti pardavimo vietoje esančius įrašus apie klientų pirkinius ir kurti produktus bei akcijas, kurios padeda organizacijai pritraukti klientą.

Sveikatos priežiūros srityje

Duomenų tyryba sveikatos priežiūroje turi puikių galimybių pagerinti sveikatos sistemą. Joje naudojami duomenys ir analizė, siekiant geresnių įžvalgų ir nustatyti geriausią praktiką, kuri pagerintų sveikatos priežiūros paslaugas ir sumažintų išlaidas. Analitikai naudoja tokius duomenų tyrybos metodus kaip mašininis mokymasis, daugiamatė duomenų bazė, duomenų vizualizavimas, minkštieji skaičiavimai ir statistika. Duomenų tyryba gali būti naudojama kiekvienos kategorijos pacientams prognozuoti. Šios procedūros užtikrina, kad pacientai gautų intensyvią priežiūrą tinkamoje vietoje ir tinkamu laiku. Duomenų tyryba taip pat leidžia sveikatos priežiūros draudikams atpažinti sukčiavimą ir piktnaudžiavimą.

Atliekant rinkos krepšelio analizę

Rinkos krepšelio analizė yra hipoteze pagrįstas modeliavimo metodas. Jei perkate tam tikrą produktų grupę, tai labiau tikėtina, kad pirksite ir kitą produktų grupę. Šis metodas gali leisti mažmenininkui suprasti pirkėjo pirkimo elgseną. Šie duomenys gali padėti mažmenininkui suprasti pirkėjo reikalavimus ir atitinkamai pakeisti prekių parduotuvėje išdėstymą. Naudojant kitą analitinį metodą galima palyginti rezultatus tarp skirtingų parduotuvių, tarp skirtingų demografinių grupių pirkėjų.

Švietime

Duomenų tyryba švietime yra naujai besiformuojanti sritis, susijusi su metodų, kuriais tiriamos žinios iš duomenų, gautų švietimo aplinkoje, kūrimu. Pripažįstama, kad EDM tikslai yra patvirtinti būsimą mokinių mokymosi elgseną, tirti švietimo pagalbos poveikį ir skatinti mokymąsi mokyti. Organizacija gali naudoti duomenų tyrybą siekdama priimti tikslus sprendimus, taip pat numatyti mokinio rezultatus. Turėdama rezultatus, įstaiga gali susitelkti į tai, ko ir kaip mokyti.

Gamybos inžinerijoje

Žinios yra geriausias gamybos įmonės turimas turtas. Duomenų tyrybos įrankiai gali būti naudingi norint rasti dėsningumus sudėtingame gamybos procese. Duomenų tyryba gali būti naudojama sistemos lygmens projektavime, siekiant gauti sąsajas tarp gaminio architektūros,

gaminų portfelio ir klientų duomenų poreikių. Be kitų užduočių, ji taip pat gali būti naudojama prognozuojant gaminio kūrimo laikotarpį, sąnaudas ir lūkesčius.

CRM (ryšių su klientais valdymo) srityje

Ryšių su klientais valdymas (CRM) yra susijęs su klientų pritraukimu ir išlaikymu, klientų lojalumo didinimu ir į klientus orientuotų strategijų įgyvendinimu. Norėdama užmegzti tinkamus santykius su klientais, verslo organizacija turi rinkti duomenis ir juos analizuoti. Naudojant duomenų tyrybos technologijas, surinktus duomenis galima naudoti analizei.

Nustatant sukčiavimą

Dėl sukčiavimo prarandama milijardai dolerių. Tradiciniai sukčiavimo aptikimo metodai reikalauja daug laiko ir yra sudėtingi. Duomenų tyryba suteikia prasmingus modelius ir paverčia duomenis informacija. Ideali sukčiavimo aptikimo sistema turėtų apsaugoti visų naudotojų duomenis. Prižiūrėjus metodus sudaro pavyzdinių įrašų rinkimas, o šie įrašai klasifikuojami kaip sukčiavimo arba nesukčiavimo. Naudojant šiuos duomenis sudaromas modelis, o pagal šį metodą nustatoma, ar dokumentas yra apgaulingas, ar ne.

Nustatant melą

Nusikaltėlio sulaikymas nėra didelė problema, tačiau išaiškinti tiesą yra labai sudėtinga užduotis. Teisėsaugos institucijos gali naudoti duomenų tyrybos metodus nusikaltimams tirti, įtariamų teroristų ryšiams stebėti ir pan. Šis metodas taip pat apima teksto tyrybą ir juo ieškoma prasmingų dėsningumų duomenyse, kurie paprastai yra nestruktūruotas tekstas. Lyginama ankstesnių tyrimų metu surinkta informacija ir sudaromas melo aptikimo modelis.

Bankininkystėje

Bankų sistemos skaitmeninimas turėtų generuoti milžinišką duomenų kiekį, kai atliekama kiekviena nauja operacija. Duomenų tyrybos metodai gali padėti bankininkams sprendžiant su verslu susijusias bankininkystės ir finansų problemas, nustatant tendencijas, atsitiktinumus ir sąsajas verslo informacijoje ir rinkos kaštus, kurie vadovams ar vadybininkams nėra iš karto akivaizdūs, nes duomenų apimtis yra per didelė arba juos ekspertai per greitai sukuria ekrane. Vadovas gali rasti šiuos duomenis, kad galėtų geriau orientuotis į pelningą klientą, jį įsigyti, išlaikyti, segmentuoti ir išlaikyti.

Duomenų tyrybos įgyvendinimo iššūkiai

<https://www.javatpoint.com/data-mining>

Nors duomenų tyryba yra labai galinga, ją vykdant susiduriama su daugybe iššūkių. Įvairūs iššūkiai gali būti susiję su našumu, duomenimis, metodais, technikomis ir kt.

Neišsamūs ir triukšmingi duomenys

Duomenų tyryba – tai naudingų duomenų išgavimo iš didelio kiekio duomenų procesas. Duomenys realiame pasaulyje yra nevienalyčiai, neišsamūs ir triukšmingi. Didžiuliais kiekiais pateikti duomenys paprastai būna netikslūs arba nepatikimi. Šios problemos gali kilti dėl duomenų matavimo prietaiso arba dėl žmoniškųjų klaidų. Tarkime, mažmeninės prekybos tinklas renka klientų, kurie išleidžia daugiau nei 500 JAV dolerių, telefono numerius, o apskaitos darbuotojai šią informaciją įtraukia į savo sistemą. Įvesdamas telefono numerį žmogus gali padaryti klaidą, todėl duomenys bus neteisingi. Net kai kurie klientai gali nenorėti atskleisti savo telefono numerių, todėl duomenys būna neišsamūs. Duomenys gali būti pakeisti dėl žmogaus ar sistemos klaidos. Visos šios pasekmės (triukšmingi ir neišsamūs duomenys) apsunkina duomenų tyrybą.

Duomenų pasiskirstymas

Realaus pasaulio duomenys paprastai saugomi įvairiose platformose paskirstytosios kompiuterijos aplinkoje. Jie gali būti duomenų bazėje, atskirose sistemose arba net internete. Praktiškai visus duomenis perkelti į centralizuotą duomenų saugyklą yra gana sudėtinga užduotis, daugiausia dėl organizacinių ir techninių priežasčių. Pavyzdžiui, įvairūs regioniniai biurai gali turėti savo serverius duomenims saugoti. Neįmanoma visų biurų duomenis saugoti centriniame serveryje. Todėl duomenų tyrybai reikia sukurti priemones ir algoritmus, kurie leistų išgauti paskirstytus duomenis.

Sudėtingi duomenys

Tai gali būti daugialypės terpės duomenys, įskaitant garso ir vaizdo įrašus, vaizdus, sudėtingus duomenis, erdvinis duomenis, laiko eilutes ir pan. Šių įvairių tipų duomenų valdymas ir naudingos informacijos išgavimas yra sudėtinga užduotis. Dažniausiai, norint gauti konkrečią informaciją, reikėtų tobulinti naujas technologijas, naujas priemones ir metodikas.

Veikimas

Duomenų tyrybos sistemos našumas visų pirma priklauso nuo naudojamų algoritmų ir metodų efektyvumo. Jei sukurtas algoritmas ir metodai neatitinka reikalavimų, tai neigiamai paveiks duomenų tyrybos proceso efektyvumą.

Duomenų privatumas ir saugumas

Duomenų tyryba paprastai sukelia rimtų problemų, susijusių su duomenų saugumu, valdymu ir privatumu. Pavyzdžiui, jei mažmenininkas analizuoja informaciją apie įsigytas prekes, jis atskleidžia duomenis apie pirkėjų pirkimo įpročius ir pageidavimus be jų sutikimo.

Duomenų vizualizavimas

Duomenų tyrybos srityje duomenų vizualizavimas yra labai svarbus procesas, nes tai yra pagrindinis metodas, kuriuo išvestis vartotojui parodoma patraukliai. Duomenys turi tiksliai perteikti tai, ką jais norima išreikšti. Tačiau daug kartų tiksliai ir paprastai pateikti informaciją galutiniam naudotojui yra sudėtinga. Kadangi įvesties duomenys ir išvesties informacija yra sudėtingi, norint, kad tai pavyktų, reikia įgyvendinti labai veiksmingus ir sėkmingus duomenų vizualizavimo procesus.

Skyrelio 1.2 medžiagai įtvirtinti siūlomos užduotys

Pasirinktinai:

1. Sukurti minčių žemėlapi, kuriame struktūrizuotai pateikiama 1.2 skyrelio informacija.

Slenkstinis pasiekimų lygis: mokiniai, padedami mokytojo, sukuria minčių žemėlapi naudodami įrankius, skirtus minčių žemėlapių kūrimui.

Patenkinamas pasiekimų lygis: mokiniai sukuria minčių žemėlapi savarankiškai, žemėlapyje gali trūkti kai kurių dalių.

Pagrindinis pasiekimų lygis: mokiniai sukuria minčių žemėlapi savarankiškai, žemėlapis pilnai atskleidžia esminius duomenų tyrybos aspektus, naudojimo tikslus, privalumus ir trūkumus, sąryšius su kitomis sritimis.

Aukštesnysis pasiekimų lygis: : mokiniai sukuria minčių žemėlapi savarankiškai, žemėlapis pilnai atskleidžia esminius duomenų tyrybos aspektus, naudojimo tikslus, privalumus ir trūkumus, sąryšius su kitomis sritimis, taikymus, įgyvendinimo iššūkius.

Nemokami minčių žemėlapių kūrimo įrankiai

Internete:

<https://products.aspose.app/diagram/lt/mindmap>

<https://www.mindonmap.com/lt/>

<https://www.canva.com/graphs/mind-maps/>

Įdiegiami kompiuteryje (nereikalauja interneto ryšio):

<https://freemind.sourceforge.net/wiki/index.php/Download>

<https://sourceforge.net/projects/freeplane/files/>

Alternatyva

2. Sukurti žodžių debesį iš 1.2 skyrelio informacijos.

Mokiniam pateikiamas failas tyryba.docx.

Slenkstinis pasiekimų lygis: mokiniai, padedami mokytojo, sukuria žodžių debesį naudodami įrankius, skirtus žodžių debesų kūrimui.

Patenkinamas pasiekimų lygis: mokiniai sukuria žodžių debesį savarankiškai.

Pagrindinis pasiekimų lygis: mokiniai sukuria žodžių debesį savarankiškai, peržiūri, rankiniu būdu iš failo pašalina parazitinius žodžius ir sukuria žodžių debesį be pašalinių žodžių.

Aukštesnysis pasiekimų lygis: mokiniai sukuria žodžių debesį generuojančią programą (pvz., Python programavimo kalba).

Nemokami žodžių debesų kūrimo įrankiai

Internete:

<https://www.wordclouds.com/> (palaiko lietuviškas raides, galima įkopi juoti norimą tekstą ir sugeneruoti žodžių debesį)

<https://freetools.textmagic.com/word-cloud-generator> (palaiko lietuviškas raides, galima įkopi juoti norimą tekstą ir sugeneruoti žodžių debesį)

<https://www.presentermedia.com/wordclouds#designs> (palaiko lietuviškas raides, labai daug šablonų, tinka meniškiems mokiniams).

Pavyzdys, kaip užduotį galėtų atlikti aukštesniojo pasiekimų lygio mokiniai


```
min_font_size = 10).generate(text)

plt.figure(figsize = (8, 8), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)
plt.show()
```

1.3. Duomenų tipai

Skiriami 3 pagrindiniai duomenų tipai: **struktūrizuoti**, **pusiau struktūrizuoti** ir **nestruktūrizuoti**. Toliau pateikiami įvairių duomenų tipų apibrėžimai iš įvairių šaltinių.

Struktūrizuoti duomenys

- Duomenys, kurių lengviausia ieškoti ir tvarkyti, nes jie paprastai pateikiami lentelėse, o jų elementus galima suskirstyti į fiksuotus iš anksto apibrėžtus laukus, vadinami struktūrizuotais duomenimis. Pagalvokite, kokius duomenis galite saugoti "Excel" skaičiuoklėje, ir turėsite struktūrizuotų duomenų pavyzdį. Struktūrizuoti duomenys gali atitikti duomenų bazės dizainerio sukurtą duomenų modelį - prisiminkite pardavimo įrašus pagal regionus, produktus ar klientus. Struktūrizuotuose duomenyse esybės gali būti sugrupuotos ir sudaryti ryšius („klientai“, kurie taip pat yra „patenkinti paslauga“). Dėl to struktūrizuotus duomenis lengva saugoti, analizuoti ir ieškoti, todėl dar visai neseniai jie buvo vieninteliai įmonėms lengvai pritaikomi duomenys. Šiandien, daugumos vertinimu, struktūrizuoti duomenys sudaro mažiau nei 20 proc. visų duomenų. Dažnai struktūrizuoti duomenys tvarkomi naudojant struktūrizuotų užklausų kalbą (SQL) - programavimo programinės įrangos kalbą, kurią praėjusio amžiaus septintajame dešimtmetyje sukūrė IBM, skirtą reliacinėms duomenų bazėms.

Struktūrizuotus duomenis gali kurti mašinos ir žmonės. Struktūrizuotų duomenų pavyzdžiai - finansiniai duomenys, pavyzdžiui, apskaitos operacijos, adresų duomenys, demografinė informacija, klientų žvaigždučių įvertinimai, mašinų žurnalai, išmaniųjų telefonų ir išmaniųjų įrenginių vietos nustatymo duomenys ir kt.

(<https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/?sh=6727d9952b4d>).

- Struktūrizuoti duomenys, paprastai priskiriami kiekybiniais duomenimis, yra labai gerai organizuoti ir lengvai iššifruojami mašininio mokymosi algoritmais. 1974 m. IBM sukurta struktūrizuotų užklausų kalba (SQL) yra programavimo kalba, naudojama struktūrizuotiems duomenims tvarkyti. Naudodami reliacinę (SQL) duomenų bazę, verslo naudotojai gali greitai įvesti, ieškoti ir tvarkyti struktūrizuotus duomenis (<https://www.ibm.com/cloud/blog/structured-vs-unstructured-data>).
- Struktūrizuoti duomenys yra kiekybiniai, labai gerai organizuoti ir lengvai analizuojami naudojant duomenų analizės programinę įrangą. Jie formatuojami į sistemas, kurios turi taisyklingą dizainą, telpa į nustatytas eilutes, stulpelius ir lenteles. Struktūrizuotų užklausų kalba (SQL) yra standartinė kalba, naudojama bendrauti su duomenų baze ir ypač naudinga tvarkant struktūrizuotus duomenis. Naudojama duomenų paieškai, pridėjimui, atnaujinimui ir pašalinimui, be kita ko, SQL palengvina struktūrizuotų duomenų tvarkymą. Pagalvokite apie viešbučio duomenų bazę, kurioje galima ieškoti svečių pagal vardą, telefono numerį, kambario numerį ir pan. Arba brūkšninius kodus, naudojamus produktams tvarkyti ir klasifikuoti gamybos, platinimo ir pardavimo vietose.

Struktūrizuoti duomenys paprastai saugomi reliacinėse duomenų bazėse (RDBMS). Duomenų bazėse esančią informaciją gali įvesti žmonės arba mašinos, o paiešką galima lengvai atlikti pagal rankiniu būdu įvestas užklausas arba algoritmus. Struktūrizuotiems duomenims saugoti ir tvarkyti taip pat naudojamos programos, pavyzdžiui, "Excel", kurias galima lengvai sujungti su kitomis analitinėmis priemonėmis tolesnei analizei atlikti. Struktūrizuoti duomenys puikiai tinka pagrindiniam organizavimui ir kiekybiniais skaičiavimams, tačiau turi atitikti griežtus, iš anksto nustatytus parametrus. Struktūrizuotų duomenų pavyzdžiai – tai duomenys, kurių lengva ieškoti pagal nustatytą struktūrą ir kuriuos galima susieti su kitomis duomenų bazėmis. Galite atlikti paiešką pagal klientų adresą, kad sužinotumėte, kurie produktai yra populiariausi tam tikroje vietovėje, arba išsiaiškinti, kuriuos produktus daug kartų užsako keli klientai (<https://monkeylearn.com/blog/structured-data-vs-unstructured-data/>).

Pusiau struktūrizuoti duomenys

- Pusiau struktūrizuotiems duomenims priskiriami duomenys turi tam tikrų apibrėžiančių ar nuoseklių savybių, tačiau neatitinka tokios griežtos struktūros, kokios tikimasi iš reliacinės duomenų bazės. Todėl yra tam tikrų organizacinių savybių, pavyzdžiui, semantinių žymų arba metaduomenų, kad būtų lengviau juos organizuoti, tačiau duomenys vis tiek yra kintantys.

Geras pavyzdys yra el. pašto pranešimai. Nors tikrasis turinys yra nestruktūrizuotas, jame yra struktūrizuotų duomenų, pavyzdžiui, siuntėjo ir gavėjo vardas ir el. pašto adresas, išsiuntimo laikas ir pan. Kitas pavyzdys - skaitmeninė nuotrauka. Pats vaizdas yra nestruktūrizuotas, bet jei nuotrauka buvo padaryta, pavyzdžiui, išmaniuoju telefonu, joje būtų nurodyta data ir laikas, geografinė žyma ir įrenginio ID. Išsaugojus nuotrauką, jai taip pat galėtų būti suteiktos žymos, kurios suteiktų struktūrą, pavyzdžiui, „šuo“ arba „augintinis“.

Daugelis duomenų, kuriuos žmonės paprastai priskiria nestruktūrizuotiems duomenims, iš tiesų yra pusiau struktūrizuoti, nes juose yra tam tikrų klasifikavimo požymių (<https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/?sh=6727d9952b4d>).

- Pusiau struktūrizuoti duomenys (pvz., JSON, CSV, XML) yra "tiltas" tarp struktūrizuotų ir nestruktūrizuotų duomenų. Jie neturi iš anksto nustatyto duomenų modelio ir yra sudėtingesni už struktūrizuotus duomenis, tačiau juos lengviau saugoti nei nestruktūrizuotus duomenis.

Pusiau struktūrizuoti duomenys naudoja metaduomenis (pvz., žymas ir semantines žymes) konkrečioms duomenų savybėms nustatyti ir duomenims suskirstyti į įrašus ir iš anksto nustatytus laukus. Metaduomenys galiausiai leidžia pusiau struktūrizuotus duomenis geriau kataloguoti, ieškoti ir analizuoti nei nestruktūrizuotus duomenis.

Metaduomenų naudojimo pavyzdys: internetiniame straipsnyje rodoma antraštė, ištrauka, rodomas paveikslėlis, paveikslėlio alt-tekstas, slug ir t. t., kurie padeda atskirti vieną žiniatinklio turinio dalį nuo panašių dalių.

Pusiau struktūrizuotų duomenų ir struktūrizuotų duomenų pavyzdys: Klientų duomenų failas su lentelėmis ir duomenų bazė su CRM lentelėmis.

Pusiau struktūrizuotų duomenų ir nestruktūrizuotų duomenų pavyzdys: Kliento Instagram komentarų sąrašas (<https://www.ibm.com/cloud/blog/structured-vs-unstructured-data>).

- Pusiau struktūrizuoti duomenys, kuriuos sudaro daugiausia nestruktūrizuotas tekstas, tačiau juos galima laisvai suskirstyti pagal metažymes. Pavyzdys galėtų būti el. paštas, kuriame galima ieškoti pagal kategorijas Gautieji, Išsiųstieji, Juodraščiai ir t. t. arba socialinė žiniasklaida, kurią galima suskirstyti į kategorijas Draugai, Žinutės, Viešos žinutės, Privačios žinutės ir t. t. Pusiau

struktūrizuotus duomenis galima lengvai suskirstyti į iš anksto nustatytas kategorijas, tačiau informacija šiose kategorijose pati savaime yra nestruktūrizuota. Analizuojant el. laiškus, ketinimų klasifikavimas gali būti naudingas automatiškai skaitant verslo el. laiškus pagal kliento ketinimus, kad būtų galima pasakyti, ar jis į užklausą atsako tikrai susidomėjęs, ar ne (<https://monkeylearn.com/blog/structured-data-vs-unstructured-data/>).

Nestruktūrizuoti duomenys

- Nestruktūrizuoti duomenys – tai duomenys, kurių negalima sutalpinti į eilutės ir stulpelio duomenų bazę ir kurie neturi susijusio duomenų modelio. Pagalvokite apie el. pašto žinutės tekstą. Dėl struktūros nebuvimo nestruktūrizuotus duomenis sunkiau ieškoti, valdyti ir analizuoti, todėl įmonės plačiai atmetė nestruktūrizuotus duomenis, kol pastaruoju metu paplitus dirbtiniam intelektui ir mašininio mokymosi algoritmams juos apdoroti tapo lengviau.

Nestruktūrizuotų duomenų pavyzdžiai: nuotraukos, vaizdo ir garso failai, tekstiniai failai, socialinės žiniasklaidos turinys, palydoviniai vaizdai, prezentacijos, PDF, atviri apklausos atsakymai, interneto svetainės ir skambučių centrų stenogramos / įrašai. (<https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/?sh=6727d9952b4d>).

- Nestruktūrizuotų duomenų, paprastai priskiriamų prie kokybinių duomenų, negalima apdoroti ir analizuoti įprastinėmis duomenų analizės priemonėmis ir metodais. Kadangi nestruktūrizuoti duomenys neturi iš anksto nustatyto duomenų modelio, juos geriausia tvarkyti nereliacinėse (NoSQL) duomenų bazėse. Kitas būdas tvarkyti nestruktūrizuotus duomenis – naudoti duomenų ežerus, kad jie būtų išsaugoti neapdoroti.

Nestruktūrizuotų duomenų svarba sparčiai didėja. Naujausios prognozės rodo, kad nestruktūrizuoti duomenys sudaro daugiau kaip 80 % visų įmonės duomenų, o 95 % įmonių teikia pirmenybę nestruktūrizuotų duomenų valdymui (<https://www.ibm.com/cloud/blog/structured-vs-unstructured-data>).

- Nestruktūrizuoti duomenys – tai informacija, kuri neturi nustatytos struktūros ir netelpa į apibrėžtą sistemą. Nestruktūrizuotų duomenų pavyzdžiai yra garso ir vaizdo įrašai, vaizdai ir įvairūs tekstai: ataskaitos, el. laišakai, įrašai socialiniuose tinkluose ir kt. Rasti išvalgų nestruktūrizuotuose duomenyse nėra lengva, tačiau tinkamai išanalizuoti tekstiniai duomenys gali būti labai vertingi siekiant išgauti kokybinius rezultatus, pavyzdžiui, klientų nuomones, arba organizuoti verslo duomenis, pavyzdžiui, klientų aptarnavimo bilietus, pagal atskiras kategorijas, kad juos būtų galima nukreipti tinkamam darbuotojui (<https://monkeylearn.com/blog/structured-data-vs-unstructured-data/>).

Struktūrizuotų, pusiau struktūrizuotų ir nestruktūrizuotų duomenų pavyzdžiai, savybės ir jų palyginimas

1. Darbo pokalbis (<https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/?sh=6727d9952b4d>)

Kad būtų lengviau suprasti duomenų klasifikavimo skirtumus, pasitelkime šią analogiją. Sakykime, kad vykstant pokalbiui dėl darbo yra trys skirtingos pokalbių klasifikacijos: struktūrizuotas, pusiau struktūrizuotas ir nestruktūrizuotas.

Struktūrizuoto pokalbio metu pokalbio vedėjas vadovaujasi griežtu scenarijumi, kurį nustatė žmoniškųjų išteklių skyrius ir kurio laikomasi su kiekvienu kandidatu. Kita pokalbio forma - nestruktūrizuotas pokalbis. Per nestruktūrizuotą pokalbį interviu vedėjas pats sprendžia, kokius klausimus ir kokia tvarka jie bus užduodami (ar apskritai bus užduodami) kiekvienam kandidatui. Iš dalies struktūrizuotas interviu turi ir struktūrizuoto, ir nestruktūrizuoto interviu elementų. Jame

naudojamas nuoseklumas ir kiekybiniai elementai, kuriuos leidžiama naudoti struktūrizuotame pokalbyje, tačiau suteikiama laisvė pritaikyti pagal aplinkybes, kurios labiau atitinka nestruktūrizuotą pokalbį.

Taigi, kalbant apie duomenis, struktūrizuotus duomenis lengva organizuoti ir jie atitinka griežtą formatą; nestruktūrizuoti duomenys yra sudėtinga ir dažnai kokybinė informacija, kurios neįmanoma redukuoti į reliacinę duomenų bazę ar organizuoti joje, o pusiau struktūrizuoti duomenys turi abiejų elementų.

2. Įvairių tipų duomenų pavyzdžiai (<https://microsoft.github.io/Data-Science-For-Beginners/#/1-Introduction/01-defining-data-science/README>)

Struktūrizuoti	Pusiau struktūrizuoti	Nestruktūrizuoti
Žmonių sąrašas su jų telefono numeriais.	Vikipedijos puslapiai su nuorodomis	Encyclopedia Britannica tekstas
Pastarųjų 20 metų visų pastato patalpų temperatūra kiekvieną minutę.	Mokslinių straipsnių rinkinys JSON formatu su autoriais, publikacijos duomenimis ir santrauka	Failų bendrinimas su įmonių dokumentais
Visų į pastatą įeinančių žmonių amžiaus ir lyties duomenys	Interneto puslapiai	Neapdorotas vaizdo įrašas iš stebėjimo kameros

3. Struktūrizuotų duomenų pavyzdžiai (<https://www.coursera.org/articles/structured-vs-unstructured-data>)

- **Skrydžio užsakymas.** Skrydžio ir užsakymo duomenys, pvz., datos, kainos ir paskirties vietos, tvarkingai saugomi "Excel" skaičiuoklės formatu. Rezervuojant skrydį ši informacija saugoma duomenų bazėje.
- **Ryšių su klientais valdymas** (CRM, Customer relationship management). CRM programinė įranga, pavyzdžiui, "Salesforce", struktūrizuotus duomenis apdoroja analitinėmis priemonėmis, kad sukurtų naujus duomenų rinkinius, skirtus įmonėms analizuoti klientų elgseną ir pageidavimus.

4. Nestruktūrizuotų duomenų pavyzdžiai (<https://www.coursera.org/articles/structured-vs-unstructured-data>)

- **Pokalbių robotai.** Pokalbių robotai suprogramuoti atlikti teksto analizę, kad atsakytų į klientų klausimus ir suteiktų reikiamą informaciją.
- **Rinkos prognozės.** Duomenys gali būti panaudoti siekiant numatyti akcijų rinkos pokyčius, kad analitikai galėtų pakoreguoti savo skaičiavimus ir investicinius sprendimus.

Struktūrizuotų ir nestruktūrizuotų duomenų privalumai ir trūkumai

<https://www.ibm.com/cloud/blog/structured-vs-unstructured-data>

Duomenų tipas	Privalumai	Trūkumai
Struktūrizuoti	• Lengvai pritaikomi mašininio mokymosi algoritmai: Specifinė ir organizuota struktūrizuotų duomenų architektūra palengvina manipuliavimą ML duomenimis ir užklausų teikimą. • Lengvai pritaikomi versle. Struktūrizuoti duomenys nereikalauja gilaus supratimo apie skirtingus duomenų tipus ir jų	• Ribotas naudojimas. Duomenys, kurių struktūra iš anksto nustatyta, gali būti naudojami tik pagal numatytą paskirtį, o tai riboja jų lankstumą ir panaudojimo galimybes. • Ribotos saugojimo galimybės: Duomenys paprastai saugomi duomenų saugojimo sistemose, turinčiose griežtas schemas (pvz.,

	veikimą. Naudotojai, turėdami pagrindinį supratimą apie su duomenimis susijusią temą, gali lengvai pasiekti ir interpretuoti duomenis. • Galimybė naudoti daugiau įrankių duomenų apdorojimui. Kadangi struktūrizuoti duomenys pradėti analizuoti anksčiau už nestruktūrizuotus duomenis, yra daugiau įrankių, skirtų naudoti ir analizuoti struktūrizuotus duomenis.	duomenų saugyklose). Todėl, pasikeitus duomenų reikalavimams, būtina atnaujinti visus struktūrizuotus duomenis, o tai reikalauja didžiulių laiko ir išteklių sąnaudų.
Nestruktūrizuoti	• Pirminis formatas. Nestruktūrizuotus duomenis, saugomus pirminiu formatu, kol jų prireikia, galima laikyti neapibrėžtus. Dėl jo pritaikomumo duomenų bazėje padaugėja failų formatų, o tai praplečia duomenų fondą ir leidžia duomenų mokslininkams rengti ir analizuoti tik tuos duomenis, kurių jiems reikia. • Greita kaupimo sparta. Kadangi duomenų nereikia iš anksto apibrėžti, juos galima kaupti greitai ir lengvai. • Duomenų ežero saugykla*. Leidžia naudoti didžiulę saugyklą ir mokėti už naudojimąsi, o tai mažina išlaidas ir leidžia lanksčiai keisti duomenų dydį.	• Reikalinga patirtis. Dėl neapibrėžto/nesuformuoto duomenų pobūdžio, norint parengti ir analizuoti nestruktūrizuotus duomenis, reikia duomenų mokslo žinių. Tai naudinga duomenų analitikams, tačiau sudaro sunkumų naudotojams, kurie gali ne iki galo suprasti specializuotas duomenų temas arba tai, kaip panaudoti savo duomenis. • Specializuotos priemonės. Duomenų tvarkytojams reikia specializuotų įrankių nestruktūrizuotiems duomenims tvarkyti, o tai apriboja produktų pasirinkimą.

*Duomenų ežero saugykla (Data lake storage): <https://www.valdas.blog/2019/03/31/duomenuezeras/#palyginkime-duomen%C5%B3-saugykl%C4%85-su-duomen%C5%B3-e%C5%BEeru>

Struktūrizuotų ir nestruktūrizuotų duomenų pagrindiniai skirtumai

<https://www.ibm.com/cloud/blog/structured-vs-unstructured-data>

Struktūrizuoti (kiekybiniai) duomenys leidžia pamatyti klientus „iš paukščio skrydžio“, o nestruktūrizuoti (kokybiniai) duomenys padeda geriau suprasti klientų elgseną ir ketinimus. Panagrinėkime kai kuriuos pagrindinius skirtumus ir jų reikšmę:

Šaltiniai: Struktūrizuoti duomenys gaunami iš GPS jutiklių, internetinių formų, tinklo žurnalų, žiniatinklio serverių žurnalų, OLTP sistemų ir t. t., o nestruktūrizuotų duomenų šaltiniai yra el. pašto pranešimai, tekstų apdorojimo dokumentai, PDF failai ir kt.

Formos: Struktūrizuotus duomenis sudaro skaičiai ir reikšmės, o nestruktūrizuotus duomenis – jutikliai, tekstiniai failai, garso ir vaizdo failai ir kt.

Modeliai: Struktūrizuoti duomenys turi iš anksto nustatytą duomenų modelį ir prieš patalpinant į duomenų saugyklą yra suformatuojami pagal nustatytą duomenų struktūrą (pvz., schema įrašant), o nestruktūrizuoti duomenys saugomi savo prigimtiniu formatu ir neapdorojami tol, kol nėra naudojami (pvz., schema skaitant).

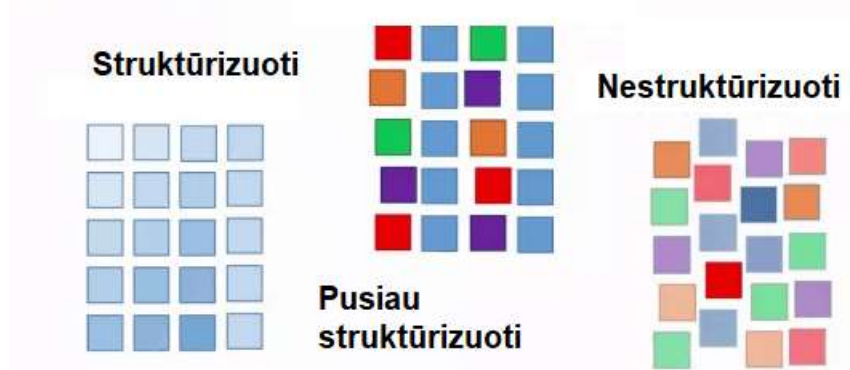
Saugojimas: Struktūrizuoti duomenys saugomi lentelių formatais (pvz., skaičiuoklių lentelėse arba SQL duomenų bazėse), kuriems saugoti reikia mažiau vietos. Juos galima saugoti duomenų saugyklose, todėl juos galima labai lengvai glaudinti. Kita vertus, nestruktūrizuoti duomenys saugomi kaip medijos failai arba NoSQL duomenų bazės, kurioms reikia daugiau vietos. Juos galima saugoti duomenų ežeruose, todėl juos sunku glaudinti.

Naudojimo būdai: Struktūrizuotus duomenis naudoja mašininis mokymasis (ML) ir jų algoritmai, o nestruktūrizuotus duomenis - natūralios kalbos apdorojimas (NLP) ir teksto gavyba.

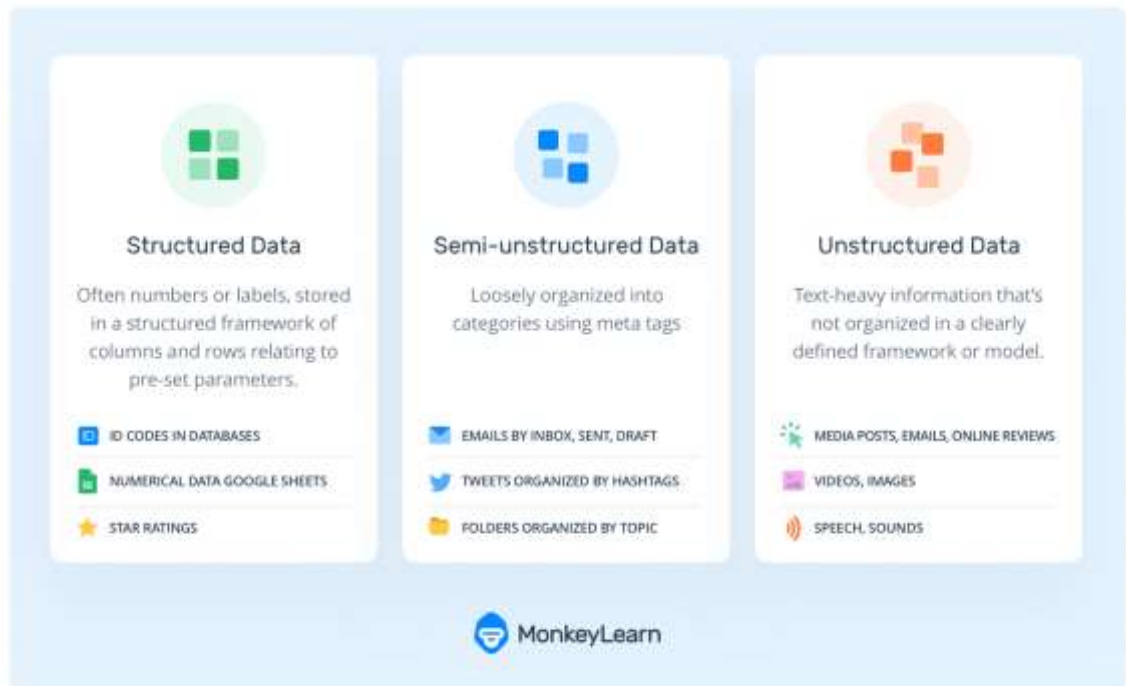
Iš šaltinio: (<https://monkeylearn.com/blog/structured-data-vs-unstructured-data/>)

Struktūrizuoti duomenys	Nestruktūrizuoti duomenys
Struktūrizuoti duomenys yra kiekybiniai ir dažnai pateikiami kaip skaičiai, datos, reikšmės ir eilutės.	Nestruktūrizuoti duomenys yra kokybiniai duomenys, apimantys tekstą, vaizdo įrašus, garso įrašus, vaizdus ir kt.
Struktūrizuoti duomenys saugomi eilutėmis ir stulpeliais.	Nestruktūrizuoti duomenys saugomi kaip garso, teksto ir vaizdo failai arba NoSQL duomenų bazės.
Apytikriai 20 proc. verslo duomenų.	Apytikriai 80-90 % verslo duomenų (ir jų vis daugėja!).
Randami uždaro tipo apklausose, pavyzdžiui, NPS ir CSAT rezultatuose, ir žiniatinklio analizėje.	Klientų aptarnavimo užklausose, atsiliepimuose internete, el. pašto žinutėse, atvirose klausimuose ir kt.
Saugomi duomenų saugyklose.	Saugomi taikomuosiose programose, NoSQL (nereliacinėse) duomenų bazėse, duomenų ežeruose ir duomenų saugyklose.
Atskleidžia modelius ir tendencijas, kurie parodo, kas vyksta.	Atskleidžia modelius ir tendencijas, kurie paaiškina, kodėl kažkas vyksta.
Reikalauja mažiau vietos saugykloje.	Reikia daugiau vietos saugykloje.
Lengva analizuoti naudojant tokias priemones kaip skaičiuoklę.	Sunku analizuoti be dirbtinio intelekto įrankių.

Įvairių tipų duomenų vaizdavimas



Šaltinis: <https://www.astera.com/type/blog/structured-semi-structured-and-unstructured-data/>



Šaltinis: <https://monkeylearn.com/blog/structured-data-vs-unstructured-data/>

Duomenų šaltiniai

<https://microsoft.github.io/Data-Science-For-Beginners/#/1-Introduction/01-defining-data-science/README>

1. Struktūrizuotų duomenų šaltiniai

- Daiktų internetas (IoT), įskaitant įvairių jutiklių, pavyzdžiui, temperatūros ar slėgio ir kt., duomenis, suteikia daug naudingų duomenų. Pavyzdžiui, jei namuose ar mokykloje įrengti daiktų interneto jutikliai, galime automatiškai valdyti šildymą ir apšvietimą, kad sumažintume išlaidas.
- Apklausos, kurias prašome naudotojų užpildyti po pirkimo arba apsilankius interneto svetainėje.
- Elgesio analizė gali, pavyzdžiui, padėti suprasti, kaip giliai naudotojas įeina į svetainę ir kokia yra tipinė svetainės palikimo priežastis.

2. Nestrukūrizuotų duomenų šaltiniai

- Tekstai gali būti turtingas išvalgų šaltinis, pavyzdžiui, bendras nuotaikų įvertinimas arba raktinių žodžių ir semantinės reikšmės išskyrimas.
- Vaizdai arba vaizdo įrašai. Stebėjimo kameros vaizdo įrašas gali būti naudojamas siekiant įvertinti eismo intensyvumą kelyje ir informuoti žmones apie galimas spūstis.
- Interneto serverio žurnalai gali būti naudojami siekiant suprasti, kurie mūsų svetainės puslapiai lankomi dažniausiai ir kiek laiko.

3. Pusiau struktūrizuotų duomenų šaltiniai

- Socialinių tinklų grafikai gali būti puikus duomenų apie naudotojų asmenybes ir galimą informacijos sklaidos veiksmingumą šaltinis.
- Kai turime krūvą nuotraukų iš vakarėlio, galime pabandyti išgauti grupės dinamikos duomenis sudarydami žmonių, fotografuojančių vienas kitą, grafiką.
- Žinodami įvairius galimus duomenų šaltinius, galite pabandyti pagalvoti apie įvairius scenarijus, kuriuose galima taikyti duomenų tyrybos metodus, kad geriau pažintumėte situaciją ir patobulintumėte verslo procesus.

1. Duomenų rinkimas

Pirmasis žingsnis – surinkti duomenis. Nors daugeliu atvejų tai gali būti nesudėtingas procesas, pavyzdžiui, duomenys, gaunami į duomenų bazę iš žiniatinklio programos, kartais turime naudoti specialius metodus. Pavyzdžiui, iš daiktų interneto jutiklių gaunamų duomenų gali būti per daug, todėl prieš tolesnį apdorojimą pravartu naudoti buferizavimo galinius taškus, pavyzdžiui, IoT Hub, kad būtų galima surinkti visus duomenis.

2. Duomenų saugojimas

Saugoti duomenis gali būti sudėtinga, ypač jei kalbame apie didelius duomenis. Nusprendžiant, kaip saugoti duomenis, tikslinga numatyti, koku būdu ateityje norėsite pateikti užklausą apie duomenis. Duomenis galima saugoti keliais būdais:

- Reliacinėje duomenų bazėje saugomas lentelių rinkinys, o užklausoms atlikti naudojama speciali SQL kalba. Paprastai lentelės yra suskirstytos į skirtingas grupes, vadinamas schemomis. Daugeliu atvejų duomenis reikia konvertuoti iš pradinės formos, kad jie atitiktų schemą.
- NoSQL duomenų bazės, pavyzdžiui, CosmosDB, duomenims schemų neįtvirtina ir leidžia saugoti sudėtingesnius duomenis, pavyzdžiui, hierarchinius JSON dokumentus ar grafikus. Tačiau NoSQL duomenų bazės neturi plačių SQL užklausų galimybių ir negali užtikrinti referencinio vientisumo, t. y. duomenų struktūros lentelėse ir ryšių tarp lentelių taisyklių.
- Duomenų ežero saugykla naudojama didelėms neapdorotų, nestruktūrizuotų duomenų kolekcijoms saugoti. Duomenų ežerai dažnai naudojami dideliems duomenims, kai visi duomenys netelpa viename kompiuteryje, todėl juos reikia saugoti ir apdoroti serverių klasteryje. Parquet yra duomenų formatas, kuris dažnai naudojamas kartu su dideliais duomenimis.

3. Duomenų apdorojimas

Tai žingsnis, apimantis duomenų konvertavimą iš pradinės formos į formą, kurią galima naudoti vizualizavimui ir (arba) modelių mokymui. Dirbant su nestruktūrizuotais duomenimis, pavyzdžiui, tekstu ar vaizdais, gali tekti naudoti tam tikrus dirbtinio intelekto metodus, kad iš duomenų būtų galima išgauti požymius ir taip juos paversti struktūrizuota forma.

4. Vizualizavimas / žmogaus įžvalgos

Dažnai, norėdami suprasti duomenis, turime juos vizualizuoti. Turėdami daug įvairių vizualizavimo metodų savo įrankių rinkinyje galime rasti tinkamą atvaizdavimą, kad padarytume įžvalgą. Dažnai duomenų mokslininkui reikia „žaisti su duomenimis“, daug kartų juos vizualizuojant ir ieškant tam tikrų sąsajų. Taip pat galime naudoti statistinius metodus, kad patikrintume hipotezes arba įrodytume ryšį tarp skirtingų duomenų.

5. Prognozavimo modelio mokymas

Kadangi galutinis duomenų mokslo tikslas – gebėti priimti sprendimus remiantis duomenimis, galime norėti pasinaudoti mašininio mokymosi metodais, kad sukurtume prognozavimo modelį. Paskui jį galėsime naudoti prognozėms atlikti naudodami naujus panašios struktūros duomenų rinkinius.

Žinoma, priklausomai nuo faktinių duomenų, kai kurių žingsnių gali nebūti (pavyzdžiui, kai duomenis jau turime duomenų bazėje arba kai mums nereikia modelio mokymo) arba kai kurie žingsniai gali būti kartojami kelis kartus (pavyzdžiui, duomenų apdorojimas).

Skaitmenizacija ir skaitmeninė transformacija

(<https://microsoft.github.io/Data-Science-For-Beginners/#/1-Introduction/01-defining-data-science/README>)

Pastarąjį dešimtmetį daugelis įmonių pradėjo suprasti duomenų svarbą priimant verslo sprendimus. Norint taikyti duomenų tyrybos principus valdant verslą, pirmiausia reikia surinkti tam tikrus duomenis, t. y. verslo procesus perkelti į skaitmeninę formą. Tai vadinama skaitmeninimu. Duomenų tyrybos metodų taikymas šiems duomenims, kuriais remiantis priimami sprendimai, gali lemti reikšmingą produktyvumo padidėjimą (ar net verslo posūkį), vadinamą skaitmenine transformacija.

Panagrinėkime pavyzdį. Tarkime, turime duomenų tyrybos kursą (kaip šis) ir norime jį patobulinti pasitelkdami duomenų tyrybą. Kaip tai galime padaryti?

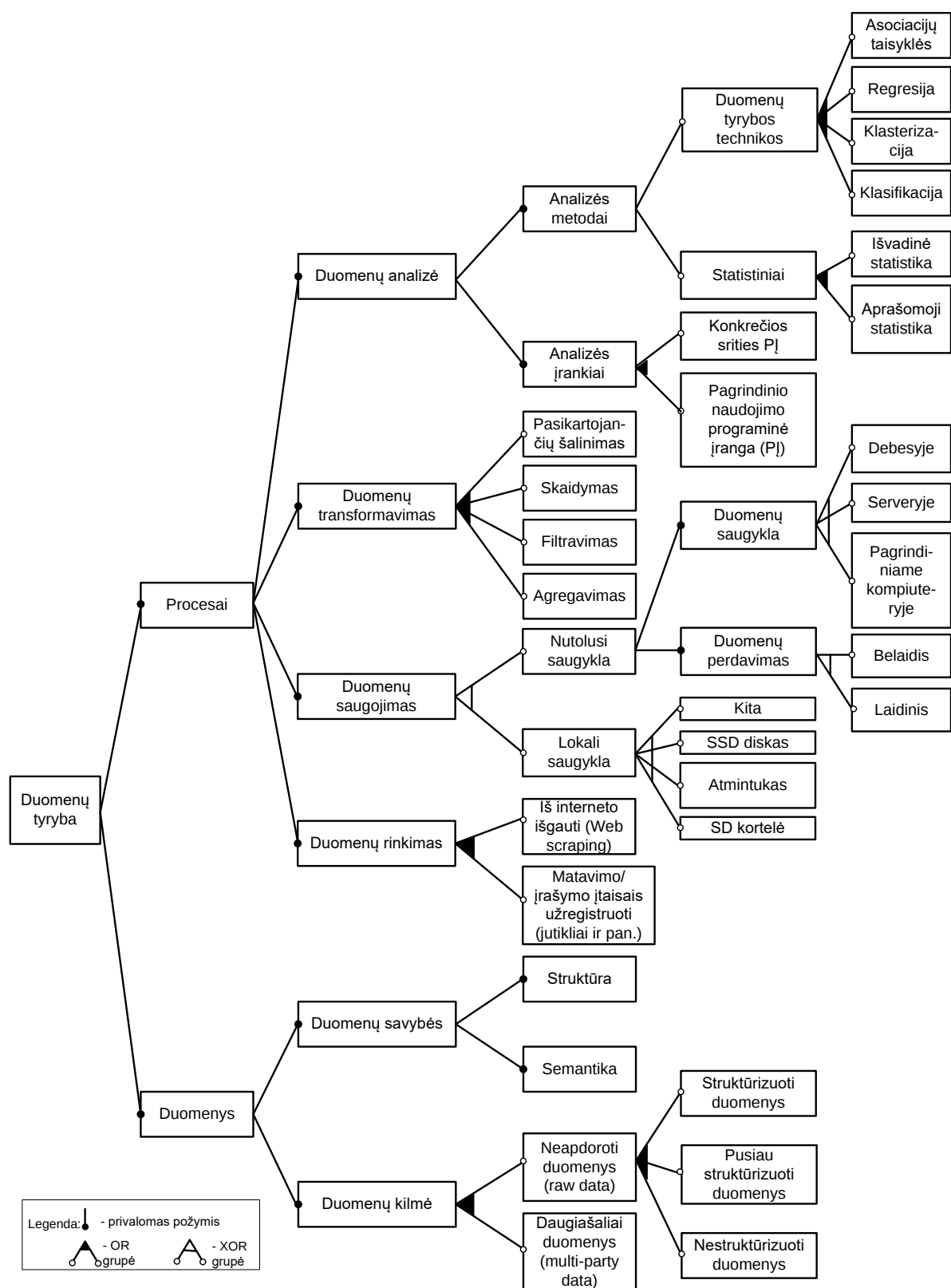
Galime pradėti nuo klausimo „Ką galima skaitmenizuoti?“. Paprasčiausias būdas būtų matuoti laiką, per kurį kiekvienas mokinys įveikia kiekvieną pamoką, ir įvertinti įgytas žinias pateikiant testą su keliais atsakymų variantais kiekvienos pamokos pabaigoje. Vidutiniškai įvertinę visų mokinių užimamą laiką, galime išsiaiškinti, kurios pamokos mokiniams kelia daugiausia sunkumų, ir jas supaprastinti.

Pradėję analizuoti testų su keliais atsakymų variantais rezultatus, galime pabandyti nustatyti, kurias sąvokas mokiniai sunkiai supranta, ir panaudoti šią informaciją turiniui tobulinti. Kad tai padarytume, turime sudaryti testus taip, kad kiekvienas klausimas atitiktų tam tikrą sąvoką ar žinių dalį.

Skyrelio 1.3 medžiagai įtvirtinti siūlomos užduotys

Analogiškai kaip ir atliekant skyrelio 1.2 užduotis, mokiniams rekomenduojama kurti minčių žemėlapių, arba žodžių debesį iš 1.3 skyrelio duomenų. Jei atlikdami 1.2 skyrelio užduotis kūrė minčių žemėlapius, tai atlikdami 1.3 skyrelio užduotį, kuria žodžių debesį ir atvirkščiai.

1.4. Duomenų tyrybos modelis



Šaltinis: Stuišys, V., Burbaite, R., Ziberkas, G., & Kubiliūnas, R. (2022). Development and Evaluation of an Approach for Integrating Data Science Concepts into High School STEM Curriculum. *INTERNATIONAL JOURNAL OF ENGINEERING EDUCATION*, 38(3), 756-773.

Testas išnagrinėjus temą

1. Kurios sritys glaudžiai susijusios su duomenų tyryba?

- a. Mašininis mokymasis.
- b. Dirbtinis intelektas.
- c. a ir b variantai teisingi.

2. Kuris iš pavyzdžių yra nestruktūrizuoti duomenys?

- a. Klasės mokinių sąrašas.
- b. Mokinių esė kolekcija.
- c. Socialinio tinklo vartotojų draugų grafas.

3. Koks yra pagrindinis duomenų tyrybos tikslas?

- a. Surinkti duomenis.
- b. Apdoroti duomenis.
- c. Gebėti sukurti duomenimis pagrįstus sprendimus.

Vertinimas

Siūloma taikyti kaupiamąjį vertinimą, įvertinant kiekvieno skyrelio užduotis: testas prieš pradedant temą – 5 proc. bendro įvertinimo, 1.1 skyrelio testas – 15 proc., 1.2 ir 1.3 skyrelių – po 35 proc., testas temos pabaigoje – 10 proc.

1 skyriaus pateiktis

<https://docs.google.com/presentation/d/1lmQZgEtQ09Qfy1mV-6h9ZPY-1KSwONzP/edit?usp=sharing&ouid=100641493981479752593&rtpof=true&sd=true>

2. Apie duomenis išsamiau

2.1. Didieji duomenys (Big Data)

Svarbiausios didžiųjų duomenų charakteristikos

Šaltinis	Svarbiausios charakteristikos
Wang, J., Xu, C., Zhang, J., & Zhong, R. (2022). Big data analytics for intelligent manufacturing systems: A review. <i>Journal of Manufacturing Systems</i> , 62, 738-752.	Sparčiai vystantis šiuolaikinėms technologijoms daugiausia dėmesio skiriama milžiniškų duomenų kiekių, vadinamų didžiais duomenimis, rinkimui. Tačiau susiduriama su dideliais iššūkiais apdorojant didžiuosius duomenis ir išgaunant iš jų informaciją. Didieji duomenys apibūdinami "3V" charakteristikomis – apimtys (Volume), įvairovės (Variety) ir greičio (Velocity).
Baig, M. I., Shuib, L., & Yadegaridehkordi, E. (2019). Big data adoption: State of the art and research challenges. <i>Information Processing & Management</i> , 56(6), 102095.	Atsiradus kompiuteriams, internetui ir mobiliosioms technologijoms duomenų kiekiai sparčiai auga. Duomenys generuojami per socialinę žiniasklaidą, apsipirkimą internetu, sandorius, tinklo įrenginius ir švietimo įrašus. Anksčiau duomenys buvo saugomi duomenų bazėse ir skaičiuoklėse, o informacija, kurią sunku konvertuoti į eilutes ir stulpelius, buvo ignoruojama. Dideli duomenys yra įvairių tipų: struktūrizuoti (duomenų bazės, sukurtos naudojant SQL serverį ir Oracle ir kt.), nestruktūrizuoti (vaizdo įrašai, garso įrašai, įvairūs dokumentai, vaizdai, komentarai, stebėtojai, patinkantys, žymos, tviteriai, paspaudimai ir pokalbiai ir kt.) ir pusiau struktūrizuoti (trečiųjų šalių duomenys, valiutos konvertavimas, orai, XML, grafiko ar teksto duomenys, elektroninė prekyba ir kt.). Didelių

	duomenų paslaugos suteikia galimybes apdoroti, analizuoti įvairių tipų duomenis ir pateikti prognozes per trumpą laiką.
Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. <i>Journal of King Saud University-Computer and Information Sciences</i>, 30(4), 431-448.	Skirtingai nuo tradicinių duomenų, terminas „didieji duomenys“ apibūdina didelius heterogeninius (nevienalyčius) duomenų rinkinius (nuolat didėjančius), apimančius struktūrizuotus, nestruktūrizuotus ir pusiau struktūrizuotus duomenis. Didieji duomenys yra sudėtingi, todėl jiems apdoroti reikalingos galingos technologijos ir pažangūs algoritmai. Dauguma duomenų mokslininkų ir ekspertų didžiuosius duomenis apibrėžia trimis pagrindinėmis charakteristikomis (vadinamos 3V): Apimtis (Volume): nuolat generuojami dideli skaitmeninių duomenų kiekiai iš milijonų įrenginių ir programų (IRT, išmaniųjų telefonų, produktų kodai, socialiniai tinklai, jutikliai, žurnalai ir kt.). Greitis (Velocity): duomenys generuojami greitai ir turi būti apdorojami greitai norint gauti naudingos ir svarbios informacijos. Įvairovė (Variety): dideli duomenys generuojami iš paskirstytų įvairių šaltinių ir įvairiais formatais (pvz., vaizdo įrašais, dokumentais, komentarais, jutiklių duomenys). Dideli duomenų rinkiniai gali būti vieši ar privatus, vietiniai ar nuotoliniai, bendrinami arba konfidencialūs, išsamūs ar neišsamūs ir pan.
Emani, C. K., Cullot, N., & Nicolle, C. (2015). Understandable big data: a survey. <i>Computer science review</i>, 17, 70-81.	Didieji duomenys – tai duomenų rinkiniai, kurių dydis viršija tipinių duomenų bazių programinės įrangos įrankių galimybes duomenis fiksuoti, saugoti, tvarkyti ir analizuoti. Taip pat akcentuojamas duomenų daugiamatiškumas (multi-dimensionality), greitas judėjimas (moves to fast). Didžiųjų duomenų charakteristikos papildytos dar 3V: Vertė (Value) - duomenų architektūros yra sukurtos ekonomiškai išgauti vertę iš labai didelių įvairių duomenų kiekių, įgalinant didelio greičio fiksavimą, atradimą ir (arba) analizę“. Vertė skirstoma į dvi kategorijas: analitiniam naudojimui (pakeitimui / palaikymui žmogaus sprendimas, poreikių atradimas, populiacijų segmentavimas pritaikyti veiksmus) ir naujų verslo modelių, produktų ir paslaugų įgalinimui. Duomenų teisingumas, tikslumas, tikrumas (Veracity) – Netikslumus gali sukelti nepilni duomenys, modelio tikslumas, neaiškumai duomenyse, apgaulė, sukčiavimas, dubliavimas, neužbaigtumas, šlamštas ir vėlavimas. Dėl teisingumo rezultatai gauti iš didelių duomenų, negali būti įrodyti. Jie yra tikimybiniai (pateikiami su tam tikra tikimybe). Vizualizacija (Visualization) – reikalinga analizės rezultatų pateikimui. Apibendrinant galima teigti, kad norint veiksmingai dirbti su didžiais duomenimis reikia kurti vertę (Value), atsižvelgiant į duomenų apimtį (Volume), įvairovę

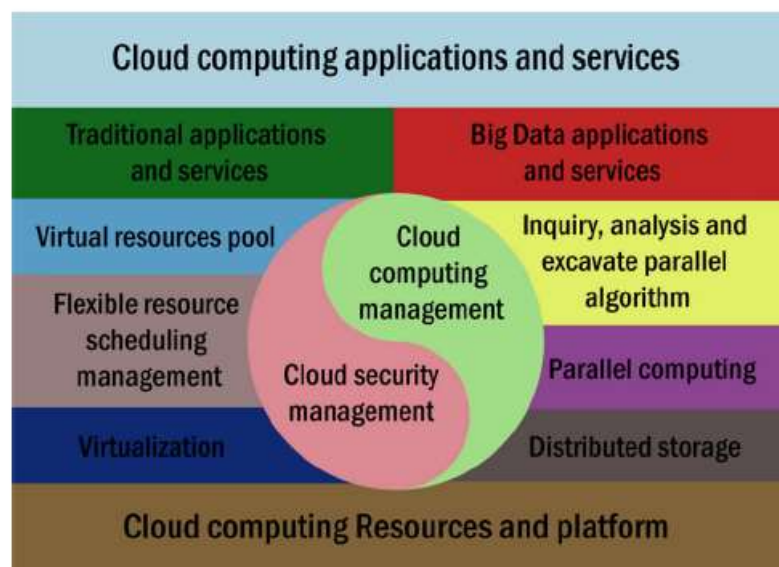
	(Variety) ir teisingumą (Veracity), kol jie vis dar juda (Velocity), o ne tik tada, kai jie jau yra (at the rest).
Ward, J. S., & Barker, A. (2013). Undefined by data: a survey of big data definitions. <i>arXiv preprint arXiv:1309.5821</i> .	Nepaisant to, kad egzistuoja įvairūs didžiųjų duomenų apibrėžimai, visuose galima išskirti šiuos pagrindinius veiksnius: Dydis: duomenų rinkinių apimtis. Sudėtingumas: duomenų struktūra, elgsena ir kombinacijos. Technologijos: priemonės ir metodai, kurie naudojami apdorojant didelius ar sudėtingus duomenų rinkinius. Suformuluotas apibrėžimas: didieji duomenys apima didelės apimties sudėtingų duomenų rinkinių saugojimą ir analizę panaudojant įvairias technologijas.
Yu, S., Liu, M., Dou, W., Liu, X., & Zhou, S. (2016). Networking for big data: A survey. <i>IEEE Communications Surveys & Tutorials</i> , 19(1), 531-549.	Tai duomenų rinkiniai, kurių dydžiai viršija įprastai naudojamų programinės įrangos priemonių galimybes, tvarkyti ir apdoroti duomenis per „leistiną laiko tarpą“.
Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. <i>National science review</i> , 1(2), 293-314.	Išskiriamos svarbiausios didžiųjų duomenų savybės: heterogeniškumas (nevienalytiškumas), triukšmų kaupimo efektas, klaidinga koreliacija, atsitiktinis endogeniškumas (kintamojo priklausomybė nuo kitų sistemos kintamųjų, https://lt1.uppercreditfieldnaturalists.org/what-is-endogeneity-what-is-an-exogenous-variable-2a66c4b). Šios savybės neleidžia panaudoti tradicinių statistinių duomenų analizės metodų.
Abdalla, H. B. (2022). A brief survey on big data: technologies, terminologies and data-intensive applications. <i>Journal of Big Data</i> , 9(1), 1-36.	Didieji duomenys išgaunami iš skirtingos kilmės ir įvairių konfigūracijų šaltinių (pvz., įrašai, ataskaitos, pastabos ir žurnalai). Didžiųjų duomenų rinkinius sudaro duomenys, kurie yra organizuoti ir nesutvarkyti, privatūs ir vieši, lokalūs ir nutolę, paskirstytieji ir slapti, pilni ir nepilni. Dauguma tyrėjų didžiuosius duomenis apibūdina penkiais atributais, vadinamais 5V: apimtis (Volume), greitis (Velocity), įvairovė (Variety), teisingumas (Veracity) ir vertė (Value).
Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. <i>Mobile networks and applications</i> , 19, 171-209.	Didieji duomenys apibūdina duomenų rinkinius, kurių negalima suvokti, įgyti, valdyti ir apdoroti naudojant tradicinius IT ir programinės/aparatinės įrangos įrankius per toleruotiną laiką. Atsižvelgiant į skirtingus aspektus, mokslo ir technologijų įmonės, mokslininkai, duomenų analitikai, technikos specialistai skirtingai apibrėžia didžiuosius duomenis. Išskiriamos pagrindinės didžiųjų duomenų savybės: 1) didelė apimtis ir įvairovė, sudėtinga struktūra; 2) duomenų išteklių našumas, kuris pasiekiamas pertvarkant ir integruojant iš skirtingų išteklių išgautus duomenis, kurie gali sukurti didesnę vertę; 3) didieji duomenys skatina kryžminį mokslo susiliejimą (cross fusion of science) apimantį ne tik debesų kompiuteriją, daiktų internetą, duomenų centrus ir mobiliuosius tinklus, bet ir duomenų išgavimo, saugumo ir analizės naujų technologijų ir metodų kūrimą; 4) vizualizacija, skirta

	efektyviam rezultatų panaudojimui; 5) orientuota į duomenis (data-oriented) programos architektūra pakeičia į algoritmus orientuotą architektūrą; 6) didžiųjų duomenų analizės rezultatai padės priimti tinkamesnius sprendimus.
https://www.europarl.europa.eu/news/lt/headlines/society/20210211STO97614/didieji-duomenys-apibrezimas-nauda-issukiai-infografikas#:~:text=Didieji%20duomenys%20%28angl.%20big%20data%29%20%E2%80%93%20tai%20%C5%BEmoni%C5%B3,sistem%C5%B3%2C%20pvz.%2C%20GPS%2C%20signalai%2C%20jutikli%C5%B3%20registruojama%20klimato%20informacija.	Didieji duomenys (angl. <i>big data</i>) – tai žmonių ar mašinų sukuriama dideli duomenų kiekiai, kaip antai pirkimo sandorių duomenys, naudojimosi socialiniais tinklais įpročiai, interesai ir pomėgiai, padėties nustatymo sistemų, pvz., GPS, signalai, jutiklių registruojama klimato informacija. Apdoroti ir išanalizuoti jie paverčiami išvalgomis, padedančiomis priimti sprendimus rinkodaros, reklamos, prekybos, miestų planavimo, sveikatos priežiūros bei transporto srityse. Pasinaudodamos šia informacija įmonės gali siūlyti prekes tik tiems, kam jų reikia, ir tik tada, kai jų reikia. Ji taip pat gali patarti, kokį maršrutą pasirinkti, kokią draudimo paslaugą pasiūlyti klientui, ar suteikti paskolą ir netgi gali padėti išsirinkti tinkamą darbuotoją.

Didžiųjų duomenų sąsajos su šiuolaikinėmis technologijomis

Šaltinis: Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, 19, 171-209.

Didieji duomenys yra glaudžiai susiję su **debesų kompiuterija** (Cloud Computing), nes jų saugojimui reikia didelių saugyklų, o apdorojimui ir analizei – daug skaičiavimų. Iš kitos pusės, didieji duomenys spartina debesų kompiuterijos vystymąsi, kuriant paskirstytąsias saugyklas ir lygiagrečiuosius skaičiavimus, skirtus didiesiems duomenims.

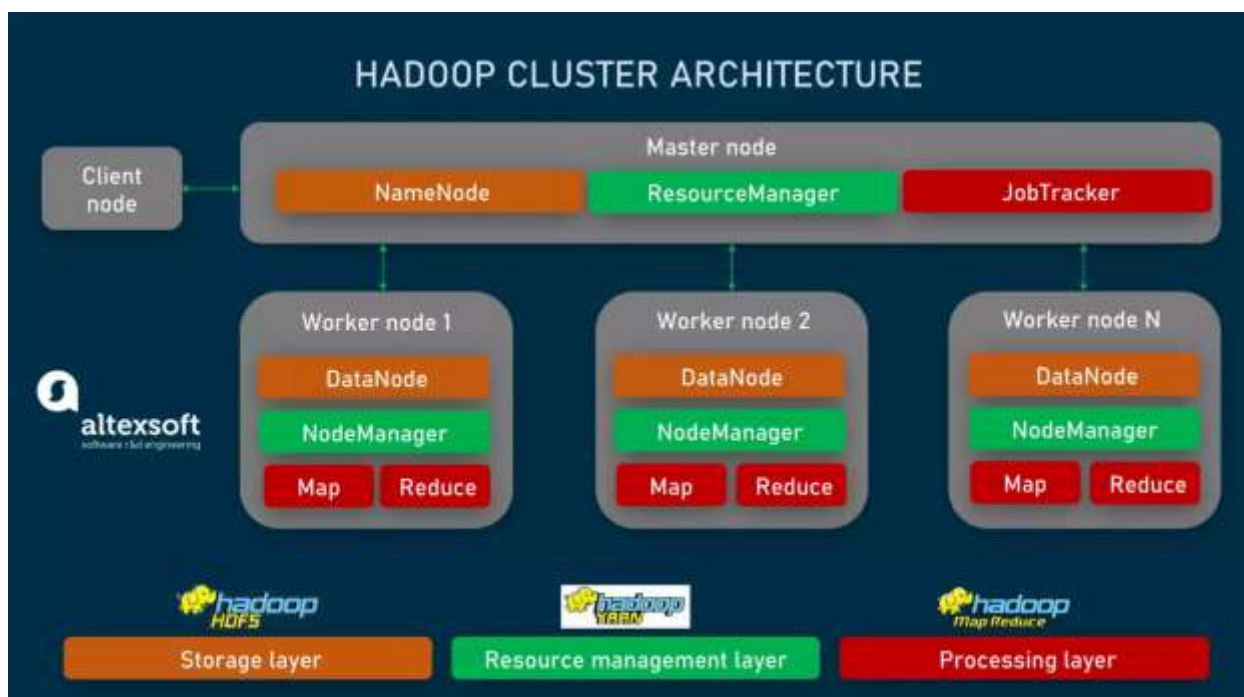


Daiktų internetas (Internet of Things) pasižymi tuo, kad jutikliai yra įmontuoti į įvairius įrenginius ir mašinas realiaame pasaulyje. Tokie jutikliai naudojami įvairiose srityse gali rinkti įvairių rūšių duomenis: aplinkos, geografinius, astronominius, logistikos ir pan. Didieji daiktų interneto generuojami duomenys palyginus su bendrais didžiaisiais duomenimis, skiriasi renkamu duomenų tipais, yra heterogeniniai, nestruktūrizuoti, triukšmingi, pasikartojantys (besidubliuojantys). Daiktų

interneto diegimas skatina spartų duomenų augimą pagal apimtį ir įvairovę. Didžiųjų duomenų technologijų perkėlimas į daiktų internetą spartina daiktų interneto pažangą ir verslo modelių kūrimą.

Didžiųjų duomenų paradigmoje **duomenų centras** (Data Center) yra ne tik platforma koncentruotam duomenų saugojimui, bet ir atlieka duomenų gavimo, tvarkymo, panaudojimo funkcijas. Dideliems duomenims reikia, kad duomenų centras užtikrintų saugojimo talpos, apdorojimo, tinklo perdavimo pajėgumus. Duomenų centras apima infrastruktūrą su daugybe mazgų, didelės spartos vidiniais tinklais, efektyviai išsklaido šilumą, daromos atsarginės duomenų kopijos. Didžiųjų duomenų taikymų plėtra spartina naujovių diegimą duomenų centruose. Toliau augant duomenų apimtims, struktūrizuoti ir nestruktūrizuoti duomenys bei jų įvairovė, analitinių duomenų šaltiniai, duomenų apdorojimas ir skaičiavimai tiesiogiai susiję su duomenų centro pajėgumų didinimu. Duomenų centrai apima ne tik techninę, bet ir programinę įrangą, skirtą didžiųjų duomenų organizavimui, analizei ir taikymui.

Šiuo metu „Hadoop“¹ plačiai naudojama didelių duomenų taikymuose: pvz., šiukšlių filtravimas, paieška tinkle, paspaudimų srauto analizė ir socialinės rekomendacijos. „Hadoop“ leidžia lengviau išnaudoti visą saugyklos ir apdorojimo talpą klasterių serveriuose ir vykdyti paskirstytus procesus, apimant didelius duomenų kiekius. „Hadoop“ ekosistema bėgant metams labai išaugo dėl savo išplėtimo. Šiandien Hadoop ekosistemoje yra daug įrankių ir programų, padedančių rinkti, saugoti, apdoroti, analizuoti ir valdyti didelius duomenis.



Šaltinis: <https://www.altexsoft.com/blog/hadoop-vs-spark/>

¹ „Hadoop“ yra atvirojo kodo paskirstytojo apdorojimo karkasas (framework), valdantis didžiųjų duomenų apdorojimą ir saugojimą keičiamo dydžio kompiuterių serverių grupėse. Tai didžiųjų duomenų technologijų, kurios pirmiausia naudojamos duomenų mokslo ir pažangios analizės iniciatyvoms, įskaitant prognostinę analizę, duomenų gavybą, mašininių ir gilųjų mokymąsi, ekosistemos centras (<https://www.techtarget.com/searchdatamanagement/definition/Hadoop>).

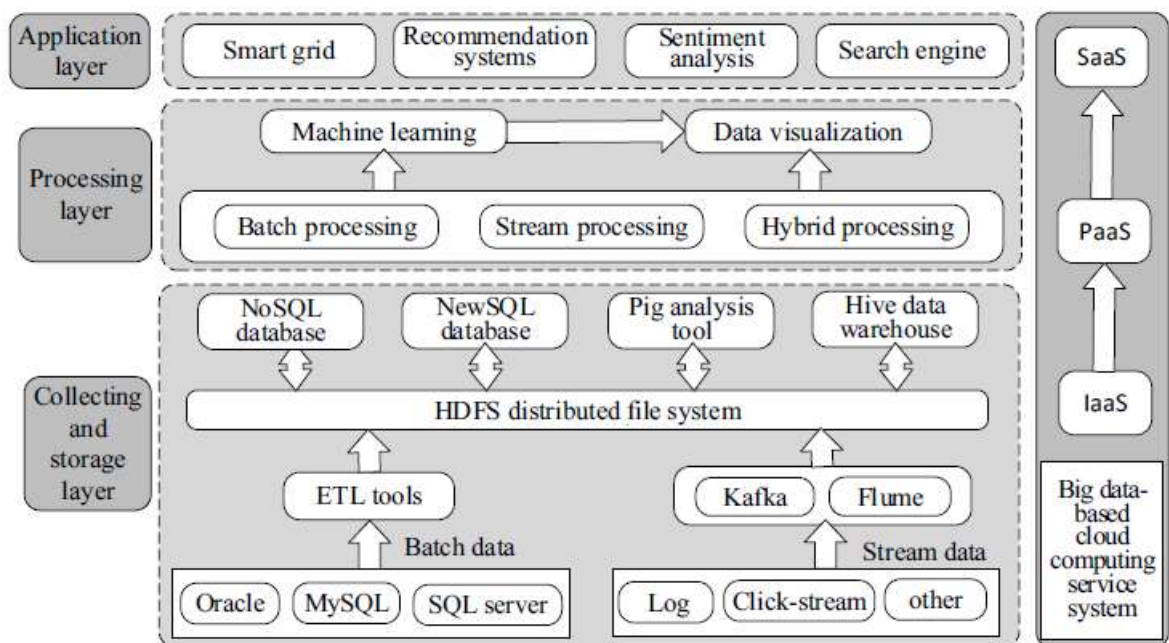


Šaltinis: <https://www.educba.com/big-data-technologies/>

Apie didžiųjų duomenų technologijas išsamiai rasite: Abdalla, H. B. (2022). A brief survey on big data: technologies, terminologies and data-intensive applications. *Journal of Big Data*, 9(1), 1-36.

Didžiųjų duomenų paslaugų architektūra

Šaltinis: Wang, J., Yang, Y., Wang, T., Sherratt, R. S., & Zhang, J. (2020). Big data service architecture: a survey. *Journal of Internet Technology*, 21(2), 393-405.

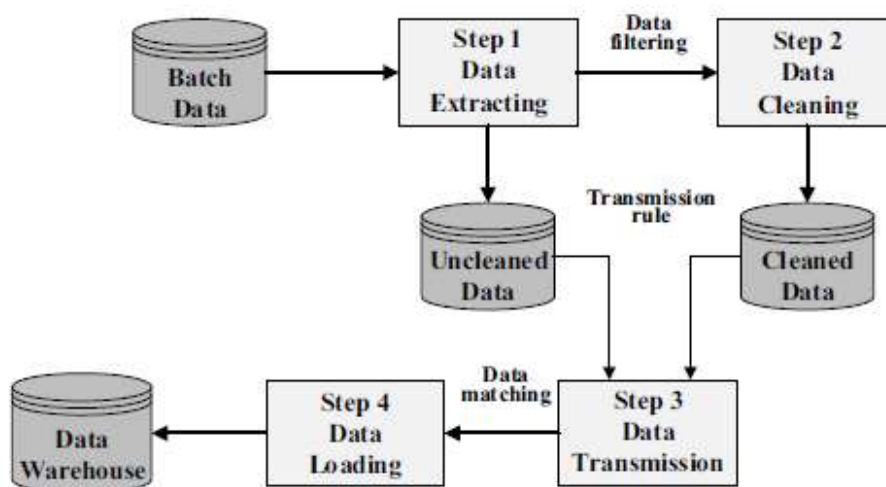


Architektūrą sudaro 3 sluoksniai. **Duomenų surinkimo ir saugojimo sluoksnyje** (Collecting and storage layer) duomenys surenkami panaudojant atitinkamą įrangą ir iš anksto apdoroti duomenys (pre-processed data) saugomi paskirstytoje failų sistemoje (HDFS (Hadoop distributed file system) distributed file system) arba duomenų bazėje.

Duomenų apdorojimo sluoksnyje (Processing layer) įvairūs duomenų apdorojimo karkasai pritaikomi skirtingiems duomenų tipams. Išsami duomenų analizė pagrįsta didelės apimties mašininio mokymosi technologijomis, kurios padidina duomenų vertę. Vizualizacijos įrankiai naudojami rezultatams pristatyti.

Taikymo sluoksnyje (Application layer) yra didžiųjų duomenų taikymai įvairiose srityse. Taip pat panaudojama debesų kompiuterija, kuri apima skaičiavimams skirtą programinę įrangą ir paslaugas (SaaS (Software as a Service) - programinės įrangos pristatymo ir licencijavimo būdas, kai programinė įranga pasiekama internetu per prenumeratą, o ne perkama ir įdiegiama atskiruose kompiuteriuose, PaaS (Platform as a service) – yra debesų kompiuterijos modelis, suteikiantis klientams visą debesų platformą – aparatinę įrangą, programinę įrangą ir infrastruktūrą, skirtą programoms kurti, paleisti ir valdyti be sąnaudų, sudėtingumo ir nelankstumo, kurie dažnai atsiranda kuriant ir palaikyti tą platformą, IaaS (infrastructure as a Service) – tai yra debesų kompiuterijos forma, teikianti pagrindinius skaičiavimo, tinklo ir saugojimo išteklius vartotojams pagal poreikį, internetu ir už mokestį „you-go“ pagrindu. IaaS suteikia galimybę galutiniams vartotojams pagal poreikį padidinti ir sumažinti išteklius, todėl sumažėja didelių išankstinių kapitalo išlaidų arba nereikalingos „priklausomos“ infrastruktūros poreikis, ypač esant „smagiam“ darbo krūviui. Priešingai nei PaaS ir SaaS, IaaS įgalina žemiausio lygio išteklių valdymą debesyje.

Išsamiau aptarsime ETL (Extract-Transform-Load) – duomenų išgavimo, transformavimo ir įkėlimo procesą.

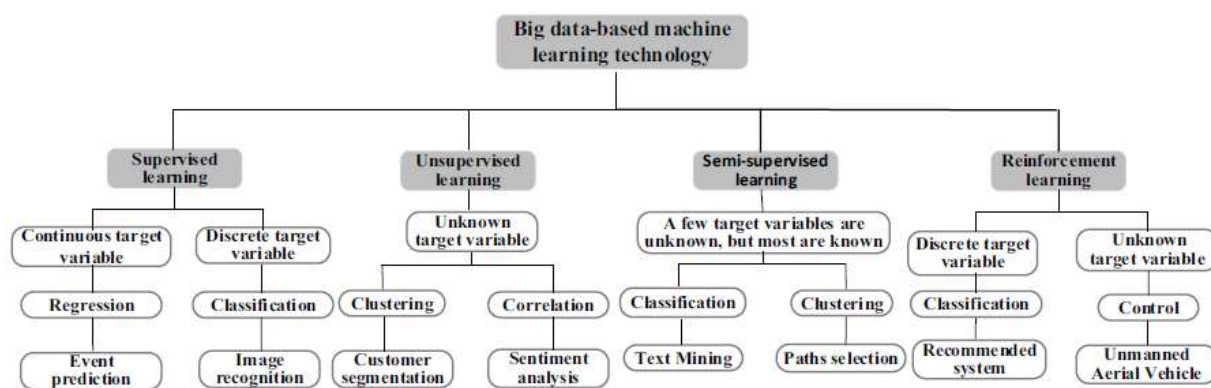


Didžiųjų duomenų formos apima **statinius paketinius duomenis** (Batch data, tai duomenų rinkinys arba grupė, kurią galima apdoroti arba analizuoti kaip vienetą) ir **dinaminis srauto duomenis** (Stream data, tai nuolatinis, besikeičiantis duomenų srautas). Dėl srauto duomenų perdavimo nestabilumo, jų rinkimas skiriasi nuo paketinių duomenų rinkimo. Paketinių duomenų surinkimui ir perdavimui (Transmission) iš skirtingų šaltinių naudojami ETL įrankiai.

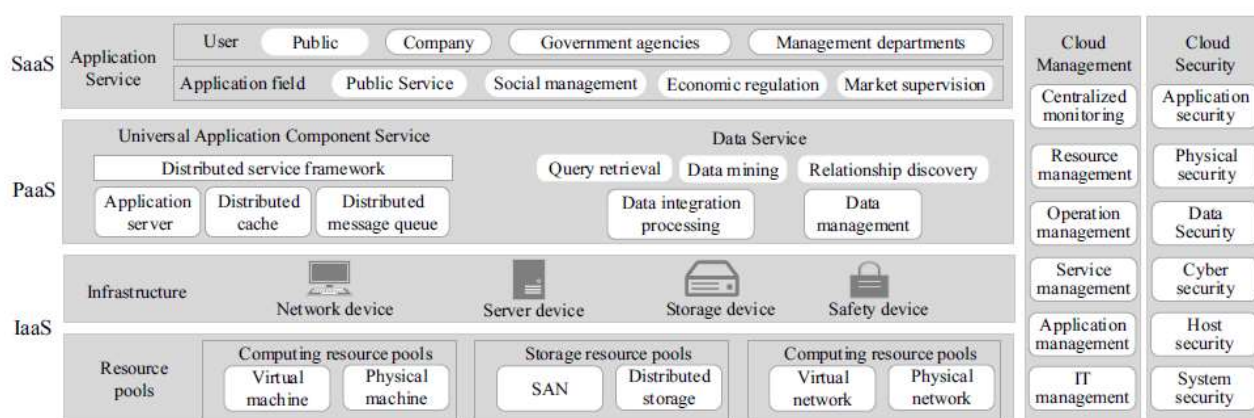
ETL pašalina sugadintus ar triukšmingus duomenis atlikdami duomenų apdorojimo operacijas, pvz., prijungimą, transformavimą ir valymą. Plačiai naudojami ETL įrankiai yra Kettle, Datastage, Informatica ir kt.

Srauto duomenims, kurie renkami realiaame laike, reikalingi įrankiai, kurie garantuoja momentišumą, atsparumą gedimams, stabilumą ir patikimumą. Flume yra patikima ir atspari gedimams paskirstyta srauto apdorojimo sistema, kuri renka, kaupia ir perduoda daug žurnalo duomenų iš skirtingų šaltinių į centralizuotą saugyklą. Kafka yra universali atviro kodo pranešimų sistema, kuri daugiausia naudojama kuriant realiu laiku duomenų srautus ir srautinio perdavimo taikymus. Siekiant dar labiau optimizuoti valdymą ir duomenų srauto apdorojimo greitį, Kafka apdorojimui naudoja eiles, kad būtų išvengta apdorojimo asinchroniškumo tarp duomenų generavimo ir apdorojimo greičio.

Apdorojant didžiuosius duomenis naudojamos **mašininio mokymosi**² technologijos.



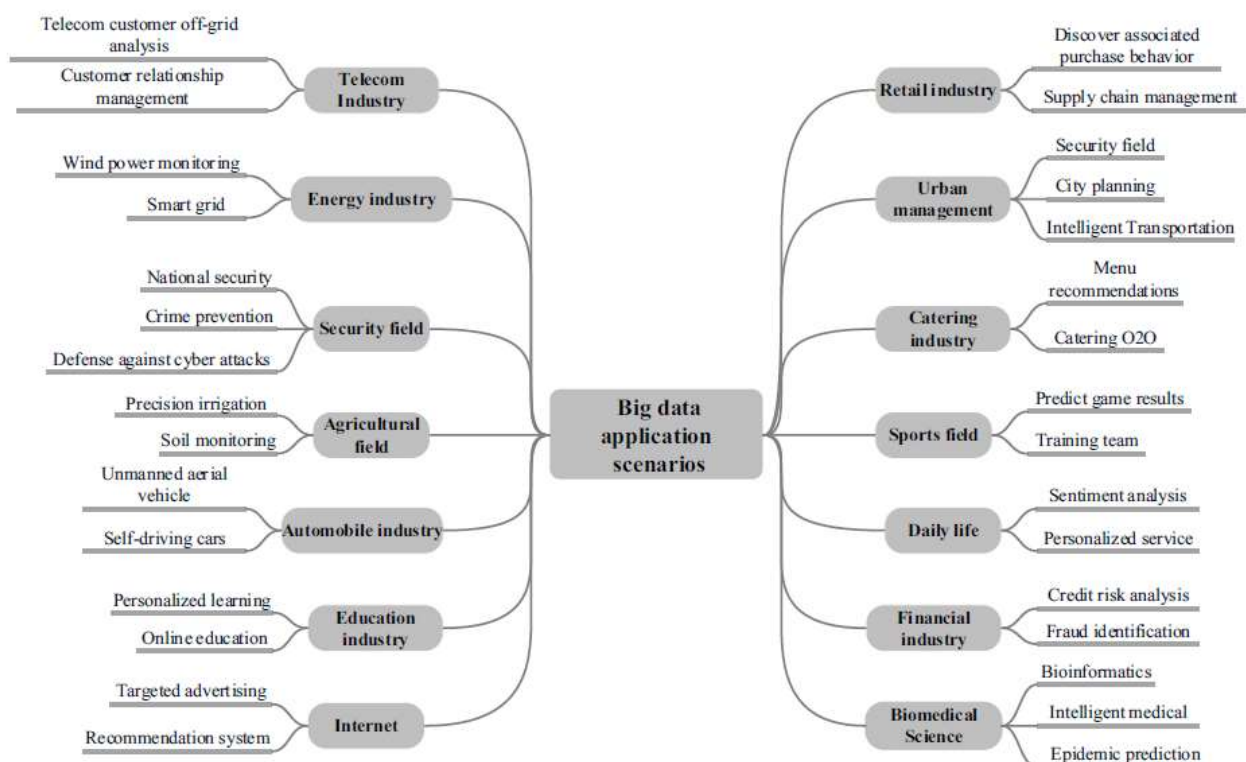
Didžiaisiais duomenimis pagrįstos **debesų kompiuterijos paslaugų sistemos** pasižymi tuo, kad turi galingus paskirstytus apdorojimo variklius, paskirstytas duomenų bazines, debesų saugyklą ir palaiko virtualizacijos technologijas.



² Mašininis mokymasis - dirbtinio intelekto algoritmų klasė, kuriai būdingas ne tiesioginis problemos sprendimas, o mokymasis, kaip pritaikyti daugelio panašių problemų sprendimus. Apima metodų, mokinančių kompiuterius „mąstyti“, kūrimą. Tai yra programų kūrimo būdas, kai sukurta sistema prisitaiko prie duomenų („apsimoko“). Šie algoritmai sugeba ilgainiui pasiekti geresnių rezultatų, patys kaupdami patyrimą.

Šaltiniai: https://lt.wikipedia.org/wiki/Ma%C5%A1ininis_mokymasis, A. Paulauskaitė-Tarasevičienė, K. Šutienė (2022). Intelektikos pagrindai, KTU leidykla „Technologija“

Didžiųjų duomenų taikymo scenarijai



Didžiųjų duomenų keliami iššūkiai

Šaltinis: Abdalla, H. B. (2022). A brief survey on big data: technologies, terminologies and data-intensive applications. *Journal of Big Data*, 9(1), 1-36.

Iššūkiai, susiję su didžiųjų duomenų taikymais

- Dauguma didelių duomenų analizės sistemų apsiriboja „Apache Hadoop“ rinkiniu. Vadinasi, duomenų saugojimo ir valdymo sluoksnių išskyrimas yra iššūkis. Sluoksnių išskyrimas padidina didžiųjų duomenų funkcionalumą taikymuose.
- Pagrindinė problema, su kuria susiduriama, yra paskirstytų ir centralizuotų funkcijų derinimas didžiųjų duomenų taikymuose, pagrįstuose **giliuoju mokymusi**³ (deep learning).

Iššūkiai, susiję su didžiųjų duomenų analitika

- Duomenų apsauga, saugumas ir pasitikėjimas. Labai aiškiai apibrėžti duomenų naudojimo tikslai, asmens duomenų apsauga.
- Dalijimasis informacija ir jos valdymas. Ypatingai daug dėmesio skiriama atviriems duomenims, jų naudojimui laikantis įstatymų.
- Inovacijos ir analitika: saugyklos, duomenų išgavimas iš saugyklų; duomenų kiekio didėjimas; struktūrizuoti ir nestruktūrizuoti duomenys; duomenų nuosavybės ir iškraipymo problemos; duomenų įvairovė; riboti resursai didžiųjų duomenų apdorojimui; ribinių duomenų (edge data) apdorojimas; energijos suvartojimas.

Šaltinis: Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review*, 1(2), 293-314.

³ Gilusis mokymasis yra mašininio mokymosi atšaka, kur kiekvienas sluoksnis išmoksta vis abstraktesnių duomenų bruožų. Tai neuroninis tinklas, turintis daugybę sluoksnių. Šaltinis: A. Paulauskaitė-Tarasevičienė, K. Šutienė (2022). Intelektikos pagrindai, KTU leidykla „Technologija“

- Sudėtingi duomenys, išgaunami iš įvairių šaltinių. Sunku numatyti priklausomybes tarp šių duomenų.
- Triukšmingi duomenys: matavimo paklaidos, nuokrypiai, trūkstamos reikšmės.
- Priklausomi duomenys: dalis duomenų yra susiję su santykinai silpnais signalais.

Šaltinis: Rakšnys, A. V., Gudelis, D., & Guogis, A. (2021). Didžiųjų duomenų ir dirbtinio intelekto technologijų pritaikymo galimybių viešojo valdymo srityje ir socialinėje politikoje analizė.

Dėl savo turimų įsipareigojimų viešojo sektoriaus institucijos, pradedant sveikatos apsauga ir socialinėmis paslaugomis, surenka daugybę jautrios informacijos apie savo piliečius. Šiuos duomenis naudojant kartu, būtų galima susidaryti beveik pilną individo gyvenimo vaizdą, tokiu būdu kėsinant į asmeninio gyvenimo privatumą. Organizacijos gali rinkti duomenis tik tada, kai to reikia įgyvendinti jų misiją, tačiau toks „informacinis delegavimas“ dažnai yra neaiškus, nestabilus ir gali būti įvairiai interpretuojamas.

Vertinant iš etinės perspektyvos, dėl vartotojų duomenų panaudojimo galėtų būti labai sudėtinga arba net neįmanoma institucijoms gauti paslaugų vartotojų sutikimą. Dar vienas klausimas kyla dėl to, kad net jeigu ir pavyktų sutikimą gauti, tai ar ir kiek vartotojai galėtų būti informuojami apie jų duomenų panaudojimą (pvz., suprantant, kam ir kaip jie bus naudojami), nes labai dažnai didieji duomenys gali būti renkami turint vieną tikslą, o panaudojami visai kitam tikslui (Gillingham and Graham, 2016). Yra grėsmė, kad, turint „didžiuosius duomenis“, gali būti peržengta „raudonoji linija“ vartotojų atžvilgiu, pavyzdžiui, pasitaikius klaidų, kai į rinką rinkodaros būdu būtų pristumiami kenksmingi jiems produktai (Coulton et. al., 2015). Didieji duomenys yra susiję ir su pačių piliečių socialiniu aktyvumu. E. Hargittai (2015) atkreipia dėmesį į situacijas, kai pasikliovimas tik didžiais duomenimis gali sukurti papildomų sunkumų, kai tam tikri svarbūs reiškiniai gali likti nepastebėti, nes jie susiję su tuo gyventojų segmentu, kuris nepalieka reikšmingo skaitmeninio pėdsako. Vyresni žmonės ir individai, gaunantys mažesnes pajamas, taip pat – individai, stokojantys išsilavinimo, nėra tinkamai reprezentuojami socialinės medijos duomenyse, o jaunimas ir aukštesniam socialiniam sluoksniui priklausantys individai yra per daug atspindimi duomenimis apie viešųjų paslaugų vartojimą (Coulton et. al., 2015). Veiksmingai naudoti didžiuosius duomenis būtinos ir tokios specifinės tarnautojų kompetencijos: programavimo, modeliavimo, statistikos, duomenų valdymo, analitikos įgūdžiai, mašininio mokymosi procesų supratimas, taip pat technologinė infrastruktūra (Maciejewski, 2016).

Analizuojant esminius trūkumus, svarbu paminėti ir pačių duomenų skaitmenizavimo procesų sudėtingumą, skirtingą duomenų pobūdį ir su jais susijusį netikrumą. Institucijos gali susidurti su iššūkiais numatydamos duomenų saugojimo vietų kūrimą, duomenų sistemų, kurios galėtų duomenis sujungti iš skirtingų šaltinių, pvz., kitų viešojo sektoriaus institucijų ar net nevyriausybinių bei verslo struktūrų, surinkimo kūrimą ir įdiegimą. Paminėtinas ir veiksmingas duomenų analizės metodų ir technikų poreikis, siekiant užtikrinti kompleksinį duomenų valdymą (Rogge et al., 2017). Dirbant su didžiais duomenimis, tradicinių duomenų valdymo metodų nepakanka (Coulton et. al., 2015).

Skyrelis 2.1. medžiagai įtvirtinti siūlomas testas su pasirenkamaisiais atsakymais (anglų k.)

Šaltinis: <https://study.com/academy/exam/topic/big-data-fundamentals.html>

Alternatyvos: minčių žemėlapių sudarymas, papildomi didžiųjų duomenų atributai remiantis straipsniu: Khan, N., Naim, A., Hussain, M. R., Naveed, Q. N., Ahmad, N., & Qamar, S. (2019, May). The 51 v's of big data: survey, technologies, characteristics, opportunities, issues and challenges. In *Proceedings of the international conference on omni-layer intelligent systems* (pp. 19-24).

2.2. Atviri duomenys

Atvirų duomenų apibrėžimas ir savybės

Šaltinis: <https://ivpk.lrv.lt/lt/atviri-duomenys-1#:~:text=Atviri%20duomenys%20-%20laisvai%20prieinami%20institucijos%20veikloje%20ar,ir%20tik%20tomis%20pa%C4%8Ddiomis%20s%C4%85lygomis%2C%20kuriomis%20buvo%20gauti.>

Atviri duomenys - laisvai prieinami institucijos veikloje ar dokumentuose užfiksuoti duomenys, informacija ar jos dalis, nepaisant jų pateikimo būdo, formos ir laikmenos, įskaitant registro duomenis, registro informaciją, registruoti pateiktų dokumentų ir (arba) jų kopijų duomenis, valstybės informacinės sistemos duomenis, kuriuos visi asmenys gali pakartotinai naudoti ir platinti bet kokių tikslų, nurodydami jų šaltinį ir tik tomis pačiomis sąlygomis, kuriomis buvo gauti.

Atvirų duomenų savybės (šaltinis: <https://opendatahandbook.org/guide/lt/what-is-open-data/>)

- **Prieinamumas ir prieiga:** duomenys turi būti prieinami pilnais rinkiniais ir ne brangiau, nei kainuoja juos pateikti. Geriausia - jei duomenis galima atsisiųsti internetu. Duomenys taip pat turi būti pateikiami patogiu ir redaguojamu formatu.
- **Pakartotinis panaudojimas (pernaudojimas) ir platinimas:** duomenys turi būti pateikti taip, kad leistų pakartotinį panaudojimą ir platinimą įskaitant ir duomenų, sujungtų su kitais duomenimis, panaudojimą ir platinimą.
- **Visuotinis dalyvavimas:** visi turi turėti galimybę naudoti, pakartotinai panaudoti ir platinti informaciją - neturi būti prieigos prie duomenų diskriminavimo jokiems asmenims ar jų grupėms. Pavyzdžiui, neturi būti “nekomercinių” apribojimų, trukdančių naudoti duomenis komerciniais tikslais, taip pat neturi būti reikalavimų naudoti duomenis tik tam tikrais tikslais (pavyzdžiui, tik akademiniais).

Svarbu aiškiai apibrėžti, ką reiškia atvirumas ir kodėl šis apibrėžimas naudojamas, tai paprastas atsakymas yra: **interoperabilumas**.

Interoperabilumas suprantamas kaip įvairių organizacijų ir sistemų sugebėjimas veikti (dirbti) kartu. Šiuo atveju kalbame apie sugebėjimą sujungti skirtingus duomenų rinkinius.

Interoperabilumas svarbus todėl, kad leidžia kartu veikti skirtingiems komponentams. Galimybė skaidyti komponentais ir tuos komponentus jungti svarbi konstruojant dideles ir sudėtingas sistemas. Be interoperabilumo tai tampa beveik neįmanoma - tą parodo ir žinomiausia Babilono bokšto legenda, kurioje nesugebėjimas komunikuoti baigėsi visišku bokšto statybos žlugimu.

Duomenų požiūriu, susiduriame su panašia problema. Esminė duomenų (ar kodo) idėja yra ta, kad vienas “atviros” medžiagos gabalas gali būti laisvai sujungiamas su kitu “atviros” medžiagos gabalu. Toks interoperabilumas yra būtinas: norint pasiekti didžiausią praktinę “atvirumo” naudą, reikia sudaryti galimybę kombinuoti skirtingus duomenų rinkinius, tokiu būdu leidžiant sukurti daugiau ir geresnių produktų bei paslaugų (ši nauda detaliau aptariama skyriuje “Kodėl reikia atvirų duomenų”).

Aiškiai apibrėžus atvirumą, užtikrinama, kad jei turite du skirtingus duomenų rinkinius iš dviejų skirtingų šaltinių, jūs galėsite juos sukombinuoti tarpusavyje. Užtikrinama, kad **išvengsime “Babilono bokšto”**: daugybės duomenų rinkinių, kurie tarpusavyje nesuderinami ir negali būti sujungti į vieną didelę sistemą, kurioje ir būtų didžiausia vertė.

Svarbiausia - atveriant duomenis, fokusuotis į nuasmenintus duomenis, tai yra, tokius, kurie nesuteikia jokios informacijos apie konkrečius asmenis.

Kai kurių rūšių valstybinė informacija taip pat gali būti neskelbiama dėl valstybės saugumo.

Atvirų duomenų šaltiniai

- Lietuvos atvirų duomenų portalas: <https://data.gov.lt/>
- <https://www.kaggle.com/datasets>
- <https://data.europa.eu/lt>
- <https://www.opendatanetwork.com/>
- <https://data.unicef.org/resources/resource-type/datasets/>
- <https://learn.g2.com/open-data-sources>
- <https://datahub.io/collections>

Skyrelio 2.2. medžiagai įtvirtinti siūlomas kursas apie atvirus duomenis:

<https://data.europa.eu/elearning/en/#/id/co-01>

Alternatyva: minčių žemėlapis sudarymas.

2.3. Kintamieji ir matavimo skalės

Šaltinis: V. Čekanavičius, G. Murauskas (2003). Statistika ir jos taikymai 1. Vilnius, TEV.

Pasirenkant duomenų analizės metodus labai svarbu nustatyti, kokia yra duomenų prigimtis. Duomenų imties elementus vienija tiriamieji požymiai. Matuodami šiuos požymius, gauname tam tikrus dydžius. Tas dydis vadinamas kintamuoju. Pagal prigimtį kintamieji skirstomi į **kiekybinius** ir **kokybinius**.

Kiekybinio kintamojo reikšmė – tai atsakymas, *kiek* tiriamo požymio turi imties elementas. Kokybiniai kintamieji nusako dydžius, kurių neįmanoma įvertinti skaičiais. Kiekybinių kintamųjų pavyzdžiai: laikas, greitis, aukštis, pažymių vidurkis ir pan. Kokybiniai kintamieji – kraujo grupė, pasiekimų lygmuo, klasė.

Kiekybiniai kintamieji skirstomi į **tolydžiuosius** ir **diskrečiuosius**. Kiekybinis kintamasis vadinamas **tolydžiuoju**, jei jo reikšmių skirtumas gali būti kiek norima mažas. Kiekybinis kintamasis, kurio reikšmių skirtumas yra ne mažesnis už tam tikrą minimalų pokytį, vadinamas **diskrečiuoju** kintamuoju. Tolydžiųjų kintamųjų pavyzdžiai: laikas, masė, aukštis, greitis, diskrečiųjų – mokinių skaičius klasėje, klaidų skaičius, mašinų skaičius.

Duomenys ir kintamieji yra klasifikuojami ir atsižvelgiant į naudotą matavimų skalę. Yra naudojamos 4 matavimų skalės: **pavadinimų, rangų, intervalų, santykių**.

Pavadinimų skalė dar vadinama **nominaliąja** arba **klasifikacine** skale. Kintamieji, kurie matuojami pavadinimų skalėje, vadinami **nominaliaisiais** kintamaisiais. Pagal kintamojo reikšmes, gautas naudojant pavadinimų skalę, duomenis galima tik klasifikuoti. Nominaliųjų kintamųjų pavyzdžiai: futbolo marškinėlių numeris, kraujo grupė, tautybė ir pan. Kiekvienas imties elementas turi turėti jam tinkamą kategoriją ir visos kategorijos turi aiškiai skirtis.

Rangų skalė dar vadinama **tvarkos** skale. Ši skalė naudojama tuomet, kai pagal tiriamo požymio skirtumus galima objektus išrikiuoti į eilę. Kintamieji, kurie matuojami rangų skalėje, vadinami **ranginiais** kintamaisiais. Pagal ranginių kintamųjų reikšmes objektus galima ne tik skirstyti į klases, bet ir sutvarkyti. Ranginių kintamųjų pavyzdžiai: vietų paskirstymas olimpiadoje, kvalifikacinės darbuotojų kategorijos.

Ranginiai ir nominalieji kintamieji vadinami **kategoriniais** (kokybiniais).

Jei naudojama intervalų skalė, objektus galima klasifikuoti, tvarkyti ir kiekybiškai įvertinti skirtumus. **Intervaliniai** duomenys visada **skaitiniai**. Skirtumas tarp dviejų kintamojo reikšmių rodo, kiek daugiau arba mažiau matuojamojo požymio yra viename elemente, palyginus su kitu

elementu. Intervalinių kintamųjų pavyzdžiai: temperatūros matavimai, kalendorinis laikas, intelekto koeficiento vertinimas.

Santykių skalė nuo intervalų skalės skiriasi tuo, kad joje yra apibrėžta **absoliuti atskaitos pradžia**. Šioje skalėje yra **absoliutusiasis nulis**, t.y. nulinis taškas, rodantis, kad tiriamojo požymio nėra. Rezultatai visais atvejais bus neneigiami skaičiai. Skaičių, gautų matuojant požymius, santykis parodo kiekybinį matuojamojo požymio santykį. Santykių skalėje jis nuo matavimo vienetų nepriklauso. Santykių skalės kintamieji: ūgis, svoris, amžius, kaina.

Kiekybiniais kintamiesiems matuoti naudojama intervalų arba santykių skalė. Kokybiniais kintamiesiems matuoti naudojama pavadinimų arba rangų skalė. Iš kiekybinių tolydžiuųjų kintamųjų galima gauti ranginius kintamuosius, o iš ranginių – nominaliuosius. Tokiu atveju prarandama dalis informacijos. Pvz., santykių skalėje apibrėžtą ūgį galime pakeisti ranginiu: aukštas, vidutinio ūgio, žemas.

Kintamuosius galima klasifikuoti pagal įgyjamų reikšmių skaičių. Kintamasis, įgyjantis tik dvi reikšmes, vadinamas **dvireikšmiu** arba **binariuoju**.

Skyrelis 2.3. medžiagai įtvirtinti siūlomi klausimai

1. Kurie kintamieji yra diskretieji, o kurie tolydieji?

- | | |
|---|---|
| a. Šeimos narių skaičius | D |
| b. Vidutinis sportui per savaitę skirtas laikas | T |
| c. Vandens kiekis stiklainyje | T |
| d. Krepšininkų komandos žaidėjų vidutinis ūgis | T |
| e. Informatikos kontrolinio darbo pažymys | D |

2. Kurie kintamieji bus kiekybiniai, o kurie kokybiniai?

- | | |
|---|------|
| a. Mokinių skaičius klasėje | KIEK |
| b. Parduotuvėje išleista pinigų suma | KIEK |
| c. Pritarimas mokslo metų ilginimui | KOK |
| d. Mokinio amžius | KIEK |
| e. Dorinio ugdymo (tikyba arba etika) grupė | KOK |
| f. Vandens talpos dydis | KIEK |
| g. Akių spalva | KOK |

3. Pagal kokią skalę matuojami kintamieji?

- | | |
|---|------------|
| a. Mašinos spalva | pavadinimų |
| b. Drabužių dydis | rangų |
| c. Metų, praleistų mokantis užsienio kalbos, skaičius | intervalų |
| d. Mokinio amžius | santykių |

4. Nustatykite, pagal kokią skalę matuojami kintamieji, jei klausimyne pateikti klausimai:

- | | |
|--|------------|
| a. Gimimo vieta | pavadinimų |
| b. Amžius | santykių |
| c. Darbo patirtis metais | intervalų |
| d. Užsienio kalbos mokėjimas (nuo 0 iki 5) | rangų |
| e. Pageidaujamas atlyginimas | santykių |
| f. Darbo kompiuteriu įgūdžiai (nuo 0 iki 10) | rangų |
| g. Pomėgiai | pavadinimų |

2.4. Vertinimas

Siūloma taikyti kaupiamąjį vertinimą, įvertinant kiekvieno skyrelio užduotis santykiu: 2.1 ir 2.2 skyrelių užduočių įvertinimai po 40 procentų, 2.3 skyrelis – 20 procentų.

3. Skaičiuoklės naudojimas duomenų tyryboje

Skaičiuoklė yra vienas iš įrankių, skirtų duomenų tvarkymui ir analizei. Skaičiuoklės panaudojimo galimybės:

Duomenų įvedimas ir organizavimas lentelėse. Galima naudoti, pavyzdžiui, įvedant rezultatus iš tyrimų ar apklausų.

Duomenų filtravimas ir rūšiavimas. Naudojant skaičiuoklės filtravimo ir rūšiavimo funkcijas, galima lengvai išskirti specifinius duomenis, pavyzdžiui, pažiūrėti tik į tam tikrą grupę tiriamųjų.

Skaičiavimai ir funkcijos. Galima atlikti skaičiavimus ir taikyti formules, kad automatiškai atliktumėte sudėtingus skaičiavimus. Tai gali būti naudinga, jei reikia apskaičiuoti vidurkius, sumas, medianas ir kt.

Diagramų ir grafikų kūrimas. Excel leidžia kurti įvairius grafikus ir diagramas, kurie padeda vizualizuoti duomenų tendencijas ir ryšius.

Duomenų analizė. Naudojant „PivotTable“ ir „PivotChart“ funkcijas galima atlikti sudėtingesnę duomenų analizę, išskirti tendencijas ir iš duomenų išgauti informaciją.

Statistinė analizė. Skaičiuoklės turi įrankių, padedančių atlikti statistinę analizę, tokią kaip vidurkio, medianos, modos, standartinio nuokrypio ir kt. skaičiavimus.

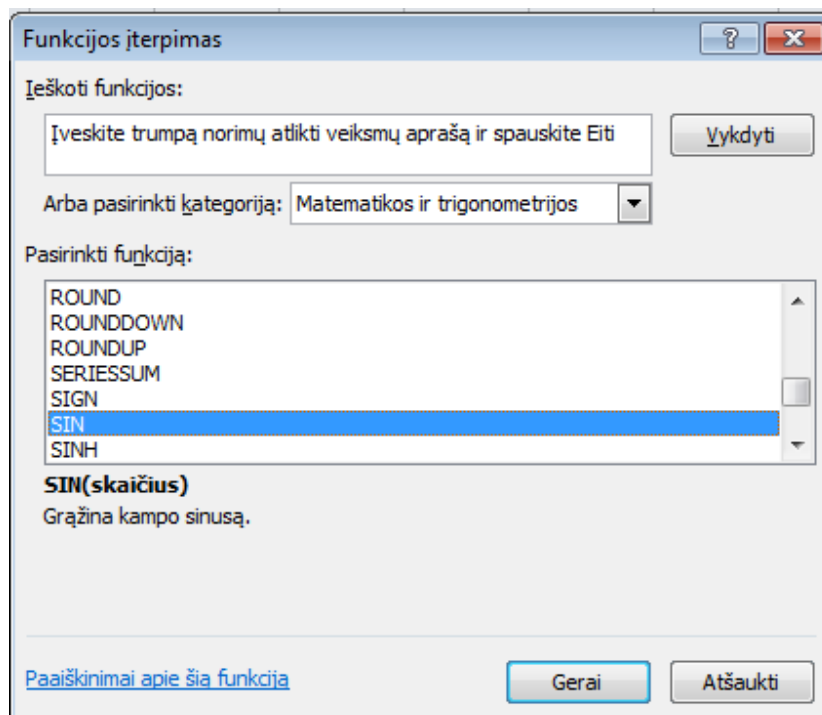
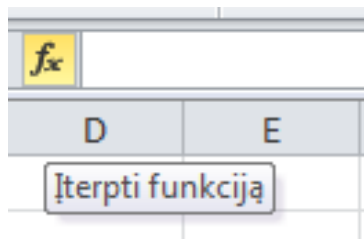
Regresinė analizė. Jei tiriate ryšį tarp kintamųjų, skaičiuoklės leidžia atlikti paprastą regresinę analizę ir pavaizduoti rezultatus grafiškai.

Duomenų vizualizacija ir interpretavimas. Galima sukurti ataskaitų šablonus su grafinėmis iliustracijomis ir lentelėmis, kad galėtumėte aiškiai pateikti savo tyrimo rezultatus ir išvadas.

Sąryšių analizė. Naudojant skaičiuoklės galima atlikti sąryšių analizę, ištirti, kaip skirtingi duomenų kintamieji gali būti susiję tarpusavyje.

Simuliacijos ir scenarijų analizė. Galima kurti modelius ir realizuoti skirtingus scenarijus, kad pamatytumėte, kaip tam tikri veiksniai gali paveikti rezultatus.

3.1. Kaip greitai susirasti norimos funkcijos aprašymą ir panaudojimo pavyzdžių?



Funkcijos SIN sintaksė turi tokius **argumentus**:

- **Skaičius** Būtinai. Kampas radianais, kurio sinusą norite apskaičiuoti.

Pastaba

Jei argumentas nurodytas laipsniais, padauginkite jį iš $\text{PI}()/180$ arba konvertuokite jį į radianus naudodamiesi funkcija RADIANS.

Pavyzdys

Pavyzdį bus lengviau suprasti, jei jį nukopijuosite į tuščią darbalapį.

☒ Kaip kopijuoti pavyzdį?

	A	B
1	Formulė	Aprašas (rezultatas)
2	=SIN(PI())	Pi sinusas radianais (0, apytiksliai)
3	=SIN(PI()/2)	Pi/2 sinusas radianais (1)
4	=SIN(30*PI()/180)	30 laipsnių sinusas (0,5)
5	=SIN(RADIANS(30))	30 laipsnių sinusas (0,5)

3.2. Kaip teisingai užrašyti funkcijos argumentus?

Kiekviena funkcija ar formulė pradedama = ženklu.

	A	B	C	D
1	14,525	=Round(		
2		ROUND(skaičius; dešimt_kiekis)		
3				

Lentelėse paryškintu šriftu pateiktos funkcijos, kurių nebuvo ankstesnėje programoje.

3.3. Funkcijos

Matematinės, trigonometrinės, statistinės funkcijos

Funkcija	Paskirtis
ABS(skaičius)	Grąžina skaičiaus modulį.
SQRT(skaitinis argumentas)	Grąžina skaičiaus kvadratinę šaknį.
POWER(laipsnio pagrindas; laipsnio rodiklis)	Grąžina skaičiaus kėlimo tam tikru laipsniu rezultatą.
Laipsnio pagrindas ^ Laipsnio rodiklis	Kėlimui laipsniu galima naudoti simbolį ^.
SIN(kampas, išreikštas radianais)	Grąžina duoto kampo sinusą. Jei argumentas nurodytas laipsniais, padauginkite jį iš $\text{PI}()/180$ arba konvertuokite jį į radianus naudodamiesi funkcija RADIANS.

COS(kampas, išreikštas radianais)	Grąžina duoto kampo kosinusą. Jei argumentas nurodytas laipsniais, padauginkite jį iš PI()/180 arba konvertuokite jį į radianus naudodamiesi funkcija RADIANS.
RADIANS(kampas laipsniais)	Laipsnius konvertuoja į radianus.
PI()	Grąžina skaičių 3,14159265358979, matematinę pi konstantą, 15 skaičių po kablelio tikslumu. Funkcija neturi argumentų.
ROUND(skaičius; dešimtinių skaitmenų kiekis)	Suapvalina skaičių iki nurodyto skaitmenų skaičiaus.
ROUNDUP(skaičius; skaitmenų_skaicius)	Suapvalina skaičių iki didesnio skaičiaus.
ROUNDDOWN(skaičius; skaitmenų_skaicius)	Suapvalina skaičių iki mažesnio skaičiaus.
SUM(Arg1;Arg2;Arg3;...)	Sudeda visus skaičius, kuriuos nurodote kaip argumentus. Kiekvienas argumentas gali būti diapazonas (diapazonas: du arba daugiau langelių darbalapyje), langelio koordinatės, konstanta, formulė, arba kitos funkcijos rezultatas.
SUMIF(diapazonas;kriterijus; sumavimo diapazonas)	Reikšmėms, esančioms diapazone ir atitinkančioms jūsų nurodytus kriterijus, sumuoti.
SUMIFS(sumos_diapazonas; kriterijų_diapazonas1; kriterijai1; kriterijų_diapazonas2; kriterijai2]; ...)	Sudeda visus savo argumentus, atitinkančius keletą kriterijų.
AVERAGE(Arg1;Arg2;Arg3;...)	Grąžina argumentų aritmetinį vidurkį.
AVERAGEIF(diapazonas; kriterijai; [vidurkio diapazonas])	Grąžina visų diapazono langelių, atitinkančių nurodytus kriterijus, vidurkį (aritmetinį vidurkį).
AVERAGEIFS(vidurkio_diapazonas; kriterijų_diapazonas1; kriterijus1; [kriterijų_diapazonas2; kriterijus2]; ...)	Grąžina visų langelių, atitinkančių kelis nurodytus kriterijus, vidurkį (aritmetinį vidurkį).
MIN(Arg1;Arg2;Arg3;...)	Grąžina mažiausią skaičių iš reikšmių aibės.
MINIFS(minimumo_diapazonas; kriterijų_diapazonas1;kriterijai1;[kriterijų_diapazonas2; kriterijai2]; ...)	Pateikia mažiausią langelių, atitinkančių tam tikrą sąlygų ar kriterijų rinkinį, reikšmę.
MAX(Arg1;Arg2;Arg3;...)	Grąžina didžiausią skaičių iš reikšmių aibės.
MAXIFS(maksimalus_diapazonas; kriterijų_diapazonas1; kriterijai1; [kriterijų_diapazonas2; kriterijai2]; ...)	Pateikia didžiausią langelių, atitinkančių tam tikrą sąlygų ar kriterijų rinkinį, reikšmę.
COUNT(Arg1;Arg2;Arg3;...)	Skaičiuoja, keli argumentai yra skaitiniai.

COUNTA(reikšmė1; [reikšmė2], ...)	Skaičiuoja, keli langeliai yra užpildyti informacija.
COUNTIF(diapazonas;kriterijus)	Skaičiuoja, kiek diapazone yra langelių, atitinkančių nurodytą kriterijų.
COUNTIFS(kriterijų_diapazonas1; kriterijai1; [kriterijų_diapazonas2; kriterijai2];...)	Funkcija COUNTIFS kriterijus taiko langeliams keliuose diapazonuose ir skaičiuoja, kiek kartų visi kriterijai yra įvykdyti.
MODE(skaičius1;[skaičius2];...)	Gražina dažniausiai pasitaikančią arba pasikartojančią reikšmę masyve arba duomenų diapazone.
MEDIAN(skaičius1; [skaičius2]; ...)	Gražina duotų skaičių medianą. Mediana yra skaičius skaičių aibės viduryje.
STDEV(skaičius1;[skaičius2];...)	Skaičiuoja standartinį imties nuokrypį. Standartinis nuokrypis rodo, kiek daug reikšmės skiriasi nuo vidurkio (vidutinės reikšmės).
SUBTOTAL(funkcijos_nr; nuoroda1;[nuoroda2];...)	Pateikia sąrašo tarpinę sumą, vidurkį, kiekį, mažiausią, didžiausią reikšmę, vidutinį kvadratinį nuokrypį.

Loginės funkcijos

Funkcija	Paskirtis
NOT(loginė reikšmė)	Loginę reikšmę TRUE keičia į FALSE ir atvirkščiai.
AND(loginis argumentas1; loginis argumentas2;...)	Gražina reikšmę TRUE, jei visi argumentai yra TRUE. Jei bent vienas argumentas yra FALSE, funkcija gražina reikšmę FALSE.
OR(loginis argumentas1; loginis argumentas2;...)	Gražina reikšmę TRUE, jei bent vienas argumentas yra TRUE. Jei visi argumentai yra FALSE, funkcija gražina reikšmę FALSE.
IF(sąlygos tikrinimas; reikšmė, jei tiesa; reikšmė, jei netiesa)	Jei sąlyga tenkinama, gražinama reikšmė, jei tiesa. Jei sąlyga netenkinama, gražinama reikšmė, jei netiesa.
=IFS([kažkas yra True1;reikšmė, jei True1;kažkas yra True2;reikšmė, jei True2;kažkas yra True3;reikšmė, jei True3;...])	Funkcija IFS tikrina, ar tenkinama viena ar daugiau sąlygų, ir pateikia reikšmę, kuri atitinka pirmą TRUE (teisingą) sąlygą. IFS gali būti vykdoma vietoj kelių įdėtųjų IF sakinių, ir yra lengviau nuskaitoma, kai yra kelios sąlygos. IFS funkcija leidžia išbandyti iki 127 skirtingų sąlygų.

Datos ir laiko funkcijos

Funkcija	Paskirtis
TODAY()	Gražina šios dienos datą. Funkcija neturi argumentų.
DATE(metai;mėnuo;diena)	Gražina nuoseklų sekos skaičių, vaizduojantį konkrečią datą.
NOW()	Gražina dabartinės datos ir laiko sekos skaičių. Funkcija neturi argumentų.
DAY(argumentas datos formatu)	Gražina mėnesio dieną, skaičių nuo 1 iki 31.
MONTH(argumentas datos formatu)	Gražina mėnesį, skaičių nuo 1 iki 12.
YEAR(argumentas datos formatu)	Gražina metus, skaičių nuo 1900 iki 9999.

Teksto funkcijos

Funkcija	Paskirtis
FIND(ieškomas_tekstas; ieškos_tekstas; [prad_num])	Randa vieną teksto eilutę kitoje ir gražina pirmosios teksto eilutės pradinės padėties numerį nuo pirmojo antrosios teksto eilutės simbolio.
RIGHT(tekstas,[simbolių_skaičius])	Gražina paskutinį simbolių ar simbolių iš teksto eilutės, remdamasi nurodytu simbolių skaičiumi.
LEFT(tekstas, [simbolių_skaičius])	Gražina pirmąjį simbolių ar simbolių iš teksto eilutės, remdamasi nurodytu simbolių skaičiumi.

Duomenų sujungimas panaudojant ampersando (&) simbolį

1. Pažymėkite langelį, į kurį norite įdėti sujungtus duomenis.
2. Įveskite = ir pasirinkite pirmą langelį, kurį norite sujungti.
3. Įveskite & ir tarpe naudokite kabutes.
4. Pažymėkite kitą langelį, kurį norite sujungti, ir paspauskite "Enter". Formulės pavyzdys gali būti =A2&" "&B2.

Funkcija VLOOKUP

=VLOOKUP(Ko norite ieškoti, kur norite jo ieškoti, stulpelio numeris diapazone, kuriame yra gražintina reikšmė, pateikti apytikslį arba tikslų atitikmenį – nurodytas kaip 1/TRUE arba 0/FALSE).

Panaudojimas:

- kai reikia rasti ką nors lentelėje arba diapazone pagal eilutę;
- kai reikia sujungti kelias lenteles į vieną, jei vienoje iš lentelių yra bendrų laukų.

3.4. Nuorodos į praktines užduotis

Atliekant praktines užduotis rekomenduojama naudoti Microsoft Excel 2019, Microsoft Excel 365, internetines Google skaičiuokles.

Užduotys funkcijų, kurios buvo ankstesnėje programoje, mokymuisi ir įgūdžių įtvirtinimui

- A. Lozdienė, I. Mackevič. „Pasaulis kompiuteryje“ IX-X klasei (2 dalis) ir XI-XII klasei (1 dalis, [pradinių duomenų failai](#)). TEV (<https://tev.lt/it/>)
- T. Balvočienė, A. Balvočius. Informacinės technologijos. Bendrasis kursas. Vadovėlis XI-XII klasei, Šviesa, 2012 (<https://www.sviesa.lt/p/informacines-technologijos-bendrasis-kursas-vadovelis-xi-xii-klasei/>)
- J. Blonskis, V. Bukšnaitis, R. Burbaitė, D. Olbutienė, R. Šiaulienė. Informacinių technologijų valstybiniam brandos egzaminui. Užduočių pavyzdžiai, TEV, 2013. [Pradinių duomenų failai](#).
- D. Jančiauskienė. Užduotys savarankiškam darbui. Informacinės technologijos. Skyrius „Skaičiuoklės naudojimas“. Skaitmeninis išteklius. [Pradinių duomenų failai](#).
- Atvirasis Kauno technologijos universiteto kursas „Pasirengimas informacinių technologijų valstybiniam brandos egzaminui“, skyrius „Skaitinės informacijos apdorojimas skaičiuokle“ (norint pasiekti medžiagą ir užduotis, reikalinga registracija“. Nuoroda: <https://open.ktu.edu/course/view.php?id=75#section-5>
- IT VBE ankstesnių metų užduotys (skaičiuoklės dalis): <https://www.nsa.smm.lt/egzaminai-ir-pasiekimu-patikrinimai/brandos-egzaminai/egzaminu-uzduotys/>
- Papildomos užduotys: [9-10 klasės kartojimas](#), [skirtingi variantai](#).

Naujai įvestų funkcijų paaiškinimai, pavyzdžiai mokymuisi

1. ROUNDUP funkcija suapvalina skaičių iki didesnio skaičiaus (<https://support.microsoft.com/lt-lt/office/roundup-funkcija-roundup-f8bc9b23-e795-47db-8703-db171d0c42a7?ns=excel&version=19&syslcid=1063&uilcid=1063&appver=zx1190&helpid=xlmain11.chm60260&ui=lt-lt&rs=lt-lt&ad=lt>)

Formulė	Aprašas (rezultatas)	Rezultatas
=ROUNDUP(3,2;0)	Skaičius 3,2 apvalinamas iki didesnio skaičiaus be dešimtainės dalies.	4
=ROUNDUP(76,9;0)	Skaičius 76,9 apvalinamas iki didesnio skaičiaus be dešimtainės dalies.	77
=ROUNDUP(3,14159; 3)	Skaičius 3,14159 apvalinamas iki didesnio skaičiaus su trimis dešimtainėmis dalimis.	3,142
=ROUNDUP(-3,14159; 1)	Skaičių -3,14159 apvalina iki didesnio skaičiaus su vienu dešimtainiu skaitmeniu po kablelio.	-3,2
=ROUNDUP(31415,92654; -2)	Skaičius 31 415,92654 apvalinamas iki didesnio skaičiaus per du dešimtainius skaičius į kairę nuo dešimtainio skaičiaus kablelio.	31500

2. ROUNDDOWN funkcija suapvalina skaičių iki mažesnio skaičiaus
(<https://support.microsoft.com/lt-lt/office/funkcija-rounddown-2ec94c73-241f-4b01-8c6f-17e6d7968f53>)

Formulė	Aprašas	Rezultatas
=ROUNDDOWN(3,2; 0)	Skaičius 3,2 apvalinamas iki mažesnio skaičiaus be dešimtainės dalies.	3
=ROUNDDOWN(76,9;0)	Skaičius 76,9 apvalinamas iki mažesnio skaičiaus be dešimtainės dalies.	76
=ROUNDDOWN(3,14159; 3)	Skaičius 3,14159 apvalinamas iki mažesnio skaičiaus su trimis dešimtainėmis dalimis.	3,141
=ROUNDDOWN(-3,14159; 1)	Skaičius -3,14159 apvalinamas iki mažesnio skaičiaus su vienu dešimtainiu skaitmeniu po kablelio.	-3,1
=ROUNDDOWN(31415,92654; -2)	Skaičius 31 415,92654 apvalinamas iki mažesnio skaičiaus per du dešimtainius skaičius į kairę nuo dešimtainio skaičiaus kablelio.	31400

3. SUMIFS funkcija sudeda visus argumentus, atitinkančius keletą kriterijų
(<https://support.microsoft.com/lt-lt/office/funkcija-sumifs-c9e748f5-7ea7-455d-9406-611cebce642b>)

Argumento pavadinimas	Aprašas
Sumos_diapazonas (būtina)	Sumuojamų langelių diapazonas.
Kriterijų_diapazonas1 (būtina)	Diapazonas, kuris bandomas naudojant Kriterijai1 . Kriterijų_diapazonas1 ir Kriterijai1 nustato ieškos porą, tuo tarpu diapazone ieškoma tam tikrų kriterijų. Diapazone radus elementus, yra įtraukiamos jų atitinkamos reikšmės iš sumos_diapazonas .
Kriterijai1 (būtina)	Kriterijus, kuris apibrėžia, kurie langeliai iš Kriterijų_diapazonas1 bus įtraukti. Pavyzdžiui, kriterijai gali būti įvesti kaip 32 , ">32" , B4 , "obuoliai" arba "32" .
Kriterijų_diapazonas2, kriterijai2, ... (pasirinktinai)	Papildomi diapazonai ir jų susiję kriterijai. Galite įvesti iki 127 diapazonų / kriterijų porų.

Pavyzdys

(https://docs.google.com/spreadsheets/d/1mUtWUb4O_5A8cARYt1yMrBJvJ3YtfBfr/edit?usp=sharing&ouid=100641493981479752593&rtpof=true&sd=true)

Parduotas kiekis	Produktas	Pardavėjas
5	Obuoliai	Tomas
4	Obuoliai	Saulė
15	Kriaušės	Tomas
3	Kriaušės	Tomas
22	Bananai	Tomas
12	Bananai	Saulė
10	Morkos	Tomas
33	Morkos	Saulė
Suma už produktus, kurių pavadinimas prasideda K raide ir kuriuos pardavė Tomas	18	Langelyje B10 įrašyta: =SUMIFS(A2:A9; B2:B9; "=K*"; C2:C9; "Tomas")
Suma už produktus, kurių pavadinimas ne Bananai ir kuriuos pardavė Saulė	37	Langelyje B11 įrašyta: =SUMIFS(A2:A9; B2:B9; "<>Bananai"; C2:C9;"Saulė")
Suma už produktus, kurių kiekis ne didesnis už 10 ir kuriuos pardavė Tomas	18	Langelyje B12 įrašyta: =SUMIFS(A2:A9;A2:A9;"<=10";C2:C9;"Tomas")

4. AVERAGEIF funkcija grąžina visų diapazono langelių, atitinkančių nurodytus kriterijus, aritmetinį vidurkį.

Pavyzdys iš <https://support.microsoft.com/lt-lt/office/funkcija-averageif-faec8e2e-0dec-4308-af69-f5576d8ac642> pateiktas skaičiuoklėje
(https://docs.google.com/spreadsheets/d/1xBiesNEJb1J0orhEBL7M4bv2q_GiJkqs/edit?usp=sharing&ouid=100641493981479752593&rtpof=true&sd=true).

Nuosavybės vertė	Komisiniai	Skaičiavimai	Rezultatas	Rezultatas gaunamas užrašius formulę
100000	7000	Visų komisinių, mažesnių nei 23000, vidurkis.	14000	=AVERAGEIF(B2:B5;"< 23 000")
200000	14000	Visų nuosavybės verčių, mažesnių nei 250000, vidurkis.	150000	=AVERAGEIF(A2:A5;"< 250000")
300000	21000	Komisinių, kai nuosavybės vertė mažesnė už 300000, vidurkis.	10500	=AVERAGEIF(A2:A5;"<300000";B2:B5)
400000	28000	Nuosavybės verčių, kai komisiniai didesni už 20000, vidurkis	350000	=AVERAGEIF(B2:B5;">20000";A2:A5)

5. AVERAGEIFS grąžina apskaičiuotą visų langelių, atitinkančių nurodytus kriterijus, aritmetinį vidurkį.

Pavyzdyje (<https://docs.google.com/spreadsheets/d/1L963YuffpOUjAS377YSrjS6Fv-FXpsNk/edit?usp=sharing&oid=100641493981479752593&rtpof=true&sd=true>) langeliuose įrašytos formulės:

E2: =AVERAGEIFS(C2:C121;A2:A121;"=1";B2:B121;">10")

E3: =AVERAGEIFS(C2:C121;A2:A121;">2";A2:A121;"<5")

	A	B	C	D	E
1	Diena	Valanda	Megavatvalandės kaina, eurai	Skaičiavimai	Rezultatas
2	1	0	95,75	Pirmos dienos vidutinė megavatvalandės kaina, kai valanda > 10	48,74
3	1	1	134,85	Trečios ir ketvirtos dienų vidutinė megavatvalandės kaina	50,38
4	1	2	132,53		

6. MINIFS pateikia mažiausią langelių, atitinkančių tam tikrą sąlygų ar kriterijų rinkinį, reikšmę.

7. Funkcija MAXIFS pateikia didžiausią langelių, atitinkančių tam tikrą sąlygų ar kriterijų rinkinį, reikšmę.

Pavyzdyje

(https://docs.google.com/spreadsheets/d/1wxnUDBPBGdZZouYMNSOpHR0_T8f3lCsB/edit?usp=sharing&oid=100641493981479752593&rtpof=true&sd=true) pateiktas funkcijų panaudojimas. Langeliuose įrašytos formulės:

E2: =MINIFS(C2:C121;A2:A121;"=1";B2:B121;">=5";B2:B121;"<=12")

E3: =MAXIFS(C2:C121;A2:A121;"=1";B2:B121;">=5";B2:B121;"<=12")

E4: =MINIFS(C2:C121;A2:A121;">=1";A2:A121;"<=4";B2:B121;">=12";B2:B121;"<=20")

E5: =MAXIFS(C2:C121;A2:A121;">=1";A2:A121;"<=4";B2:B121;">=12";B2:B121;"<=20")

	A	B	C	D	E
1	Diena	Valanda	Megavatvalandės kaina, eurai	Skaičiavimai	Rezultatas
2	1	0	95,75	Pirmos dienos mažiausia megavatvalandės kaina valandų intervale [5;12]	19,77
3	1	1	134,85	Pirmos dienos didžiausia megavatvalandės kaina valandų intervale [5;12]	111,74
4	1	2	132,53	Pirmos-ketvirtos dienų mažiausia megavatvalandės kaina valandų intervale [12;20]	-0,08
5	1	3	84,69	Pirmos-ketvirtos dienų didžiausia megavatvalandės kaina valandų intervale [12;20]	166,41
6	1	4	19,96		
7	1	5	19,77		
8	1	6	84,69		
9	1	7	111,74		
10	1	8	111,21		

- COUNTA skaičiuoja langelius, kuriuose yra bet kokio tipo informacijos, įskaitant klaidų reikšmes ir tuščią tekstą (""). Duomenų tyryboje funkcija naudojama tuomet, kai reikia suskaičiuoti, kiek tuščių langelių yra tam tikrame duomenų bloke (blokuose).
- COUNTIFS kriterijus taiko langeliams keliuose diapazonuose ir skaičiuoja, kiek kartų visi kriterijai yra įvykdyti.

Pavyzdžiuose (<https://support.microsoft.com/lt-lt/office/countifs-funkcija-countifs-dda3dc6e-f74e-4aee-88bc-aa8c2a866842>) pateikti funkcijos panaudojimo atvejai (<https://docs.google.com/spreadsheets/d/1mIWQqGcuiThqao27lvzVyKMQfw4zcZcm/edit?usp=sharing&ouid=100641493981479752593&rtpof=true&sd=true>).

	A	B	C	D
	Pardavėjas	Viršija 1 ktv. kvotą	Viršija 2 ktv. kvotą	Viršija 3 ktv. kvotą
1				
2	Paulaitis	Taip	Ne	Ne
3	Antanaitis	Taip	Taip	Ne
4	Jurgaitis	Taip	Taip	Taip
5	Juozaitis	Ne	Taip	Taip
6	Apskaičiuoja, kiek kartų Paulaitis viršijo pardavimo kvotą laikotarpiais 1ktv., 2 ktv. ir 3 ktv. (tik 1 ktv.).	1		
7	Apskaičiuoja, kiek pardavėjų viršijo tiek 1 ktv., tiek 2 ktv. kvotas (Antanaitis ir Jurgaitis).	2		
8	Apskaičiuoja, kiek kartų Juozaitis ir Antanaitis viršijo tą pačią kvotą laikotarpiais 1ktv., 2 ktv. ir 3 ktv. (tik 2 ktv.).	1		

B6 langelyje įrašyta: =COUNTIFS(B2:D2;"=Taip")

B7 langelyje įrašyta: =COUNTIFS(B2:B5;"=Taip";C2:C5;"=Taip")

B8 langelyje įrašyta: =COUNTIFS(B5:D5;"=Taip";B3:D3;"=Taip")

	A	B
1	Duomenys	Data
2		1 2011-05-01
3		2 2011-05-02
4		3 2011-05-03
5		4 2011-05-04
6		5 2011-05-05
7		6 2011-05-06
8	Apskaičiuoja, kiek skaičių nuo 1 iki 6 (neįtraukiant 1 ir 6) yra langeliuose nuo A2 iki A7.	4
9	Apskaičiuoja, keliose eilutėse langeliuose nuo A2 iki A7 yra skaičiai, mažesni nei 5, bei keliose eilutėse langeliuose nuo B2 iki B7 yra datos, ankstesnės nei 2011-05-03.	2
10	Apskaičiuoja, keliose eilutėse langeliuose nuo A2 iki A7 yra skaičiai, mažesni nei 5, bei keliose eilutėse langeliuose nuo B2 iki B7 yra datos, ankstesnės nei 2011 05 03.	2

B8: =COUNTIFS(A2:A7;"< 6";A2:A7;"> 1")

B9: =COUNTIFS(A2:A7; "< 5";B2:B7;"<"&B4)

B10: =COUNTIFS(A2:A7; "< 5";B2:B7;"<2011-05-03")

10. MODE grąžina dažniausiai pasitaikančią arba pasikartojančią reikšmę nurodytame duomenų diapazone (<https://support.microsoft.com/lt-lt/office/mode-funkcija-mode-e45192ce-9122-4980-82ed-4bdc34973120>).

Duomenys		
5,6		
4		
4		
3		
2		
4		
Formulė	Aprašas	Rezultatas
=MODE(A2:A7)	Moda, arba dažniausiai pasitaikantis iš aukščiau pateiktų skaičių	4

11. MEDIAN gražina duotų skaičių medianą. Mediana yra skaičius skaičių aibės viduryje (<https://support.microsoft.com/lt-lt/office/median-funkcija-median-d0916313-4753-414c-8537-ce85bdd967d2>).

Duomenys		
1		
2		
3		
4		
5		
6		
Formulė	Aprašas	Rezultatas
=MEDIAN(A2:A6)	Penkių skaičių iš diapazono A2:A6 mediana. Kadangi reikšmės yra penkios, mediana yra trečioji.	3
=MEDIAN(A2:A7)	Penkių skaičių iš diapazono A2:A6 mediana. Kadangi reikšmės yra šešios, mediana yra tarpinis skaičius tarp trečiojo ir ketvirtojo skaičių.	3,5

12. STDEV skaičiuoja standartinę imties nuokrypį. Standartinis nuokrypis rodo, kiek daug reikšmės skiriasi nuo vidurkio (vidutinės reikšmės).

Pavyzdžio idėja iš knygos: V. Čekanavičius, G. Murauskas „Statistika ir jos taikymai 1“, TEV, 2000. Pavyzdžio failas:

https://docs.google.com/spreadsheets/d/1Fvun5LT7QXn44suqXl_TrPgQuAcgQ2Qb/edit?usp=s_haring&ouid=100641493981479752593&rtpof=true&sd=true

	A	B	C
		Pirmosios firmos programuotojų atlyginimai	Antrosios firmos programuotojų atlyginimai
1			
2		3000	5000
3		3000	5000
4		4000	5000
5		6000	5000
6		9000	5000
7	Vidutinis atlyginimas	5000	5000
8	Standartinis nuokrypis	2280,35085	0

13. SUBTOTAL pateikia sąrašo tarpinę sumą, vidurkį, kiekį, mažiausią, didžiausią reikšmę, vidutinį kvadratinį nuokrypį. Funkcijos pirmųjų argumentų lentelė:

funkcijos_nr (įskaitant paslėptas eilutes)	funkcijos_nr (nepaiso paslėptų eilučių)	Funkcija
---	--	----------

funkcijos_nr (įskaitant paslėptas eilutes)	funkcijos_nr (nepaiso paslėptų eilučių)	Funkcija
1	101	AVERAGE
2	102	COUNT
3	103	COUNTA
4	104	MAX
5	105	MIN
7	107	STDEV
9	109	SUM

Pavyzdyje

(https://docs.google.com/spreadsheets/d/1_EK1a1URZYRJPU4K36dAKrutZaBsFpvK/edit?usp=sharing&ouid=100641493981479752593&rtopof=true&sd=true) pateiktos 2023 m. liepos mėnesio pirmųjų penkių dienų elektros energijos kainos kiekvieną valandą. Skaičiuoklės lentelės fragmentas su nefiltruotais duomenimis:

Die- na	Val an- da	Mega- vatvalan- dės kaina, eurais	Skaičiavimai	Rezul- tatas	Rezultatas gaunamas užrašius formulę
1	0	95,75	Pirmos dienos vidutinė megavatvalandės kaina	66,35	=SUBTOTAL(1;C2:C25)
1	1	134,85	Antros dienos mažiausia megavatvalandės kaina	-0,10	=SUBTOTAL(5;C26:C49)
1	2	132,53	Trečios dienos didžiausia megavatvalandės kaina	84,40	=SUBTOTAL(4;C50:C73)
1	3	84,69	Pirmos ir penktos dienų vidutinė megavatvalandės kaina	79,14	=SUBTOTAL(1;C2:C25;C98:C121)
1	4	19,96			
1	5	19,77			
1	6	84,69			
1	7	111,74			
1	8	111,21			
1	9	99,06			
1	10	64,51			
1	11	43,8			
1	12	24,61			
1	13	16,83			

1	14	4,43			
---	----	------	--	--	--

Išfiltravus septintą kiekvienos dienos valandą:

Die- na	Valan- da	Mega- vatva- landės kaina, eurais	Skaičia- vimai (7 valanda)	Rezulta- tas	Rezultatas gaunamas užrašius formulę
1	7	111,74	Vidutinė megavatva- landės kaina	70,56	=SUBTOTAL(101;C9:C121)
2	7	-0,02	Mažiausia megavatva- landės kaina	-0,02	=SUBTOTAL(105;C2:C121)
3	7	39,94	Didžiausia megavatva- landės kaina	116,09	=SUBTOTAL(104;C2:C121)
4	7	116,09			
5	7	85,06			

Išfiltravus pirmos ir penktos dienos duomenis ir apskaičiavus pirmos ir penktos dienų vidutinę elektros energijos megavatvalandės kainą, gaunamas toks pat rezultatas, kaip ir apskaičiavus nefiltruotame sąraše. Išfiltruotame sąraše skaičiavimams panaudota formulė:

=SUBTOTAL(101;C2:C121)

A	B	C	D	E
Diena	Valanda	Megavatvalandės kaina, eurais	Skaiciavimai	Rezultatas
1	7	95,75	Pirmos ir penktos dienų vidutinė megavatvalandės kaina	79,14
2	7	34,85		
3	7	32,53		
4	7	84,69		
5	7	19,96		
6	7	19,77		
7	7	84,69		
8	7	11,74		
9	7	11,21		
10	7	99,06		
11	7	64,51		
12	7	43,8		
13	7	24,61		
14	7	16,83		
15	7	4,43		
16	7	0,07		
17	7	0,99		
18	7	12,67		
19	7	52,4		
20	7	67,88		
21	7	84,7		

14. IFS tikrina, ar tenkinama viena ar daugiau sąlygų. Gali būti vykdoma vietoj kelių įdėtųjų IF sakinių.

Pavyzdys (https://docs.google.com/spreadsheets/d/1bOamhVnB-Hw0ngJUzaKSmiw-K6gD_CmV/edit?usp=sharing&oid=100641493981479752593&rtpof=true&sd=true)

	A	B	C	D	E	F
1	Vardas	Pažymys	Pasiekimų lygis		Pažymys	Pasiekimų lygis
2	Romas	5	Patenkinamas		1	Mokytis ir papildomai atsiskaityti
3	Tomas	2	Mokytis ir papildomai atsiskaityti		2	Mokytis ir papildomai atsiskaityti
4	Rimas	3	Mokytis ir papildomai atsiskaityti		3	Mokytis ir papildomai atsiskaityti
5	Simas	4	Slenkstinis		4	Slenkstinis
6	Pranas	5	Patenkinamas		5	Patenkinamas
7	Antanas	1	Mokytis ir papildomai atsiskaityti		6	Patenkinamas
8	Irena	6	Patenkinamas		7	Pagrindinis
9	Janina	7	Pagrindinis		8	Pagrindinis
10	Regina	8	Pagrindinis		9	Aukštesnysis
11	Rita	7	Pagrindinis		10	Aukštesnysis
12	Rima	9	Aukštesnysis			
13	Roma	7	Pagrindinis			
14	Toma	10	Aukštesnysis			
15						

Langelyje C2 įrašyta:

=IFS(B2=\$E\$2;\$F\$2;B2=\$E\$3;\$F\$3;B2=\$E\$4;\$F\$4;B2=\$E\$5;\$F\$5;B2=\$E\$6;\$F\$6;B2=\$E\$7;\$F\$7;B2=\$E\$8;\$F\$8;B2=\$E\$9;\$F\$9;B2=\$E\$10;\$F\$10;B2=\$E\$11;\$F\$11)

15. FIND randa vieną teksto eilutę kitoje ir grąžina pirmosios teksto eilutės pradinės padėties numerį nuo pirmojo antrosios teksto eilutės simbolio.

Pavyzdys

(https://docs.google.com/spreadsheets/d/19BPhOmcW3tRNF1a7mYmAxOSkxNvE_KH/edit?usp=sharing&oid=100641493981479752593&rtpof=true&sd=true)

	A	B	C
1	Vardas, pavardė	Rezultatas1	Rezultatas2
2	Kiškis Kiškelis	1	8
3	Romas Romelis	2	8

B2: =FIND("Kišk";A2)

C2: =FIND("Kišk";A2;5)

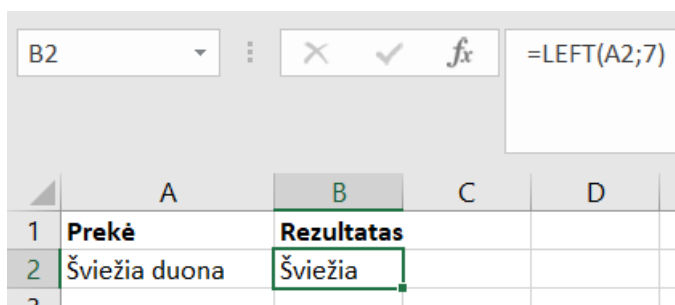
B3: =FIND("om";A3)

C3: =FIND("om";A3;4)

16. RIGHT grąžina paskutinį simbolį ar simbolius iš teksto eilutės, remdamasi nurodytu simbolių skaičiumi.

B2	=RIGHT(A2;5)			
	A	B	C	D
1	Prekė	Rezultatas		
2	Šviežia duona	duona		

17. LEFT grąžina pirmąjį simbolį ar simbolius iš teksto eilutės, remdamasi nurodytu simbolių skaičiumi.



RIGHT ir LEFT pavyzdžiai:

<https://docs.google.com/spreadsheets/d/1QKwY6HjdeFikievOozMdDIslR0l2R0fM/edit?usp=sharing&ouid=100641493981479752593&rtpof=true&sd=true>

18. VLOOKUP funkciją aiškinanti pamoka su pradinių duomenų failais:

<https://www.timeatlas.com/vlookup-tutorial/>

19. Suvestinės lentelės. Lentelių jungimas panaudojant VLOOKUP. Sukinio lentelės ir diagramos.

Skaidrės:

https://docs.google.com/presentation/d/1bfsIjhHdxjq_iBCRTGqyO0D3spH7Ufy/edit?usp=sharing&ouid=100641493981479752593&rtpof=true&sd=true

Pradinių duomenų failai:

https://drive.google.com/file/d/1z0T4tF_bJlmeE8KntkapzAvwCfJZbYdhp/view?usp=sharing

20. Skaitinių duomenų apdorojimas ir vizualizavimas

https://docs.google.com/presentation/d/1Ugx4X-DcEKYB4gehImxhB_JxbWEPiL7C/edit?usp=sharing&ouid=100641493981479752593&rtpof=true&sd=true

21. Sąlyginis formatavimas, rūšiavimas, filtravimas, sukinio lentelės, sudėtingesnės diagramos

Skaidrės:

https://docs.google.com/presentation/d/1GTzy_4Lg8skWVjJDHGJtZ3mrS3VxVrdA/edit?usp=sharing&ouid=100641493981479752593&rtpof=true&sd=true

Pradinių duomenų failai:

<https://drive.google.com/file/d/1dKdDW1OzPeQH7EVGGVCYNHagGevLSHE3/view?usp=sharing>

Šaltiniai ir idėjos užduotims ir projektams

- V. Šakys. Veiklos analitika su Microsoft Excel Power Pivot ir Power BI. Kaunas, Vitae Litera, 2020.
- V. Šakys. Veiklos analitikos praktikumas. Kaunas, Vitae Litera, 2021.
- <https://excelexercises.com/lessons.html> - pagrindinių funkcijų naudojimo įgūdžiams įgyti skirtos pamokos. Labai tinka treniravimuisi prieš tarpinį patikrinimą.
- <https://www.w3schools.com/excel/index.php>

- https://www.w3schools.com/excel/excel_functions.php
- Trigonometrinių funkcijų panaudojimas temperatūros prognozavimui: https://www.khanacademy.org/math/algebra2/x2ec2f6f830c9fb89:trig/x2ec2f6f830c9fb89:sinusoidal-models/v/modeling-temperature-fluctuations?fbclid=IwAR3SA1ljLjpbE5_RklEbovJBekx8VusiJE9LQqEaxS2S-jyp1o232LpRAv0
- https://www.w3schools.com/excel/excel_exercises.php
- <https://edu.gcglobal.org/en/excelformulas/excel-formulas-quiz/1/>
- <https://study.com/academy/practice/quiz-worksheet-using-excel-formulas-functions.html>
- <https://www.propofs.com/quiz-school/story.php?title=excel-formulas-functions>
- <https://study.com/academy/practice/quiz-worksheet-if-function-in-excel.html>
- <https://www.excel-demy.com/top-100-mcq-of-excel/>
- <https://www.excel-exercise.com/excel-quizzes/>
- <https://www.ncertbooks.guru/mcq-questions-on-ms-excel/>
- <https://www.automateexcel.com/tutorials-practice-tests/>
- <https://exceljet.net/formulas>

Chat GPT pasiūlymai (kalba netaisyta)

Skaičiuoklės absoliutinės reikšmės (modulio) funkcijos ABS taikymo galimybės duomenų tyrtyboje vidurinėje mokykloje sprendžiant realaus pasaulio uždavinius

ChatGPT

Žinoma, ABS funkcija gali būti taikoma vidurinės mokyklos lygmeniu, kad mokiniai galėtų išspręsti realaus pasaulio uždavinius, susijusius su duomenų tyrimu. Tai padeda jiems įgyti praktinių įgūdžių analizuoti, interpretuoti ir manipuliuoti skaičiais bei duomenimis. Štai keli pavyzdžiai, kaip ABS funkcija gali būti naudojama sprendžiant realaus pasaulio uždavinius:

1. Atstumų analizė: Mokiniai gali išspręsti uždavinį, kuriame reikia apskaičiuoti atstumą tarp dviejų skirtingų vietų ar objektų. ABS funkcija padės jiems ignoruoti atstumo kryptį ir susikoncentruoti tik į absoliučią reikšmę.
2. Laiko analizė: Mokiniai gali analizuoti laiko skirtumus tarp dviejų įvykių arba laikotarpių. ABS funkcija gali padėti nustatyti, kiek laiko praėjo nepriklausomai nuo to, ar tai buvo laikas praeities ar ateities.
3. Skaičių skirtumai: Mokiniai gali sprendžiant uždavinį apie skaičių skirtumus, pavyzdžiui, temperatūros pokyčius per kelias dienas. ABS funkcija padės jiems įvertinti, kiek reikšmės iš tikrųjų pasikeitė.
4. Vidurkio ir medianos analizė: Mokiniai gali būti priversti analizuoti duomenų rinkinį ir apskaičiuoti vidurkį ar medianą. ABS funkcija padėtų jiems ignoruoti skaičių ženklus ir apskaičiuoti absoliučią vidurkio nuokrypį.
5. Matematiniai modeliai: Mokiniai gali būti iššūkiami sukurti matematinį modelį, kuris paaiškintų realaus pasaulio situaciją. ABS funkcija galėtų būti naudojama modelyje, kuriame vertinama absoliuti reikšmė ar skirtumas.
6. Sprendimų priėmimas: Mokiniai gali analizuoti situacijas, kur reikia pasirinkti tarp dviejų galimų variantų. ABS funkcija padės jiems nustatyti, kuris variantas yra artimesnis absoliučiu reikšmių skirtumu.
7. Netolydžių reikšmių analizė: Mokiniai gali analizuoti netolydžias reikšmes, pavyzdžiui, objekto greičio pasikeitimus laike. ABS funkcija padės suvokti, kaip skaičiai gali keistis netolygiai ir kokie yra absoliutūs pokyčiai.

Taikant ABS funkciją realaus pasaulio uždaviniuose, mokiniai gali įgyti pritaikomų įgūdžių, kurie padės jiems geriau suprasti ir analizuoti skaičius bei duomenis. Tai taip pat suteiks jiems gebėjimą vertinti reikšmes nepriklausomai nuo ženklo ir naudoti matematiką sprendžiant praktinius uždavinius.

Skaičiuoklės funkcijų ROUND, ROUNDUP, ROUNDDOWN taikymo galimybės duomenų tyryboje vidurinėje mokykloje sprendžiant realaus pasaulio uždavinius

ChatGPT

Skaičiuoklėse funkcijos ROUND, ROUNDUP ir ROUNDDOWN gali būti labai naudingos vidurinės mokyklos mokiniams, kai jie sprendžia realaus pasaulio uždavinius, susijusius su skaičių apvalinimu ir tikslumu. Šios funkcijos padeda tvarkingai suapvalinti skaičius ir atlieka apvalinimo operacijas, kurios yra labai svarbios tiksliai analizei. Štai keletas galimybių, kaip šios funkcijos gali būti taikomos duomenų tyryboje vidurinėje mokykloje:

1. Finansinė analizė: Mokiniai gali išmokti naudoti ROUND funkciją, kad būtų galima apvalinti finansinius skaičius, pvz., pinigų sumas ar palūkanas. Tai padėtų tinkamai apvalinti skaičius pagal finansinės analizės reikalavimus.
2. Matavimai ir vienetai: Mokiniai gali taikyti ROUND funkciją apvalinant matavimus ir kitus dydžius, kad būtų pasiektas norimas tikslumas. Tai labai svarbu, kai analizuojami matavimai, pvz., ilgis, masė, plotas ar tūris.
3. Laiko analizė: ROUND funkcija gali būti naudojama apvalinant laiko reikšmes. Pavyzdžiui, mokiniai gali analizuoti laiko trukmę ir apvalinti laiko intervalus arba laiko tarpus.
4. Statistika: Mokiniai gali mokytis naudoti ROUND, ROUNDUP ir ROUNDDOWN funkcijas analizuojant statistinius rodiklius. Tai padėtų pasiekti tinkamą tikslumą, kai apvalinami vidurkiai, medianos, kvartilai ir kitos statistikos reikšmės.
5. Projektų planavimas: Mokiniai gali taikyti šias funkcijas apvalindami skaičius, susijusius su projektų planavimu ir trukme. Tai padėtų planuojant užduotis ir resursus pagal tikslų tikslumą.
6. Prekyba ir nuolaidos: Mokiniai gali išmokti naudoti ROUNDUP ir ROUNDDOWN funkcijas, sprendžiant uždavinius apie prekių kainas ir nuolaidas. Tai padėtų apvalinti kainas pagal prekybinės praktikos reikalavimus.
7. Mokėjimai ir skaičiavimai su mokesčiais: Apvalinimo funkcijos gali būti taikomos skaičiuojant mokėjimus ir apmokestinimą. Tai padėtų gauti tikslias mokėjimo sumas arba apmokestinimo reikšmes.

Šios funkcijos suteikia mokiniams gebėjimų tvarkingai ir tiksliai tvarkyti skaičius, atliekant būtinus apvalinimus ir tinkamai analizuojant duomenis. Tai yra svarbus įgūdis, reikalingas siekiant tikslumo ir tikslaus sprendimų priėmimo realaus pasaulio uždaviniais.