

The background is a solid dark blue. It is decorated with various hand-drawn style elements: a green squiggly line in the top left, a yellow square in the top center, a blue square in the top right, a yellow square with a blue checkmark in the middle left, a yellow circle in the middle right, a green square in the bottom left, a blue square in the bottom center, a red square in the bottom right, and a green squiggly line in the bottom right.

Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems

Predict applicants' academic performance at university

2,093(students)



FROM 2016 TO 2019
ENROLLED IN A COMPUTER
SCIENCE AND INFORMATION COLLEGE

EARLY UNIVERSITY PERFORMANCE

- HIGH SCHOOL GRADE AVERAGE
- SCHOLASTIC ACHIEVEMENT
ADMISSION TEST SCORE
- GENERAL APTITUDE TEST SCORE

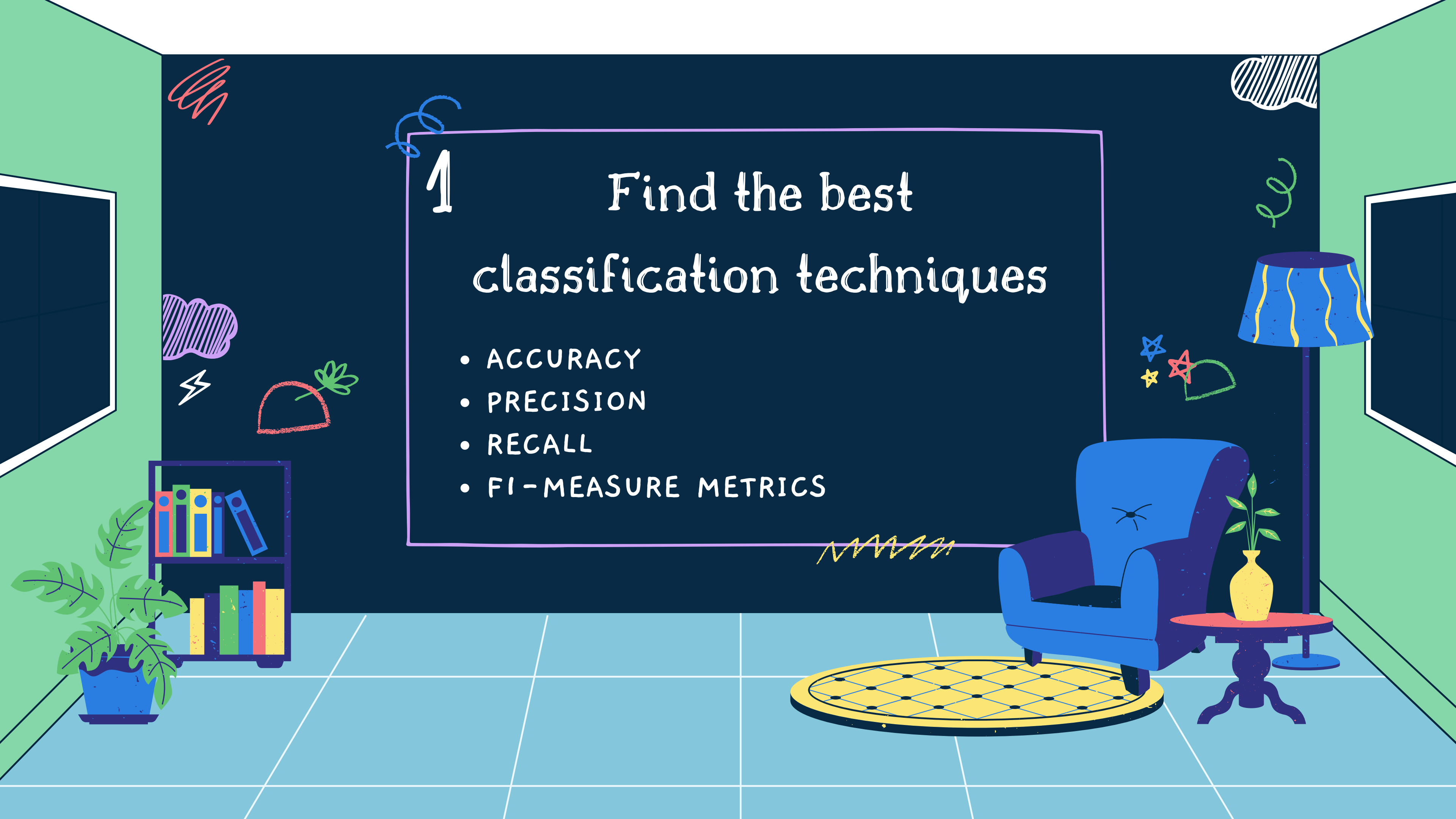


SCHOLASTIC ACHIEVEMENT
ADMISSION TEST SCORE

1

Find the best classification techniques

- ACCURACY
- PRECISION
- RECALL
- F1-MEASURE METRICS



2

Find relation

- ADMISSIONS CRITERIA
- STUDENTS' CUMULATIVE GRADE POINT AVERAGE (CGPA) DURING THEIR FIRST YEAR OF STUDY AT UNIVERSITY

3

After change the weighting of its admission criteria



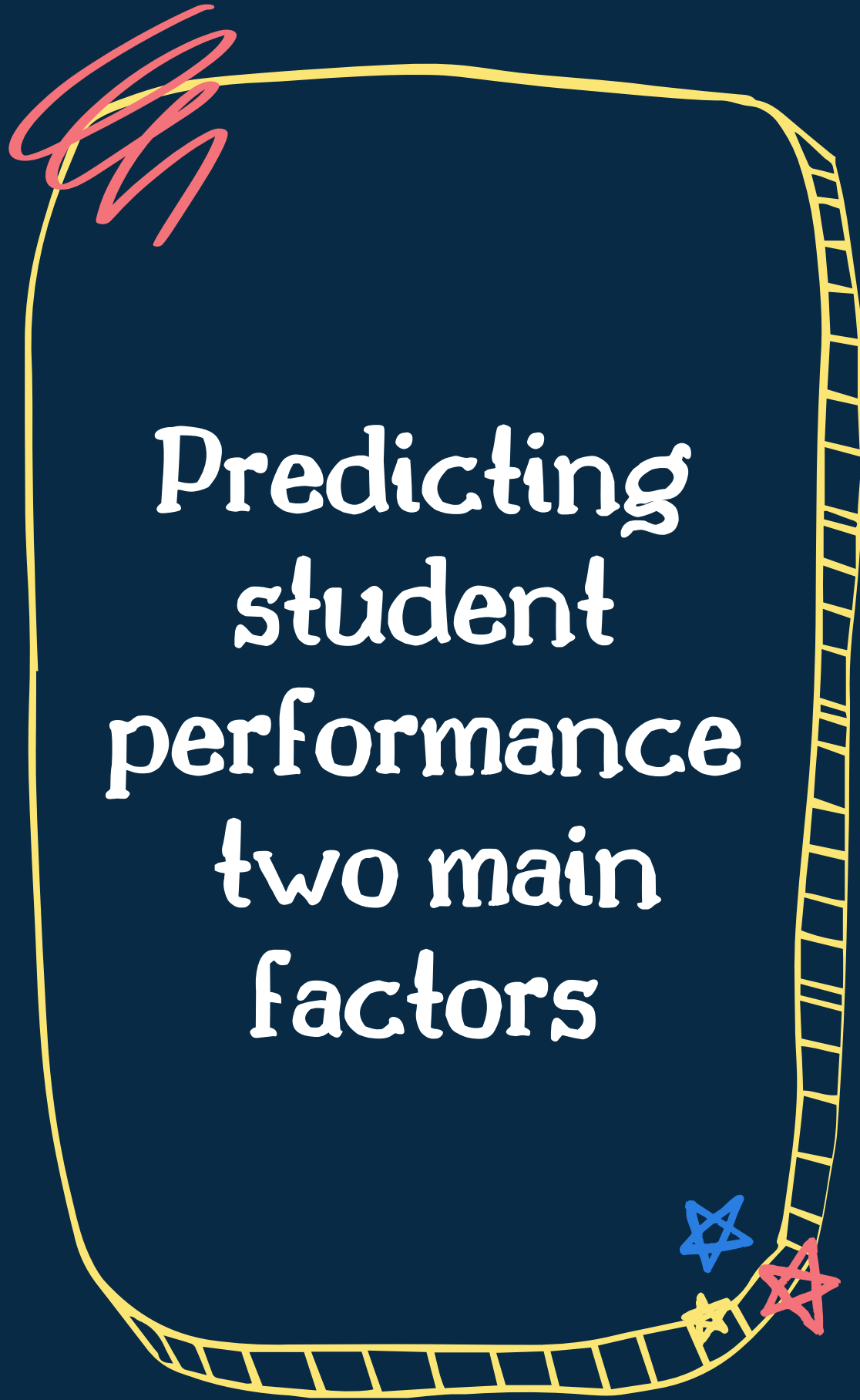
31%




18%





New CGPA



Predicting student performance two main factors




ATTRIBUTES

- 
- STUDENT CGPA
 - ASSESSMENTS, QUIZ GRADES, LAB WORK, AND FINAL EXAM GRADES
 - EXTRA ACTIVITIES, STUDENT DEMOGRAPHY, AND SOCIAL INTERACTION NETWORK
 - PRE-ADMISSION TESTS 



PREDICTION METHODS

- 
- ANN : predict the academic performance of 505 students in their eighth semester.
 - Decision Trees : predict student performance in certain courses. (small student sample sizes)
 - Naïve Bayes : on a set of 1,600 students to predict performance in a particular course.
 - SVM : predict students at risk performance in their first year of study.

UNDERSTANDING THE DOMAIN OF THE STUDY

* 2004-2005 TO 2016-2017

In KSA, enrollment in the computer science, technology colleges has increased

2,959
TO
27,089
(STUDENTS)



* IN 2017-2018

THREE CRITERIA (HSGA, SAAT, AND GAT)

- 30%, 40%, 30% at KSU
- 50%, 20%, 30% at King Abdulaziz University
- 20%, 50%, 30% at King Fahd University of Petroleum & Minerals
- 60%, 20%, and 20% at PNU


* SINCE THE EARLY 2000S

General Aptitude Test (GAT)

Scholastic Achievement Admission Test (SAAT)

- biology
- chemistry
- physics
- mathematics
- English






The
answer
of
the
question




LINEAR REGRESSION TECHNIQUE



the three admission criteria as the independent variables,
the CGPA through students' first two semesters as the
dependent variable.

correlation coefficient

PREDICTION MODELS



Neural Network (ANN), Decision Tree,
Support Vector Machine (SVM), and Naive Bayes

METHODOLOGY



ARTIFICIAL NEURAL NETWORK (ANN)

detect all possible interactions among variables

learn from a limited set of examples.

Multilayer Percep tion (MLP) topology



DECISION TREE

predicts values in a simple and straightforward way

understood and interpreted easily by users



NAIVE BAYES

simplicity,

very good performance for real-world

problems,


computational efficiency



SUPPORT VECTOR MACHINE (SVM)

small datasets

Faster than the other techniques



The
answer
of
the
question



LINEAR



accuracy, recall, precision, and
F1-Measure metrics.

PREDICTION MODELS



developed the second stage of the
study



EXPERIMENTAL EVALUATION

1. EXPERIMENTAL SETUP
 2. EXPERIMENTAL RESULTS AND
DISCUSSION
- 

EXPERIMENTAL SETUP

DATASET

THE FIRST STAGE

1,569 student records
Saudi females
monthly salary from the
government

HSGA, SAAT, and GAT
60%, 20%, and 20%,
SAAT score : most
accurately
HSGA, SAAT, and GAT
30%, 40%, and 30%

THE SECOND STAGE

470 student
the three admission
criteria and CGPA after
the first two semesters

WEKA tool

Our school grading
system applies.

Passing rate is 60.

PREPARED THE DATA SET

- eliminating certain irrelevant attributes
 - identifying outliers
 - deleting records with missing values
 - removing duplicate
- maximized the total
number of observation
used for testing.

EVALUATION METRICS



TRUE POSITIVE RATE
(TP)

the number of
instances that are
correctly predicted as
positive.



FALSE POSITIVE RATE
(FP)

the number of
instances that are
incorrectly predicted as
positive



TRUE NEGATIVE RATE
(TN)

the number of
instances that are
correctly predicted as
negative.



FALSE NEGATIVE RATE
(FN)

the number of
instances that are
incorrectly predicted as
negative.



ACCURACY = $(TP + TN) / (TP + TN + FP + FN)$



RECALL = $TP / (TP + FN)$



PRECISION = $TP / (TP + FP)$



F1 MEASURE = $2 \times \text{RECALL} \times \text{PRECISION} / (\text{RECALL} + \text{PRECISION})$

A decorative border surrounds the slide, featuring various colorful geometric shapes and patterns. At the top, there is a red square, a blue swirl, a yellow diamond, a red circle, a green triangle, a red spiral, a pink semi-circle, and a blue arrow. At the bottom, there is a red swirl, a yellow circle, a blue heart, a red swirl, a green flower, a yellow circle, and several red and yellow stars.

EXPERIMENTAL RESULTS AND DISCUSSION

- 1.MOST SIGNIFICANT CRITERION IDENTIFICATION
- 2.CLASSIFICATION TECHNIQUE PERFORMANCE
- 3.STUDENT PERFORMANCE IMPROVEMENT

result of the
correlation
coefficient
between the
variables

the coefficient
of determination
between the
variables

SAAT



TECHNIQUE PERFORMANCE

ANN

accuracy rate
precision rate

DT

recall rate

NB

worst

Add more weight to the SAAT criterion

★ THE THREE ADMISSION CRITERIA (HSGA, SAAT, GAT)
60%, 20%, 20%
CHANGE TO 30%, 40%, AND 30%

★ EXCELLENT OR VERY GOOD INCREASED TO 65%

★ ACCEPTABLE OR POOR DECREASED TO 6%



CONCLUSION

FOUR DATA MINING

Artificial Neural Network
Decision Tree
Support Vector Machine
Naive Bayes

PERFORMANCE MODEL

ANN is the best

RECOMMENDATIONS

colleges should give
serious consideration to
students' pre admission
SAAT scores before
accepting them



WEKA

1.DECISION TREE

2.SUPPORT VECTOR MACHINE (SVM)

3.NAIVE BAYES

4.ARTIFICIAL NEURAL NETWORK (ANN)

Table

seniorhighgrade

academicexam

apititudeexam

Result

grade

65

55

40

VERYGOOD

150

79

20

12

POOR

0

81

79

30

EXCELLENT

200

46

63

50

ACCEPTABLE

80

STEP 1 CREATE DATASET

	A	B	C	D
1				
2	senior high grade	academic exam	apititude exam	Result
3	73	6	71	very good
4	83	100	92	excellent
5	42	57	55	very good
6	48	17	98	very good
7	23	2	7	poor
8	40	27	82	acceptable
9	43	25	44	acceptable
10	67	45	89	excellent
11	19	46	65	acceptable
12	81	54	50	very good
13	1	78	2	acceptable
14	39	27	77	acceptable
15	92	87	71	excellent
16	51	49	69	very good
17	9	41	85	acceptable
18	57	42	35	acceptable
19	27	21	90	acceptable
20	50	100	90	excellent
21	26	16	21	poor
22	26	23	96	acceptable
23	48	71	2	acceptable
24	20	7	88	acceptable
25	43	50	54	acceptable
26	2	20	95	acceptable
27	61	80	4	acceptable
28	87	43	89	excellent
29	55	27	5	acceptable
30	57	30	60	acceptable
31	4	97	36	acceptable
32	69	89	57	excellent
33	71	20	75	very good
34	87	87	97	excellent

76	81	82	67	excellent
77	11	45	34	acceptable
78	45	15	6	poor
79	88	41	11	acceptable
80	41	57	89	very good
81	99	81	51	excellent
82	39	82	65	very good
83	74	1	60	acceptable
84	31	5	80	acceptable
85	94	78	30	excellent
86	35	84	59	very good
87	61	15	87	very good
88	43	100	62	excellent
89	90	26	61	very good
90	83	34	9	acceptable
91	5	39	21	poor
92	95	22	30	acceptable
93	62	19	67	acceptable
94	80	38	67	very good
95	96	6	98	excellent
96	72	32	85	very good
97	81	67	8	very good
98	16	19	15	poor
99	16	26	74	acceptable
100	85	70	49	excellent
101	24	40	17	acceptable
102	54	31	48	acceptable

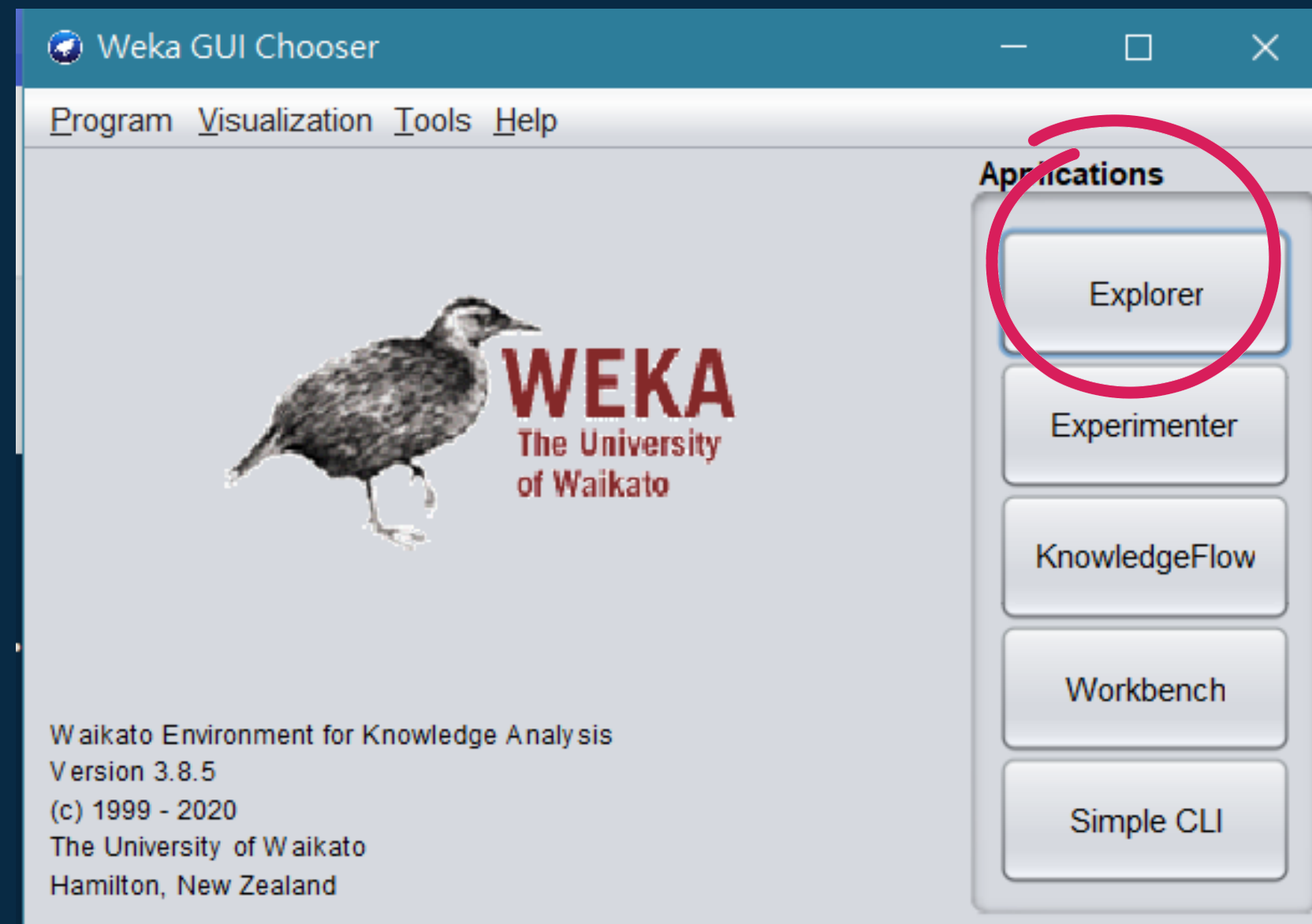
STEP 2 CHANGE DATASET TO .ARFF

```

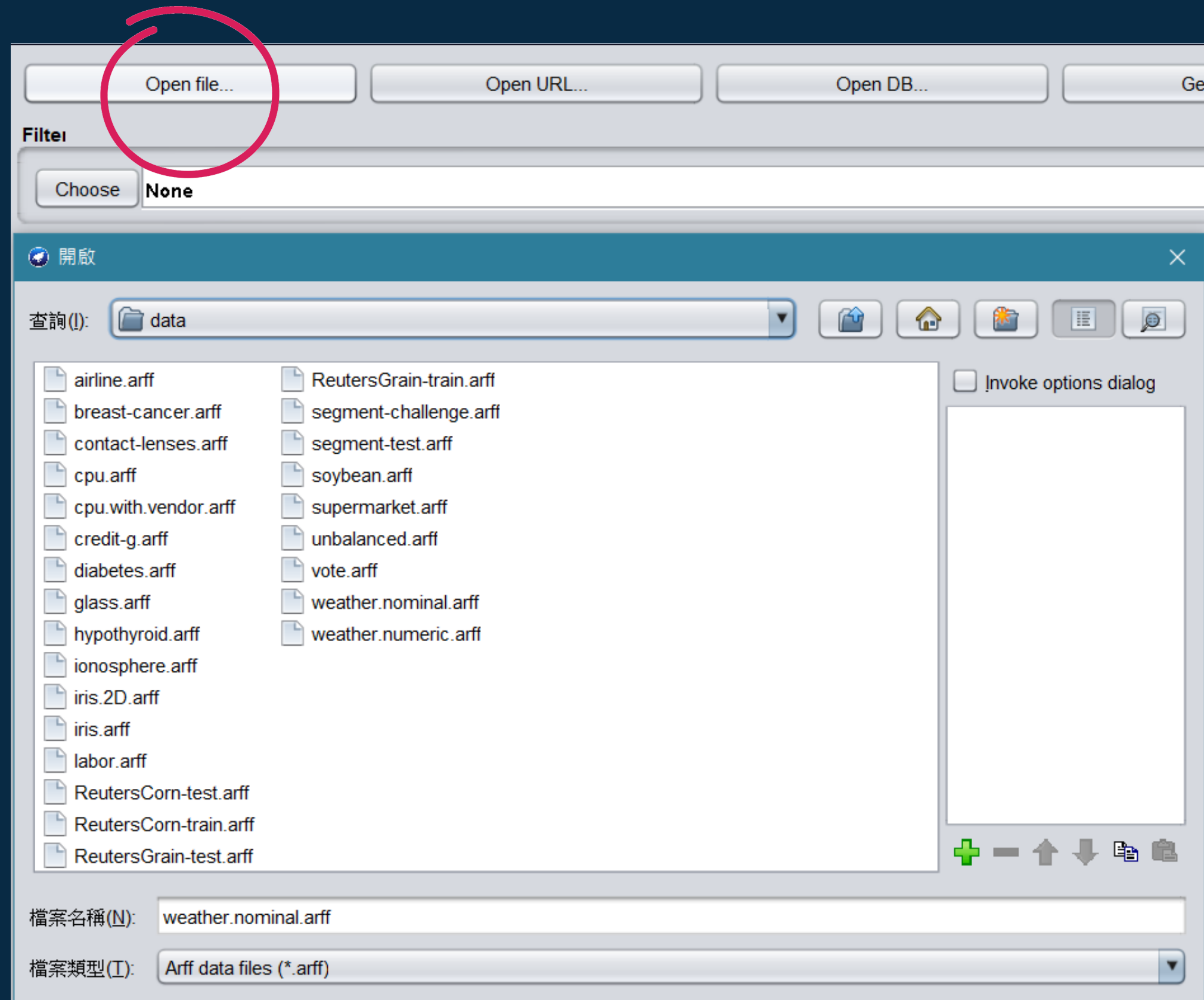
1 @relation weather.symbolic
2
3 @attribute seniorhighgrade{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36}
4 @attribute aademicexam{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36}
5 @attribute apititudeexam{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36}
6 @attribute Result{excellent,verygood,acceptable,poor}
7 @attribute grade{200,150,80,0}
8
9
10 @data
11 73,6,71,verygood,150
12 83,100,92,excellent, 200
13 42,57,55,verygood,150
14 48,17,98,verygood,150
15 23,2,7,poor, 0
16 40,27,82,acceptable,80
17 43,25,44,acceptable,80
18 67,45,89,excellent, 200
19 19,46,65,acceptable,80
20 81,54,50,verygood,150
21 1,78,2,acceptable,80
22 39,27,77,acceptable,80
23 92,87,71,excellent, 200
24 51,49,69,verygood,150
25 9,41,85,acceptable,80
26 57,42,35,acceptable,80
27 27,21,90,acceptable,80
28 50,100,90,excellent, 200
29 26,16,21,poor,0

```

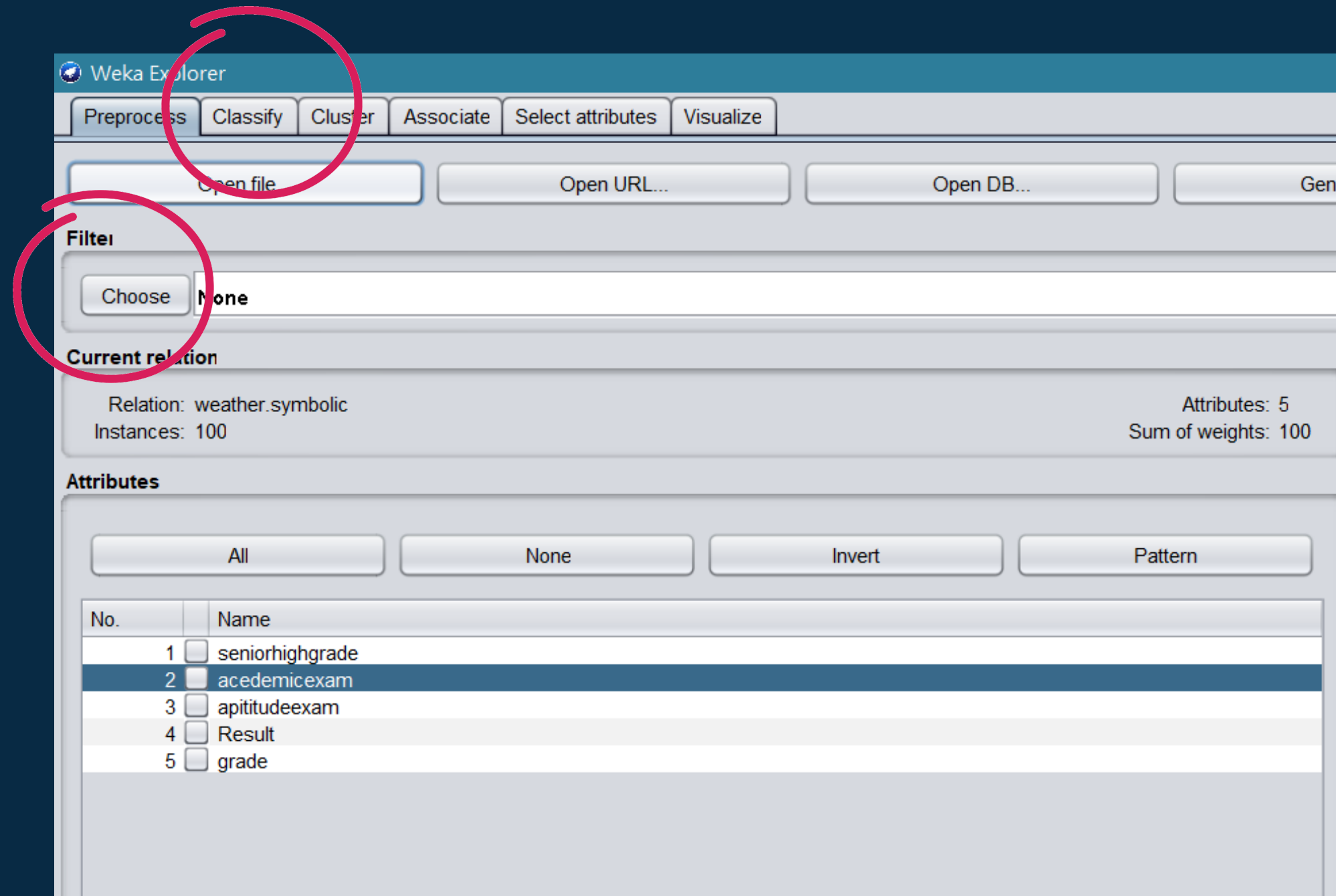

STEP 3 OPEN WEKA & CLICK EXPOLER



STEP 4 CLICK OPEN FILE



STEP 5 CHOOSE CLASSIFY AND CLICK WHICH FILE TO CHOOSE



STEP 6 CLICK START AND WAIT FOR THE RESULT

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 3 -G -R'. The 'Test options' section shows 'Cross-validation' selected with 'Folds' set to 50. The 'Result list' on the left shows the MultilayerPerceptron model selected, circled in red. The 'Classifier output' on the right displays the results for the 'apititudeexam' and 'grade' attributes, including a Sigmoid Node 7 and a Class poor node.

Classifier

Choose **MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H 3 -G -R**

Test options

☐ Use training set
☐ Supplied test set Set...
☒ Cross-validation Folds **50**
☐ Percentage split % **66**
More options...

(Nom) Result

Start Stop

Result list (right-click for options)

- 16:12:02 - trees.M8
- 16:16:13 - NaiveBayes
- 16:16:18 - functions.SMO
- 16:16:51 - functions.MultilayerPerceptron
- 16:21:23 - functions.MultilayerPerceptron**

Classifier output

```
Attrib apititudeexam=88 0.03010345649749896
Attrib apititudeexam=89 -0.008151567050619882
Attrib apititudeexam=90 -0.042347327778797884
Attrib apititudeexam=91 -0.045619664639321045
Attrib apititudeexam=92 -0.010688317220185609
Attrib apititudeexam=93 -0.02230582578664261
Attrib apititudeexam=94 -0.03566485398771346
Attrib apititudeexam=95 0.02250882938268467
Attrib apititudeexam=96 0.027757858334092597
Attrib apititudeexam=97 0.03744408022435754
Attrib apititudeexam=98 0.02124926578271112
Attrib apititudeexam=99 -0.012225082890087183
Attrib apititudeexam=100 0.03141580594504839
Attrib grade=200 -0.0489028077631411
Attrib grade=150 0.01765167055562543
Attrib grade=80 0.001909298469315604
Attrib grade=0 0.028542372158497906

Sigmoid Node 7
Inputs Weights
Threshold 0.008660920341785117

Class excellent
Input
Node 0

Class verygood
Input
Node 1

Class acceptable
Input
Node 2

Class poor
Input
Node 3

Time taken to build model: 12.88 seconds
```

weka3.csv - Excel

檔案 常用 插入 版面配置 公式 資料 校閱 檢視 告訴我您想要執行的動作...

貼上 剪貼 字型 對齊方式 數值 樣式 儲存格 編輯

	A	B	C	D	E	F	G	H	I	J	K
1											
3	73	6	71	very good							
61	36	6	44	acceptable							
95	96	6	98	excellent							

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

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

Filter: Choose None Apply Stop

Current relation: Relation: weather.symbolic Attributes: 5 Instances: 100 Sum of weights: 100

Selected attribute: Name: acedemicexam Type: Nominal Missing: 0 (0%) Distinct: 61 Unique: 33 (33%)

No.	Label	Count	Weight
1	1	1	1.0
2	2	1	1.0
3	3	1	1.0
4	4	0	0.0
5	5	1	1.0
6	6	3	3.0
7	7	1	1.0
8	8	0	0.0
9	9	1	1.0
10	10	0	0.0
11	11	0	0.0
12	12	0	0.0
13	13	1	1.0

Class: grade (Nom) Visualize All

Attributes: All None Invert Pattern

No.	Name
1	<input type="checkbox"/> seniorhighgrade
2	<input checked="" type="checkbox"/> acedemicexam
3	<input type="checkbox"/> apititudeexam
4	<input type="checkbox"/> Result
5	<input type="checkbox"/> grade

Remove

Status

Current relation

Relation: weather.symbolic
Instances: 100

Attributes: 5
Sum of weights: 100

Attributes

AllNoneInvertPattern

No.		Name
1	<input type="checkbox"/>	seniorhighgrade
2	<input type="checkbox"/>	acedemicexam
3	<input type="checkbox"/>	apititudeexam
4	<input type="checkbox"/>	Result
5	<input checked="" type="checkbox"/>	grade

Remove

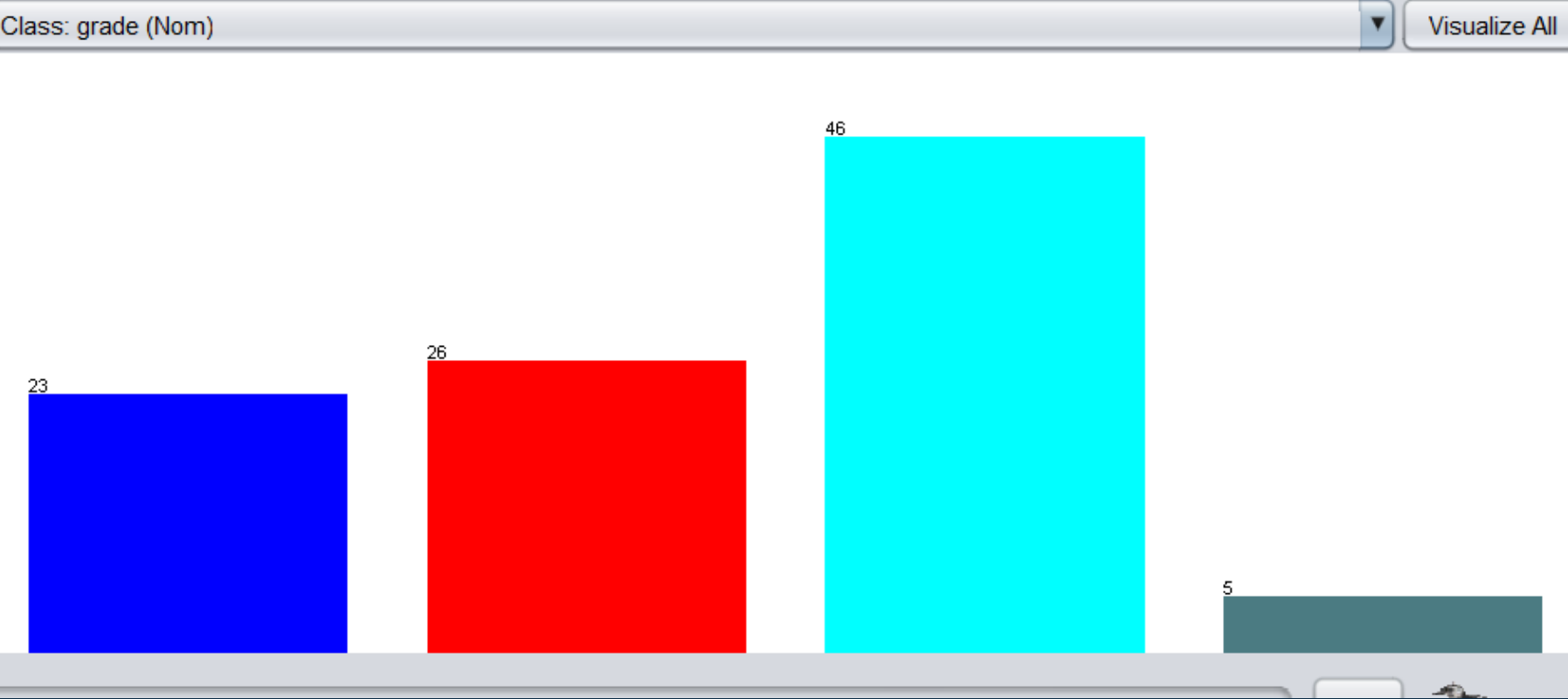
Selected attribute

Name: grade
Missing: 0 (0%)

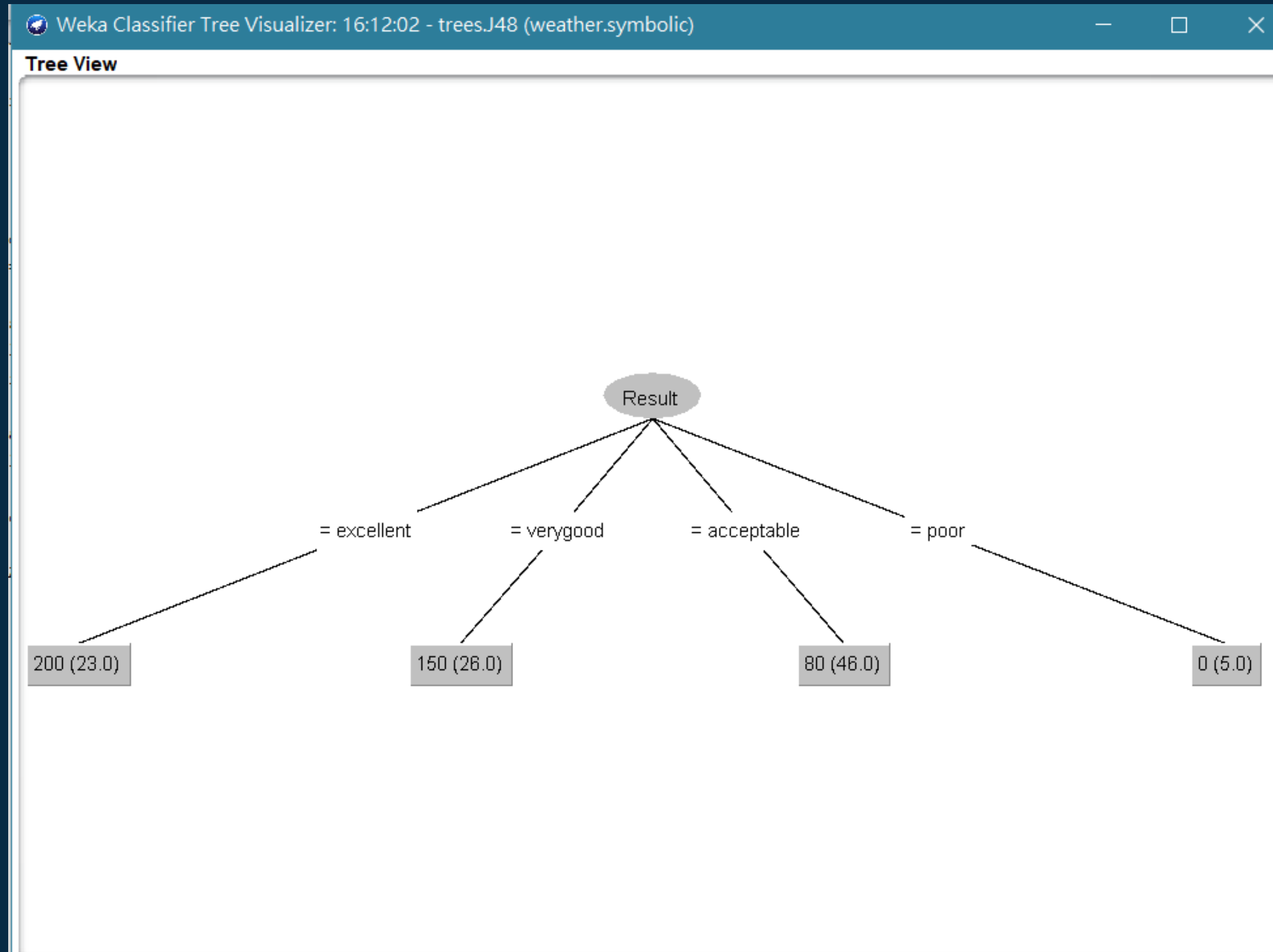
Distinct: 4

Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	200	23	23.0
2	150	26	26.0
3	80	46	46.0
4	0	5	5.0



1. DECISION TREE



=== Summary ===

Correctly Classified Instances	100	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	100		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	200
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	150
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	80
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
23	0	0	0	a = 200
0	26	0	0	b = 150
0	0	46	0	c = 80
0	0	0	5	d = 0

2.SUPPORT VECTOR MACHINE (SVM)

=== Summary ===

Correctly Classified Instances	100	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.25		
Root mean squared error	0.3118		
Relative absolute error	74.6269 %		
Root relative squared error	76.3194 %		
Total Number of Instances	100		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	200
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	150
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	80
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
23	0	0	0	a = 200
0	26	0	0	b = 150
0	0	46	0	c = 80
0	0	0	5	d = 0

3.NAIVE BAYES

=== Summary ===

Correctly Classified Instances	100	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0555		
Root mean squared error	0.0838		
Relative absolute error	16.5556	%	
Root relative squared error	20.4999	%	
Total Number of Instances	100		

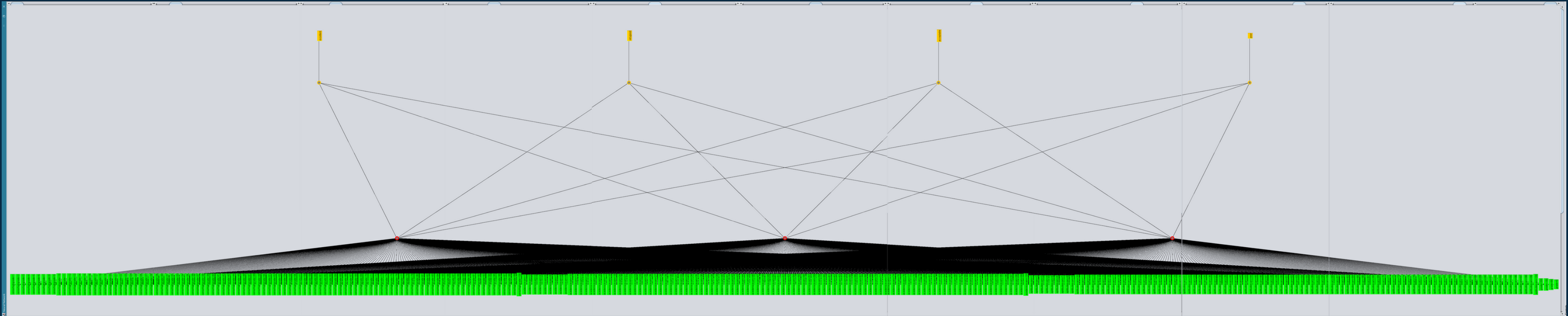
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	200
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	150
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	80
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	0
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
23	0	0	0	a = 200
0	26	0	0	b = 150
0	0	46	0	c = 80
0	0	0	5	d = 0

4. ARTIFICIAL NEURAL NETWORK (ANN)



Sigmoid Node 7

Inputs	Weights
--------	---------

Threshold	0.008660920341785117
-----------	----------------------

Class excellent

Input

Node 0

Class verygood

Input

Node 1

Class acceptable

Input

Node 2

Class poor

Input

Node 3

THANKS FOR LISTENING

