DWM PARTICAL NO 5

Assignment No 5: Demonstrate performing Regression on data sets

house_price.csv file

Note: house_price.csv file is made in Mysql Database and exported to weka in csv format

Algorithm: In Classify "Choose" *→ classifier → functions →LinearRegression*

Code:

```
CREATE DATABASE IF NOT EXISTS house_data;
USE house_data;

CREATE TABLE house_prices (
    id INT AUTO_INCREMENT PRIMARY KEY,
    size INT,
    bedrooms INT,
    bathrooms INT,
    floors INT,
    age INT,
    garage BOOLEAN,
    location_score INT,
    price INT
);

INSERT INTO house_prices (size, bedrooms, bathrooms, floors, age, garage, location_score, price) VALUES
(2100, 3, 2, 2, 20, 1, 8, 399900),
(1600, 3, 2, 1, 15, 1, 7, 329900),
(2400, 3, 3, 2, 30, 1, 6, 369000),
(1416, 2, 1, 1, 20, 0, 5, 232000),
```

(3000, 4, 3, 2, 8, 1, 9, 539900),

(1985, 4, 2, 2, 12, 1, 8, 299900),

(1534, 3, 2, 1, 10, 0, 6, 314900),

(1427, 3, 1, 1, 18, 0, 5, 198999),

(1380, 3, 2, 1, 14, 0, 5, 212000),

(1494, 3, 2, 1, 12, 1, 6, 242500),

(1800, 4, 3, 2, 7, 1, 9, 355000),

(1750, 3, 2, 2, 5, 1, 9, 339000),

(2200, 4, 3, 2, 3, 1, 10, 459900),

(1950, 3, 2, 2, 6, 1, 8, 310000),

(1300, 2, 1, 1, 25, 0, 4, 190000),

(1700, 3, 2, 1, 15, 1, 6, 299000),

(1650, 3, 2, 1, 10, 0, 7, 278000),

(2500, 4, 3, 2, 5, 1, 9, 488000),

(1450, 3, 1, 1, 17, 0, 5, 210000),

(1550, 3, 2, 1, 12, 1, 6, 235000),

(2000, 4, 3, 2, 9, 1, 9, 350000),

(1725, 3, 2, 1, 11, 0, 7, 290000),

(1350, 2, 1, 1, 22, 0, 4, 185000),

(1900, 3, 2, 2, 7, 1, 8, 320000),

(1850, 3, 2, 1, 13, 1, 6, 298000),

(1600, 3, 2, 1, 18, 0, 5, 275000),

(1550, 3, 2, 1, 20, 0, 6, 265000),

(1780, 3, 2, 2, 6, 1, 7, 310000),

(2100, 4, 3, 2, 4, 1, 9, 420000),

(1250, 2, 1, 1, 30, 0, 3, 175000),

(2600, 4, 3, 2, 2, 1, 10, 499000),

(1950, 3, 2, 2, 10, 1, 8, 330000),

(1430, 3, 2, 1, 16, 0, 5, 225000),

(1540, 3, 2, 1, 14, 0, 6, 238000),

(1700, 3, 2, 1, 8, 1, 7, 289000),

(2150, 4, 3, 2, 5, 1, 9, 430000),

(1875, 3, 2, 2, 7, 1, 8, 320000),

(1350, 2, 1, 1, 22, 0, 4, 180000),

(1400, 2, 1, 1, 20, 0, 4, 195000),

(1980, 4, 2, 2, 9, 1, 8, 340000),

(2250, 4, 3, 2, 6, 1, 9, 455000),

(1600, 3, 2, 1, 18, 0, 6, 275000),

(1900, 3, 2, 2, 5, 1, 8, 325000),

(1425, 3, 2, 1, 16, 0, 6, 230000),

(2000, 3, 2, 2, 10, 1, 8, 349000),

(1700, 3, 2, 1, 14, 1, 7, 285000),

(2500, 4, 3, 2, 3, 1, 10, 495000),

(1450, 2, 1, 1, 25, 0, 5, 200000),

(1650, 3, 2, 1, 12, 0, 6, 255000),

(1850, 4, 2, 2, 7, 1, 8, 310000),

(1750, 3, 2, 1, 10, 1, 7, 295000);

SELECT * FROM house_prices;

OUTPUT:

```
=== Run information ===

Scheme:       weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4
Relation:     house_prices
Instances:    51
Attributes:   9
              id
              size
              bedrooms
              bathrooms
              floors
              age
              garage
              location_score
              price
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===


Linear Regression Model

price =

    158.5829 * size +
  20596.5752 * bathrooms +
 -18330.8409 * floors +
  16069.423  * location_score +
-101963.7692

Time taken to build model: 0.03 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.9557
Mean absolute error                  21177.2641
Root mean squared error              26513.0826
Relative absolute error                 29.6242 %
Root relative squared error             28.7203 %
Total Number of Instances               51
```