

Duomenų tyryba (patikra 1 (iš 4 – 5 - 6))

Duomenys gali būti paveikslėliai, skaičiai, tekstas be konteksto

Kai duomenys tampa reikšmingais gavėjui, tai vadinama informacija.

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.

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 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.