

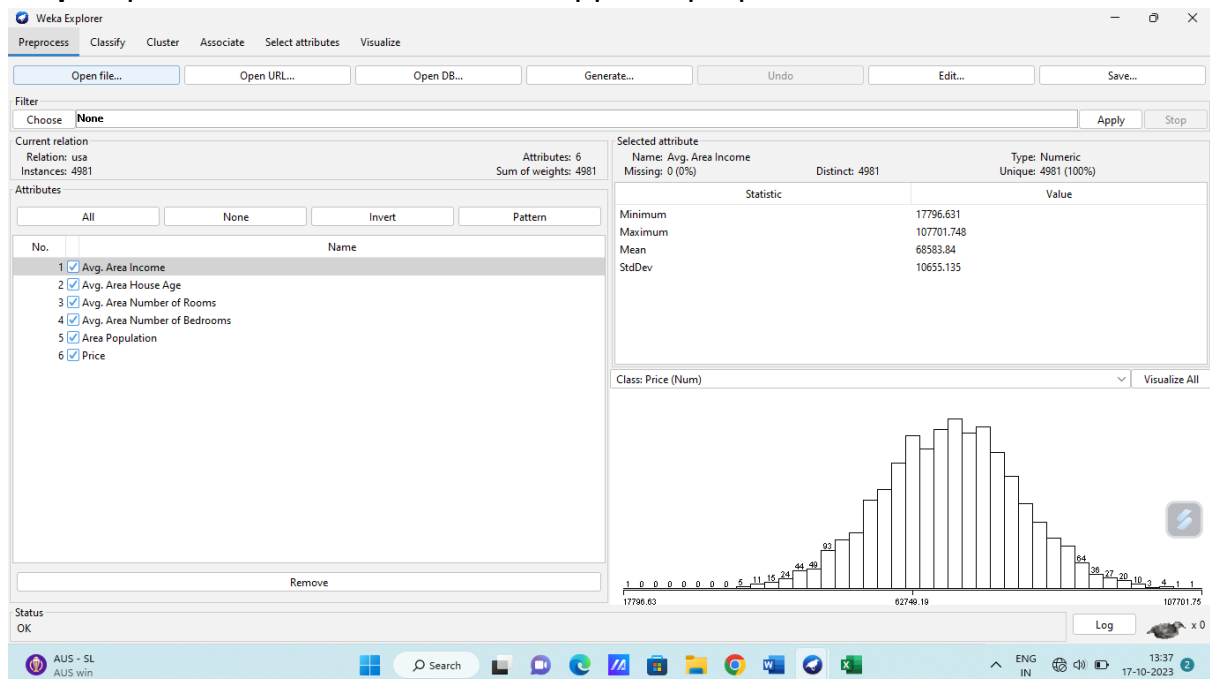
AI_PHASE3: PREDICTING HOUSE PRICES USING MACHINE LEARNING:

PREPROCESSING AND LOADING

Step1: Collect the dataset from the Kaggle website.

Step2:download and install the weka app

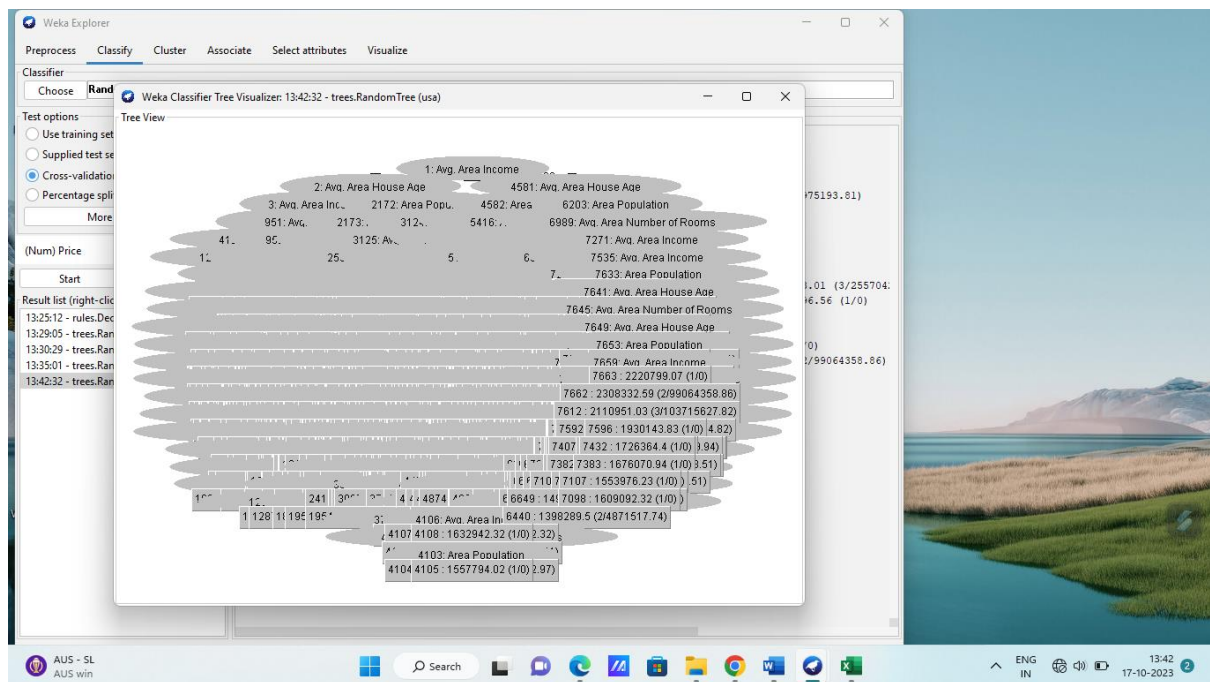
Step3:open the dataset in the weka app and preprocess.



CLASSIFY :

Step1:go to classifier and choose random forest tree.

Step2: click start to run the code and visualize the tree.



CLUSTER:

=== Run information ===

Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100

Relation: usa

Instances: 4981

Attributes: 6

Avg. Area Income

Avg. Area House Age

Avg. Area Number of Rooms

Avg. Area Number of Bedrooms

Area Population

Price

Test mode: evaluate on training data

=== Clustering model (full training set) ===

EM

==

Number of clusters selected by cross validation: 16

Number of iterations performed: 100

		Cluster							
Attribute			0	1	2	3	4	5	6
7	8	9	10	11	12	13	14	15	
			(0.04)	(0.12)	(0.03)	(0.02)	(0.13)	(0.09)	
(0.08)	(0.02)	(0)	(0.03)	(0.13)	(0.08)	(0.13)	(0.07)		
(0.01)	(0.02)								

=====
=====
=====
=====

Avg. Area Income

mean	73914.2749	62923.1802	81979.8821	86352.3168
72886.2674	65980.9841	65784.8849	54959.0076	61837.7316
52382.8165	63100.9233	63443.5285	73055.0793	74312.0321
75379.2191	86706.1966			

std. dev.	5895.2499	8561.0431	8985.153	6515.9877
7845.4864	8903.1647	9587.7086	3736.3265	11845.9935
8329.1034	8377.9475	7477.2597	7878.4218	6158.2273
				6202.1902

Avg. Area House Age

mean		6.539	5.5758	6.9357	5.8368	6.1689
5.882	5.7363	7.2674	5.4289	5.0518	5.6631	5.7246
6.1271	6.2115	8.0439	6.4245			
std. dev.		0.7007	0.9192	0.7945	0.8198	0.846
1.0054	0.9675	0.59	0.9493	0.6694	0.9271	0.9606
0.8272	0.9112	0.4831	0.7981			

Avg. Area Number of Rooms

mean		8.0239	6.6386	8.1124	7.5287	7.0592
7.724	7.7214	7.2889	8.2191	6.1071	6.7156	6.0981
6.9592	6.3279	6.9796	6.4782			
std. dev.		0.6458	0.9705	0.8837	0.3727	1.0016
0.5421	0.5293	0.9249	0.4316	0.8974	0.929	0.6546
0.9575	0.5283	1.0722	0.6356			

Avg. Area Number of Bedrooms

mean		5.6207	4.2533	4.4279	5.8792	3.2436
6.2582	5.2421	3.8157	3.6677	3.307	3.2592	2.2436
4.243	2.2529	3.6184	3.2359			
std. dev.		0.5655	0.1449	1.0201	0.4773	0.1431
0.1431	0.1465	0.7814	0.5379	0.7799	0.1439	0.1475
0.1501	0.1485	0.7165	0.7105			

Area Population

mean		38941.1543	33125.2642	44674.2572	42145.275
38492.4719	33814.7516	34194.7219	43334.2135	60569.0238	28356.55
33080.9574	33785.5471	37851.9045	38520.5599	36604.2998	
43233.1232					
std. dev.		9340.4261	9349.4529	8740.0965	8506.5246
9019.5946	9094.0187	9767.8057	8679.8437	4855.1501	9344.4575
9435.2353	9372.8245	8464.8893	9511.8492	10070.2796	9052.0164

Price

mean 1627173.7545 939650.9967 1985375.442 1774954.7551
1406013.2133 1206075.7312 1184909.4988 1316230.9424 1566322.4636
452589.2999 963231.2016 921097.0108 1393815.3978 1358487.011
1755599.5515 1789815.1504

std. dev. 117200.2835 181093.6703 160980.9993 153434.8518
172815.0136 248183.1701 250113.7039 131695.683 271090.4544
145461.8848 179507.6007 172232.445 177660.4082 170276.1761
96537.2487 118567.1376

Time taken to build model (full training data) : 310.17 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 192 (4%)
1 589 (12%)
2 133 (3%)
3 71 (1%)
4 646 (13%)
5 469 (9%)
6 421 (8%)
7 68 (1%)
8 23 (0%)

9	124 (2%)
10	636 (13%)
11	394 (8%)
12	683 (14%)
13	363 (7%)
14	58 (1%)
15	111 (2%)

Log likelihood: -38.83307

ASSOCIATE:

=== Run information ===

Scheme: weka.associations.FilteredAssociator -F "weka.filters.MultiFilter -F \"weka.filters.unsupervised.attribute.ReplaceMissingValues \" -S 1" -c -1 -W weka.associations.Apriori -- -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: usa

Instances: 4981

Attributes: 6

Avg. Area Income

Avg. Area House Age

Avg. Area Number of Rooms

Avg. Area Number of Bedrooms

Area Population

Price

ATTRIBUTE:

=== Run information ===

Evaluator: weka.attributeSelection.ClassifierAttributeEval -execution-slots 1
-B weka.classifiers.rules.ZeroR -F 5 -T 0.01 -R 1 -E DEFAULT --

Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1

Relation: usa

Instances: 4981

Attributes: 6

Avg. Area Income

Avg. Area House Age

Avg. Area Number of Rooms

Avg. Area Number of Bedrooms

Area Population

Price

Evaluation mode: evaluate on all training data

=== Attribute Selection on all input data ===

Search Method:

Attribute ranking.

Attribute Evaluator (supervised, Class (numeric): 6 Price):

Classifier feature evaluator

Using Wrapper Subset Evaluator

Learning scheme: `weka.classifiers.rules.ZeroR`

Scheme options:

Subset evaluation: RMSE

Number of folds for accuracy estimation: 5

Ranked attributes:

- 0 5 Area Population
- 0 2 Avg. Area House Age
- 0 3 Avg. Area Number of Rooms
- 0 4 Avg. Area Number of Bedrooms
- 0 1 Avg. Area Income

Selected attributes: 5,2,3,4,1 : 5

VISUALIZE:

