

- Data.gov. Репозиторій відкритих наборів даних від урядових установ США з таких тем, як клімат, споживачі, освіта, екосистеми, енергетика, фінанси, виробництво, наука.

Серед великого різноманіття джерел датасетів також можна виділити українські, хоча їх невелика кількість, але вони користуються популярністю. Одним із найвідоміших українських джерел наборів даних є Єдиний державний веб-портал відкритих даних [5]. Він містить 35 тисяч наборів даних, що розбиті на групи. Будь-яка особа може вільно копіювати, публікувати, поширювати, використовувати їх, у тому числі в комерційних цілях, у поєднанні з іншою інформацією або шляхом включення до складу власного продукту, публічну інформацію у формі відкритих даних з обов'язковим посиланням на джерело отримання такої інформації. Структура набору відкритих даних оприлюднюється у форматах XSD, JSON, CSV або інших аналогічних форматах.

Підводячи підсумок, можна сказати, що завдяки стрімкому розвитку інформаційних технологій створюється все більша кількість наборів даних, що провокує збільшення джерел для їх зберігання. Набори даних все частіше використовують для проведення data-driven досліджень, задоволення інтересу до тенденцій в певних сферах нашого життя та у вирішенні низки проблем у різних областях.

Список літературних джерел

1. Neskorodieva T., Fedorov E. Method of spectral clustering of payments and raw materials supply for the compliance audit planning. *Радіоелектроніка, інформатика, управління*. 2021. №1. pp.127-135. doi: 10/15588/1607-3274-2021-1-13
2. Vunivere.ru [Електронний ресурс]. Режим доступу – <https://vunivere.ru/work35003>
3. Loc.gov[Електронний ресурс]. Режим доступу – https://www.loc.gov/preservation/digital/formats/fdd/dataset_fdd.shtml
4. Nuancesprog.ru [Електронний ресурс]. Режим доступу – <https://nuancesprog.ru/p/4258/>
5. Gov.ua [Електронний ресурс]. Режим доступу – <https://data.gov.ua/en/pages/about>

УДК 004.82:004:85

*Грущенко В. Ю., студент 1 курсу
СО Магістр*

*Нескородєва Т. В., канд. техн. наук, доцент,
завідувач кафедри комп'ютерних наук та
інформаційних технологій*

АНАЛІЗ ДАНИХ РИНКУ ПЕРСОНАЛЬНИХ КОМП'ЮТЕРІВ

Донецький національний університет імені Василя Стуса, м. Вінниця

Аналіз даних про ринок персональних комп'ютерів, дозволяє ознайомитись з тим, які пропозиції на ньому переважають, та зробити висновок

які фактори впливають на формування ціни.

Актуальність полягає в тому, що пропозицій на ринку комп'ютерів багато, та виявити закономірність для прийняття рішень щодо покупки або встановлення ціни неможливо без аналізу великих обсягів даних [1,2].

Розглянемо приклад такого аналізу на основі даних ринку персональних комп'ютерів. Ми використовували набір даних Computers [3], що містить ціни на персональні комп'ютери у розрізі 1993 по 1995 роки. Нижче наводиться опис 5 основних показників, що були відібрані для аналізу з 11 змінних. В якості відгука був обраний показник «price», та створений якісний показник бюджетних та небюджетних моделей «price01»:

- price – вартість комп'ютера (представлена у доларах США \$).
- speed – тактова частота в МГц .
- hd – розмір жорсткого диска в МБ.
- ram – розмір оперативної пам'яті в МБ.
- screen – розмір екрану в дюймах.

Об'єктами спостереження є комп'ютери, що пропонувались на ринку за даний період. На першому етапі було отримано описову статистику для кожного з показників та матрицю розсіювань для змінних набору даних «Computers» (рис 1).

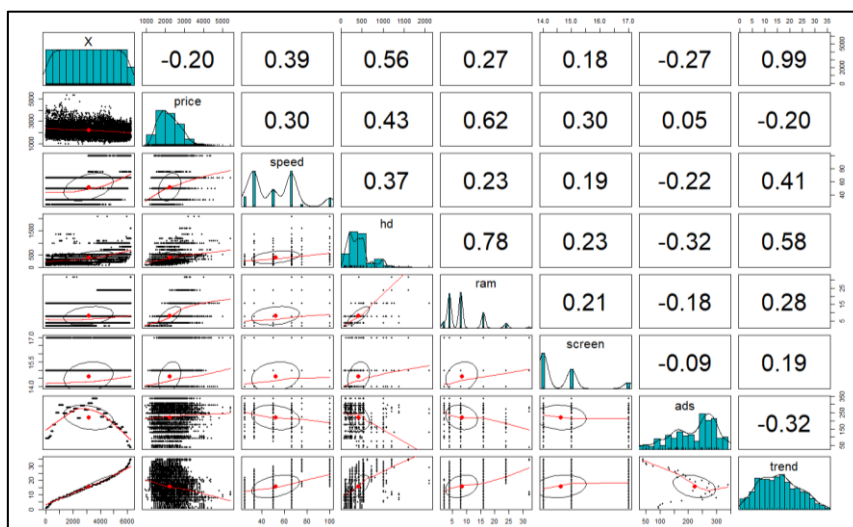


Рисунок 1 – Побудова матриць діаграм розсіювання

Наступним кроком необхідно визначити кількість бюджетних моделей комп'ютерів, що представлені на ринку. Орієнтуючись на показник медіани створимо змінну «price01» та виведемо кількість бюджетних комп'ютерів, що менші за медіану.

Отже на ринку було представлено дещо більше моделей небюджетного характеру, але їх різниця була не надто великою, що дозволяло обрати модель відповідно до матеріальних можливостей.

Наступним етапом будуть створені: LDA модель [4], QDA модель, GLM модель та KNN модель (Рисунок 2).

```

Coefficients of linear discriminants:
LD1
speed  0.0156870022
hd     -0.0009666769
ram     0.2343060844
screen  0.0757581542
ads     0.0054987741
> pred.lda <- predict(fit.lda, Computers.test)
> table(pred.lda$class, price01.test)
      price01.test
      0      1
0 1592  566
1  283 1061
> mean(pred.lda$class != price01.test)
[1] 0.2424329

> library(class)
> train.X <- cbind(speed, hd, screen, ads)[train, ]
> test.X <- cbind(speed, hd, screen, ads)[!train, ]
> train.price01 <- price01[train]
> set.seed(1)
> pred.knn <- knn(train.X, test.X, train.price01, k = 1)
> table(pred.knn, price01.test)
      price01.test
pred.knn  0      1
0 1369  215
1  506 1412
> mean(pred.knn != price01.test)
[1] 0.2058824

```

Рисунок 2 – Звіти по LDA та KNN моделях (в якості відгука price01)

Точність побудованих моделей наведена нижче:

LDA – 24 %.

QDA – 22 %.

GLM – 25 %.

KNN (при K (кількість найближчих сусідів)) = 1 – 20%.

KNN (10) – 24%.

KNN (20) – 25%.

KNN (50) – 26%.

KNN (100) – 29%.

Отже найбільш точною є модель KNN при K = 1.

Висновок. Отже використання даних моделей дає можливість виявити фактори, які впливають на ціну та класифікувати комп'ютери по категоріям бюджетний/небюджетний в залежності від його технічних характеристик.

Список літературних джерел

1. Кузьома Ю.В., Нескородева Т.В. Розробка інтернет-магазину з продажу деталей та аксесуарів для мобільних телефонів. Матеріали всеукраїнської науково-практичної конференції для студентів, аспірантів та молодих вчених "Прикладні інформаційні технології" (29 квітня 2020 року) - Вінниця: ДонНУ імені Василя Стуса. С.217-219.
2. Сніжинський М.В., Нескородева Т.В. Аналіз даних про рейтинг фільмів на платформі оцінювання IMDb засобами мови R. Матеріали всеукраїнської науково-практичної конференції для студентів, аспірантів та молодих вчених "Прикладні інформаційні технології" (29 квітня 2020 року) - Вінниця: ДонНУ імені Василя Стуса. С.57- 58.
3. «Computers» статистика – [Електронний ресурс]. Режим доступу: <http://qed.econ.queensu.ca/jae/>
4. Джеймс Г., Уиттон Д., Хасті Т., Тибширани Р. Введение в статистическое обучение с примерами на языке R. Пер. с англ. С. Э. Мастицкого - М.: ДМК Пресс, 2017. - 456 с.: ил.