**Daugiamačių duomenų vizualizacija**

**1 laboratorinio darbo ataskaita**

**Ignas Jatulis 1815628**

**Pirmoji užduotis:**

*Pasirinkite daugiamačių duomenų rinkinį - iš magistro tyrimų, darbo, pomėgių ar Interneto (pvz.* [*UCI Machine Learning Repository*](http://archive.ics.uci.edu/ml/)*). Aprašykite duomenų rinkinį: prasmė, objektų ir parametrų skaičius bei savybės.*

**Sprendimas:**

**Pasirinktas duomenų šaltinis:**

[Boston Housing Data](https://archive.ics.uci.edu/ml/machine-learning-databases/housing/)

**Prasmė:**

Šiame duomenų rinkinyje pateikiama JAV surašymo tarnybos surinkta informacija apie būstą Bostono rajonuose. Ji buvo gauta iš StatLib archyvo (http://lib.stat.cmu.edu/datasets/boston) .

Šis duomenų rinkinys buvo naudojamas namų kainos prognozavimui tūkstančiais dolerių, pateikiant informaciją apie namą ir jo apylinkes.

Duomenis paskelbė: Harrison, D. and Rubinfeld, D.L. `Hedonic prices and the demand for clean air', J. Environ. Economics & Management, vol.5, 81-102, 1978.

**Objektų ir parametrų skaičius:**

Įrašų kiekis: 506.

Atributai: 13 skaitinių atribūtų (įskaitant ir „klasės“ atributą „MEDV“), vienas binarinis atributas. Iš viso 14 atributų, trūkstamų atributų reikšmių šiame rinkinyje nėra.

**Savybės:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Nr.** | **Pavadinimas** | **Reikšmė** | **Tipas** |
| 1. | CRIM | Nusikaltymų skaičius vienam gyventojui. | Realusis skaičius |
| 2. | ZN | Proporcija gyvenamajai žemei, skirta zonoms, kuriose yra daugiau kaip 25 000 kv. | Realusis skaičius |
| 3. | INDUS | Ne mažmeninės prekybos akrų skaičius mieste | Realusis skaičius |
| 4. | CHAS | Charles River (upės) požymis (= 1 jei ribojasi su upe; 0 kitu atveju). | Binarinė reikšmė |
| 5. | NOX | Azoto oksidų koncentracija (10-mis milijonų dalių). | Realusis skaičius |
| 6. | RM | Vidutinis kambarių skaičius vienam būstui. | Realusis skaičius |
| 7. | AGE | Savininkų naudojamų būstų dalis, pastatyta iki 1940 m. | Realusis skaičius |
| 8. | DIS | Svertinis atstumas iki penkių Bostono įdarbinimo centrų. | Realusis skaičius |
| 9. | RAD | Greitkelių prieinamumo indeksas. | Sveikasis skaičius |
| 10. | TAX | Visos vertės turto mokesčio tarifas 10 000-ių dolerių. | Realusis skaičius |
| 11. | PTRATIO | Mokinių ir mokytojų santykis. | Realusis skaičius |
| 12. | B | 1000 (Bk - 0,63) ^ 2, kur Bk yra juodaodžių dalis miesto dalyje. | Realusis skaičius |
| 13. | LSTAT | Procentas, kiek gyventojų sudaro žemo statuso gyventojai. | Realusis skaičius |
| 14. | MEDV | Vidutinė savininkų namų vertė 1000-iais dolerių. | Realusis skaičius |

**Duomenų rinkinio aprašymas anglų kalba:** <https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html>