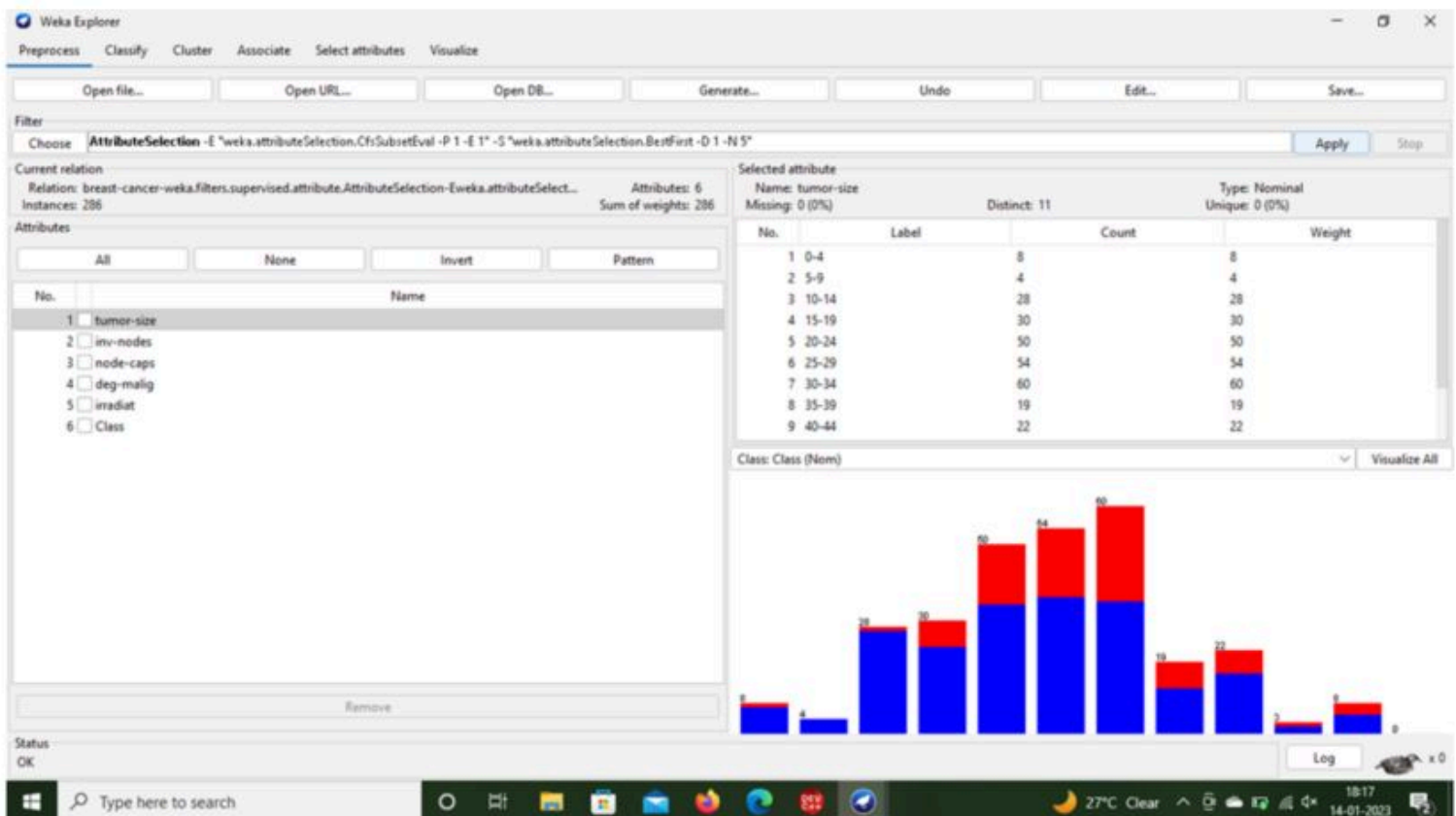


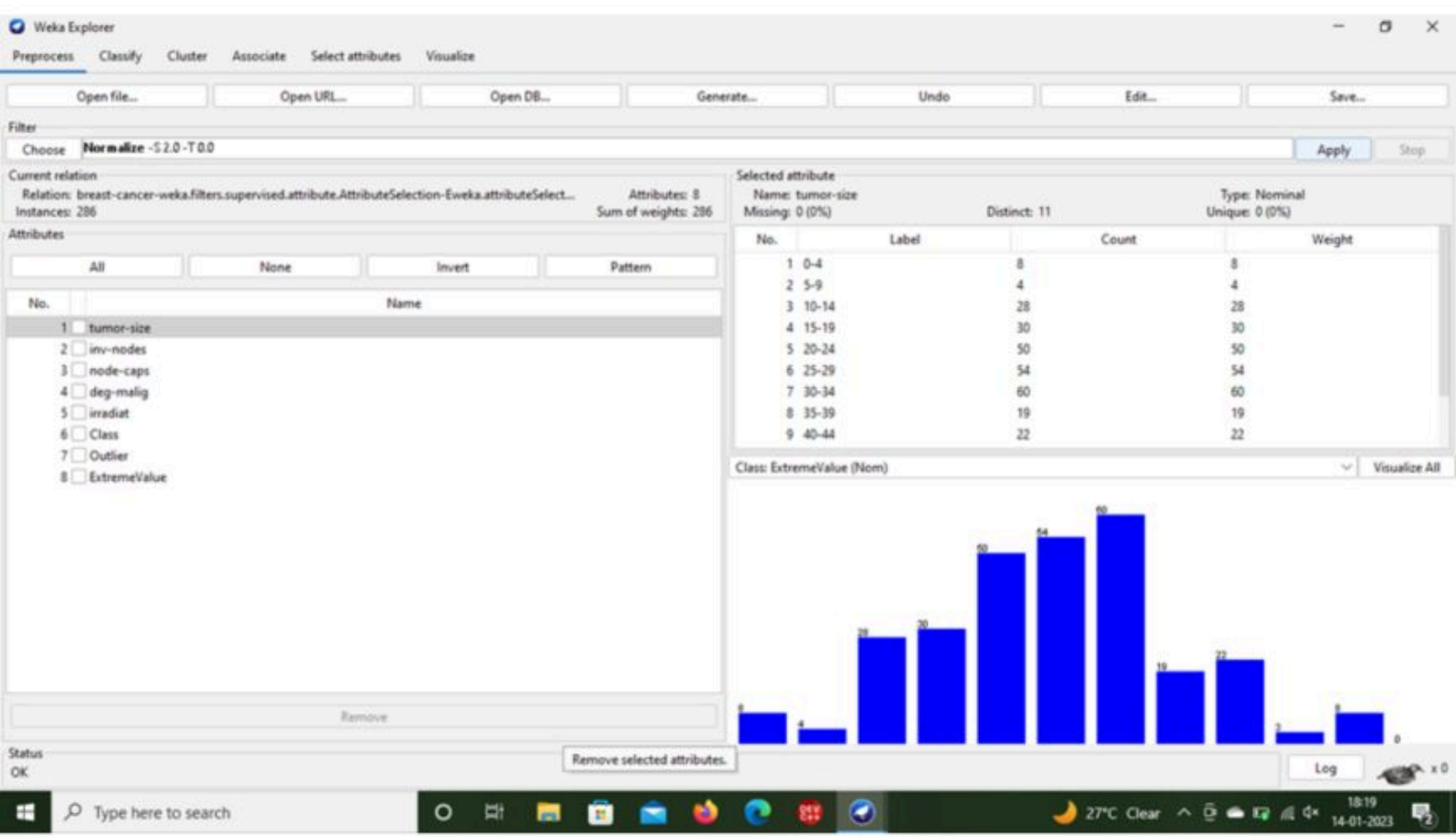
CSA1668-DATA WAREHOUSING AND DATA MINING

K.SRIKANTH
192110218

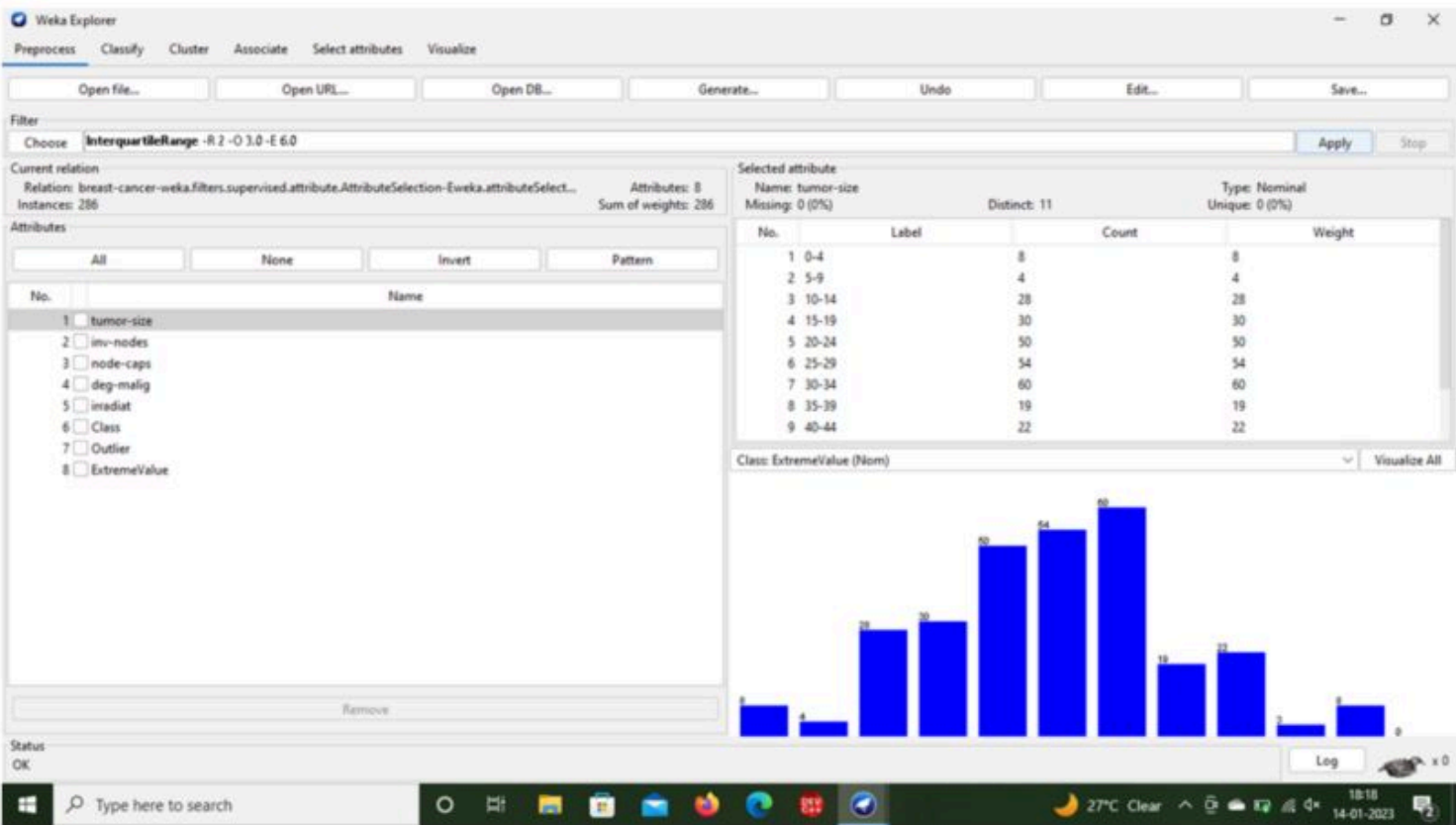
Data Pre processing



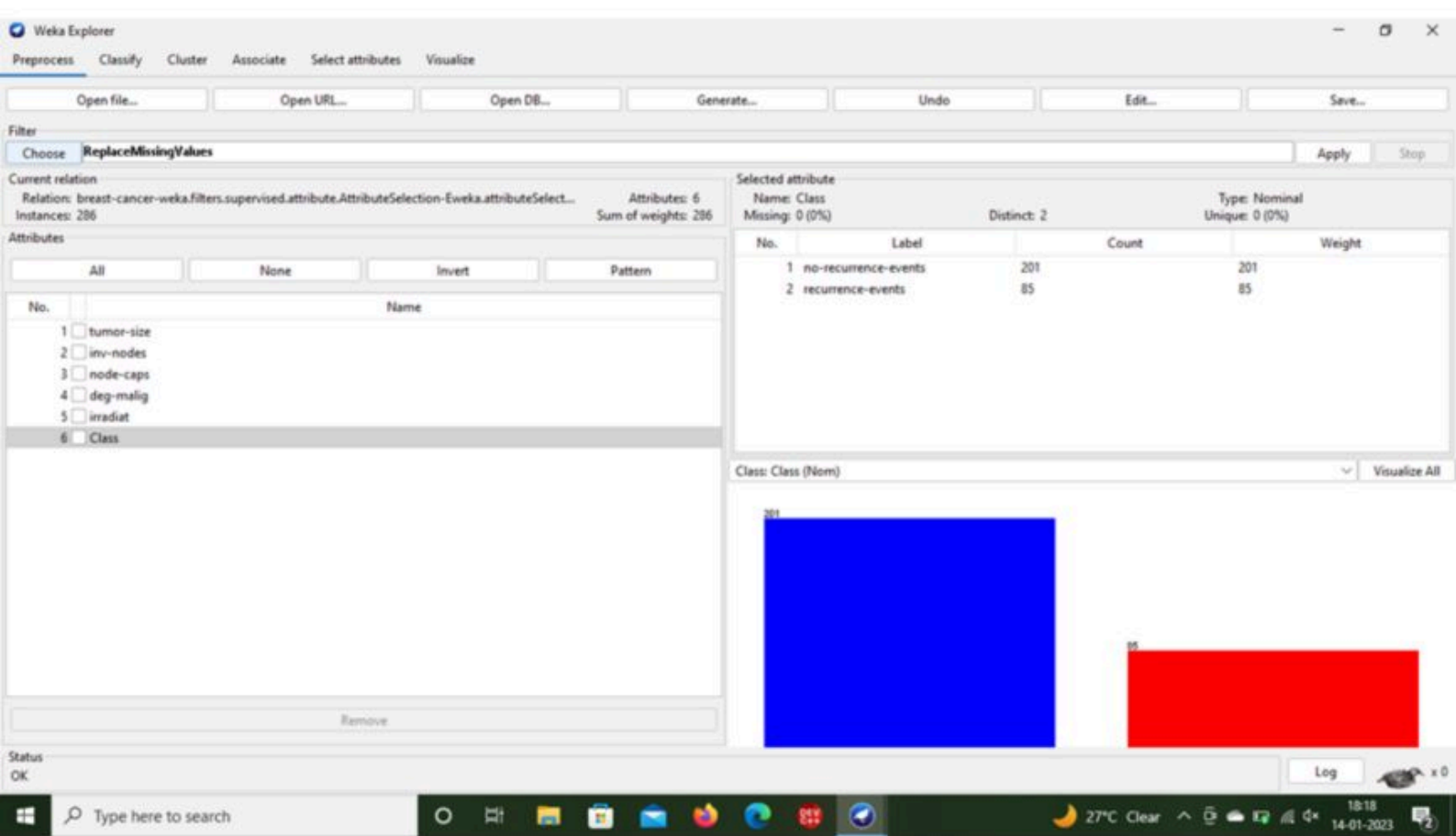
NORMALIZE



INTERQUARTILE RANGE



REPLACE MISSING VALUES



DISCRETIZE

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter
Choose **Discretize** -B 10 -M -1.0 -R first-last -precision 6 Apply Stop

Current relation
Relation: breast-cancer-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelect...
Instances: 286 Attributes: 8 Sum of weights: 286

Attributes
All None Invert Pattern

No.	Name
1	<input checked="" type="checkbox"/> tumor-size
2	<input type="checkbox"/> inv-nodes
3	<input type="checkbox"/> node-caps
4	<input type="checkbox"/> deg-malig
5	<input type="checkbox"/> irradiat
6	<input type="checkbox"/> Class
7	<input type="checkbox"/> Outlier
8	<input type="checkbox"/> ExtremeValue

Remove

Status
OK

Selected attribute
Name: tumor-size
Missing: 0 (0%) Distinct: 11 Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	0-4	8	8
2	5-9	4	4
3	10-14	28	28
4	15-19	30	30
5	20-24	50	50
6	25-29	54	54
7	30-34	60	60
8	35-39	19	19
9	40-44	22	22

Class: ExtremeValue (Nom) Visualize All

Log x 0

Type here to search

27°C Clear 18:20 14-01-2023

CLUSTERING

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 1.25 -t2 1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode

☒ Use training set

☐ Supplied test set

☐ Percentage split %

☐ Classes to clusters evaluation (Nom) ExtremeValue

☒ Store clusters for visualization

Result list (right-click for options)

18:21:22 - SimpleKMeans

Cluster output

Cluster 0: 10-14,0-2,no,2,no,no-recurrence-events,no,no

Cluster 1: 15-19,0-2,yes,3,no,recurrence-events,no,no

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (286.0)	Cluster# 0 (230.0)	1 (56.0)
tumor-size	30-34	20-24	30-34
inv-nodes	0-2	0-2	0-2
node-caps	no	no	yes
deg-malig	2	2	3
irradiat	no	no	no
Class	no-recurrence-events	no-recurrence-events	recurrence-events
Outlier	no	no	no
ExtremeValue	no	no	no

Time taken to build model (full training data) : 0.02 seconds

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

Clustered Instances

0	230 (80%)
1	56 (20%)

Status OK

Log

Type here to search

27°C Clear 18:21 14-01-2023

ASSOCIATE

Weka Explorer

Preprocess Classify Cluster **Associate** Select attributes Visualize

Associate

Choose **Apriori - N 10 - T 0 - C 0.9 - D 0.05 - U 1.0 - M 0.1 - S - 1.0 - c - 1**

Start Stop

Result list (right-click for ...)

18:21:55 - Apriori

Associate output

=== Associate model (full training set) ===

Apriori

Minimum support: 0.75 (214 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 5

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4
Size of set of large itemsets L(2): 5
Size of set of large itemsets L(3): 2

Best rules found:

1. ExtremeValue=no 286 ==> Outlier=no 286 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. Outlier=no 286 ==> ExtremeValue=no 286 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. node-caps=no 230 ==> Outlier=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. node-caps=no 230 ==> ExtremeValue=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. node-caps=no ExtremeValue=no 230 ==> Outlier=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. node-caps=no Outlier=no 230 ==> ExtremeValue=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. node-caps=no 230 ==> Outlier=no ExtremeValue=no 230 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. irradiat=no 218 ==> Outlier=no 218 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. irradiat=no 218 ==> ExtremeValue=no 218 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. irradiat=no ExtremeValue=no 218 ==> Outlier=no 218 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

Status OK

Log x 0

Type here to search

27°C Clear 18:21 14-01-2023

EM

The screenshot shows the Weka Explorer application with the 'Cluster' tab selected. The 'Cluster mode' section has 'Use training set' selected. The 'Clusterer' dropdown is set to 'EM'. The 'Clusterer output' section displays the results of the clustering process, including a table of cluster statistics and a list of clustered instances.

Cluster mode

- ☒ Use training set
- ☐ Supplied test set
- ☐ Percentage split
- ☐ Classes to clusters evaluation
- ☒ Store clusters for visualization

Clusterer

Choose **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

Clusterer output

[total]	212.8894	79.1106
irradiat		
yes	27.0213	42.9787
no	184.8681	35.1319
[total]	211.8894	78.1106
Class		
no-recurrence-events	149.7021	33.2979
recurrence-events	42.1873	44.8127
[total]	211.8894	78.1106
Outlier		
no	210.8894	77.1106
yes	1	1
[total]	211.8894	78.1106
ExtremeValue		
no	210.8894	77.1106
yes	1	1
[total]	211.8894	78.1106

Time taken to build model (full training data) : 0.93 seconds

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

Clustered Instances

0	211 (74%)
1	75 (26%)

Log likelihood: -5.35071

CLASSIFICATION

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier Choose **J48 -C 0.25 -M 2**

Test options

- ☐ Use training set
- ☐ Supplied test set
- ☒ Cross-validation Folds
- ☐ Percentage split %

(Nom) ExtremeValue

Result list (right-click for options)

- 18:17:21 - trees.J48
- 18:20:14 - trees.J48**

Status OK

Classifier output

Size of the tree : 1

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	286	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	286		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	?	1.000	1.000	1.000	1.000	?	?	1.000	no
?	0.000	?	?	?	?	?	?	?	yes
Weighted Avg.	1.000	?	1.000	1.000	1.000	?	?	1.000	

=== Confusion Matrix ===

a	b	<-- classified as	
286	0	a = no	
0	0	b = yes	

OneDrive

Screenshot saved

The screenshot was added to your OneDrive.

27°C Clear 18:20 14-01-2023