



---

## المشروع النهائي لمقرر التنقيب عن البيانات لذكاء الاعمال MIS342

---

إعداد الطالبة : أماني سليمان الجهني  
الرقم الجامعي : 4051377

أستاذ المقرر د/ سهام ساسي



## المطلوب في المشروع

أن يتم تطبيق نماذج للبيانات من خلال اختيار احدى المجموعات التالية ٨

١- المعالجة المسبقة

٢- الأظهار المرئي

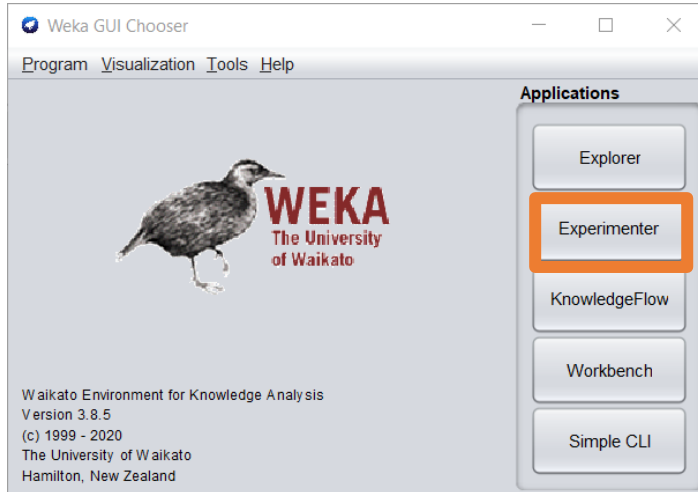
٣- التصنيف وأداة التنبؤ

٤- الشبكات العصبية

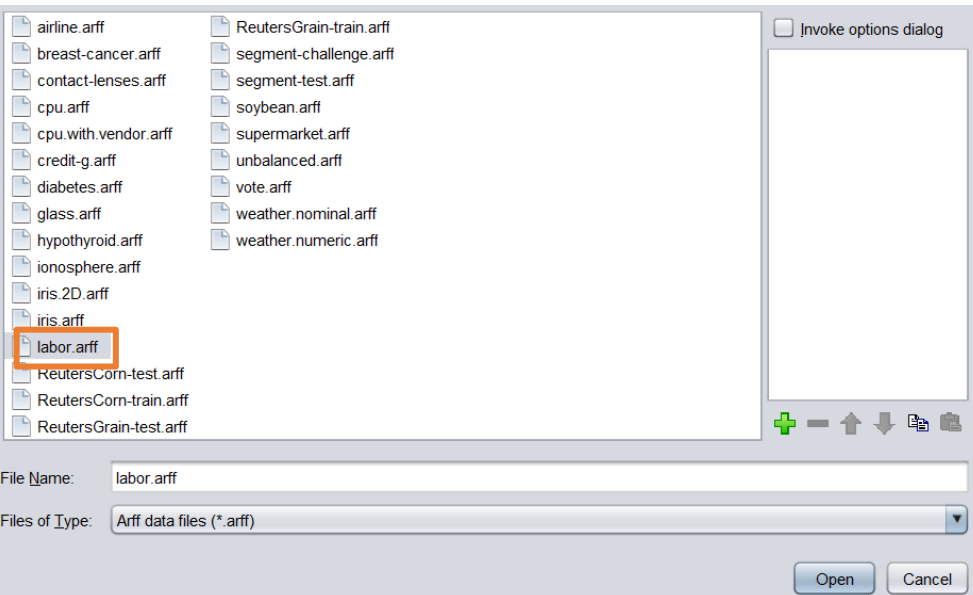
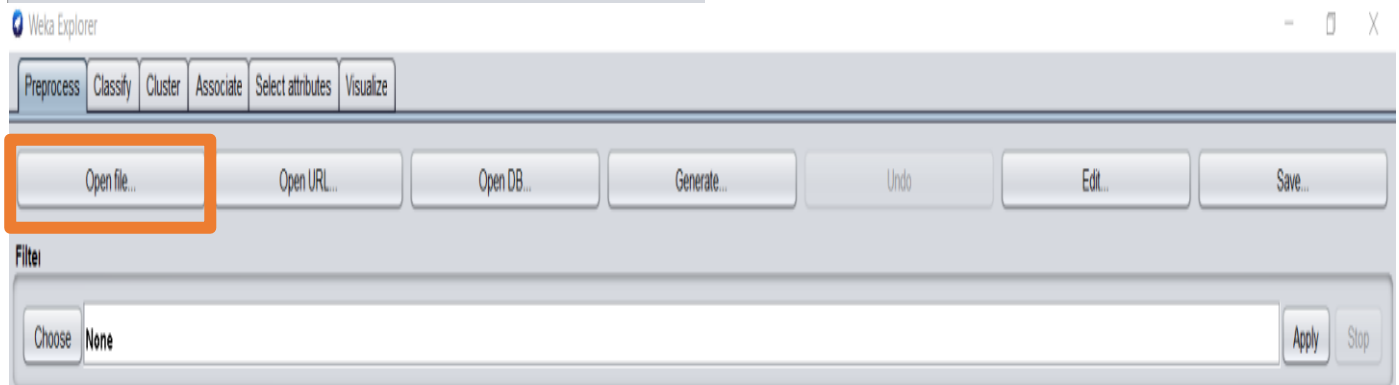
٥- التحليل العنقودي

## أولاً: المعالجة المسبقة

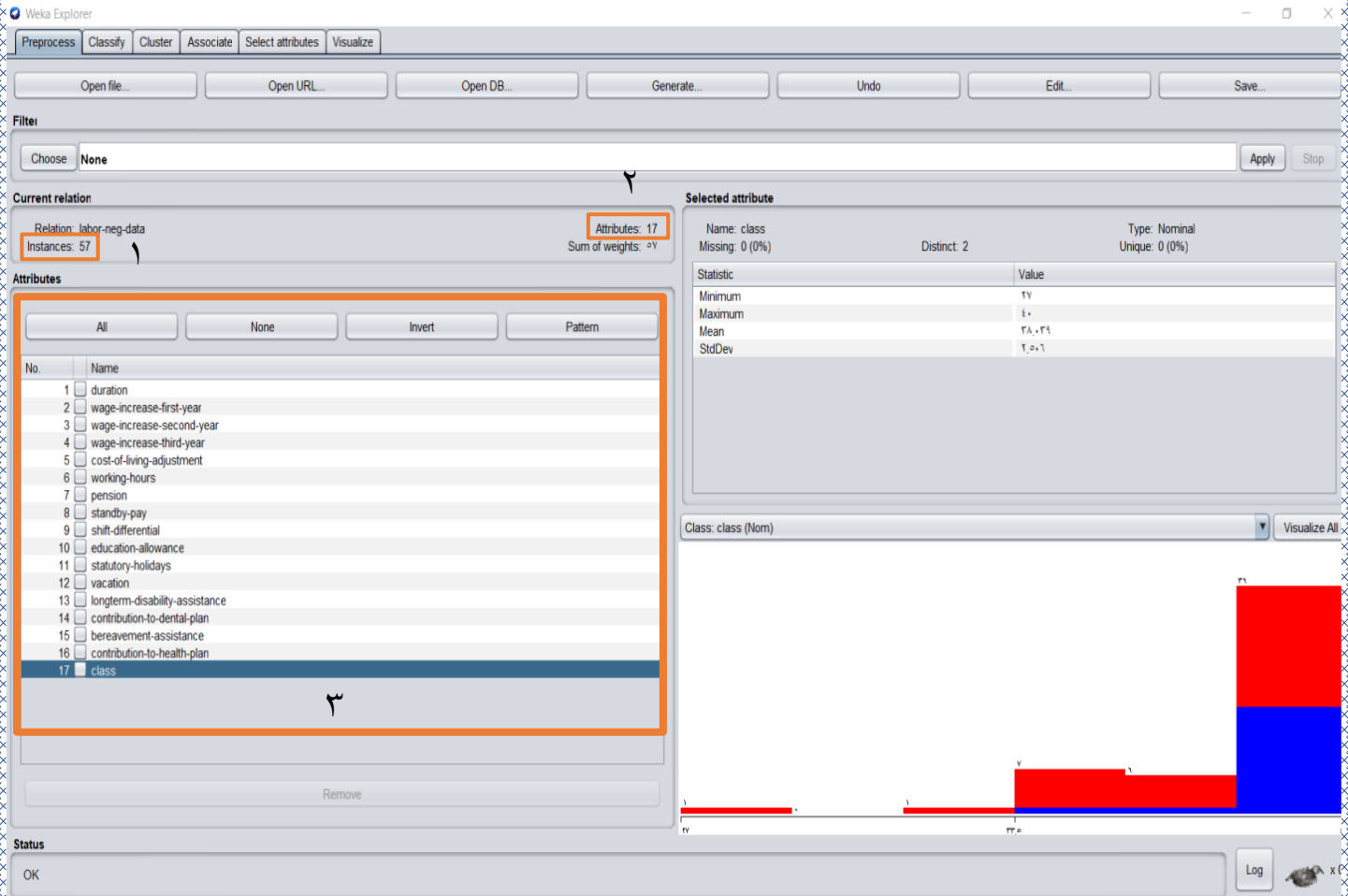
هناك ثلاث طرق لتحميل البيانات للمعالجة المسبقة ونحن سوف نختار أمر open file



```
@relation 'labor-neg-data'
@attribute 'duration' real
@attribute 'wage-increase-first-year' real
@attribute 'wage-increase-second-year' real
@attribute 'wage-increase-third-year' real
@attribute 'cost-of-living-adjustment' {'none','tcf','tc'}
@attribute 'working-hours' real
@attribute 'pension' {'none','ret_allw','empl_contr'}
@attribute 'standby-pay' real
@attribute 'shift-differential' real
@attribute 'education-allowance' {'yes','no'}
@attribute 'statutory-holidays' real
@attribute 'vacation' {'below_average','average','generous'}
@attribute 'longterm-disability-assistance' {'yes','no'}
@attribute 'contribution-to-dental-plan' {'none','half','full'}
@attribute 'bereavement-assistance' {'yes','no'}
@attribute 'contribution-to-health-plan' {'none','half','full'}
@attribute 'class' {'bad','good'}
@data
```



نختار البيانات المراد  
معالجتها وسوف نختار  
في مشروعنا بيانات  
(Labor.arff)



(١) هناك ٥٧ حالة - عدد الصفوف في الجدول

(٢) يحتوي الجدول على ١٧ سمات - الحقول Attributes

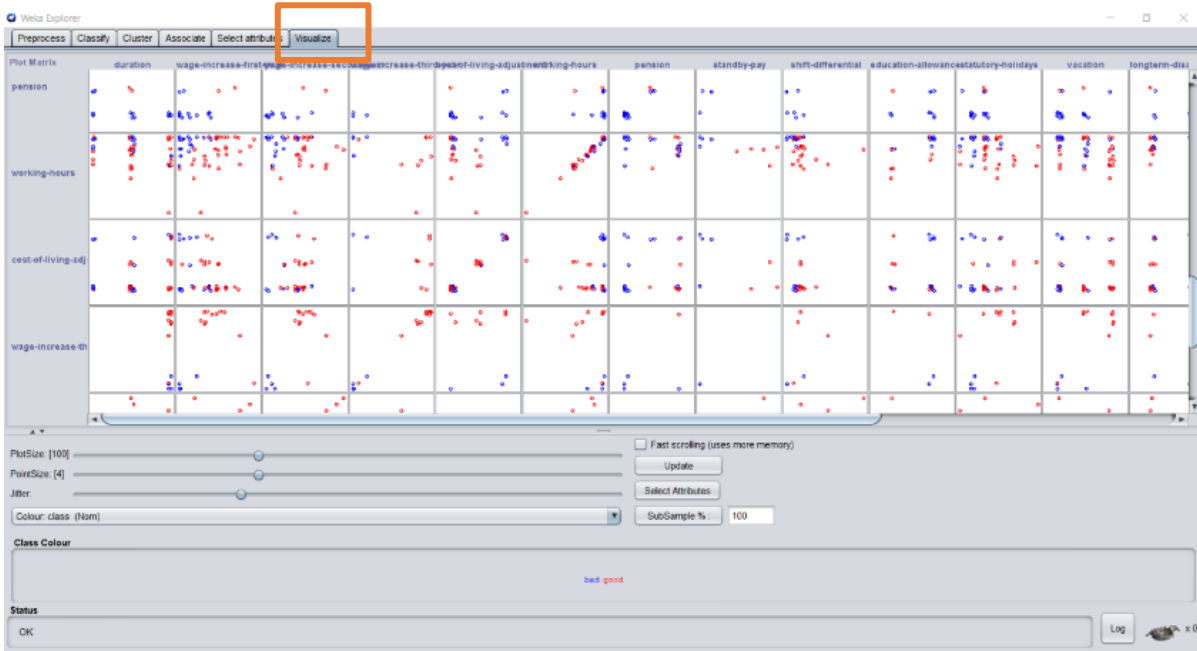
(٣) نافذة السمات الفرعية التي تعرض الحقول المختلفة في قاعدة البيانات.

## ثانياً: الإظهار المرئي

- هو المكون الذي يمكن أن يظهر على لوحة panel لتصور البيانات برسم مبعثر واحد وكبير

ثنائي الأبعاد

مكون ScatterPlotMatrix يمكن أن يظهر على لوحة تحتوي على مصفوفة من رسم مبعثر

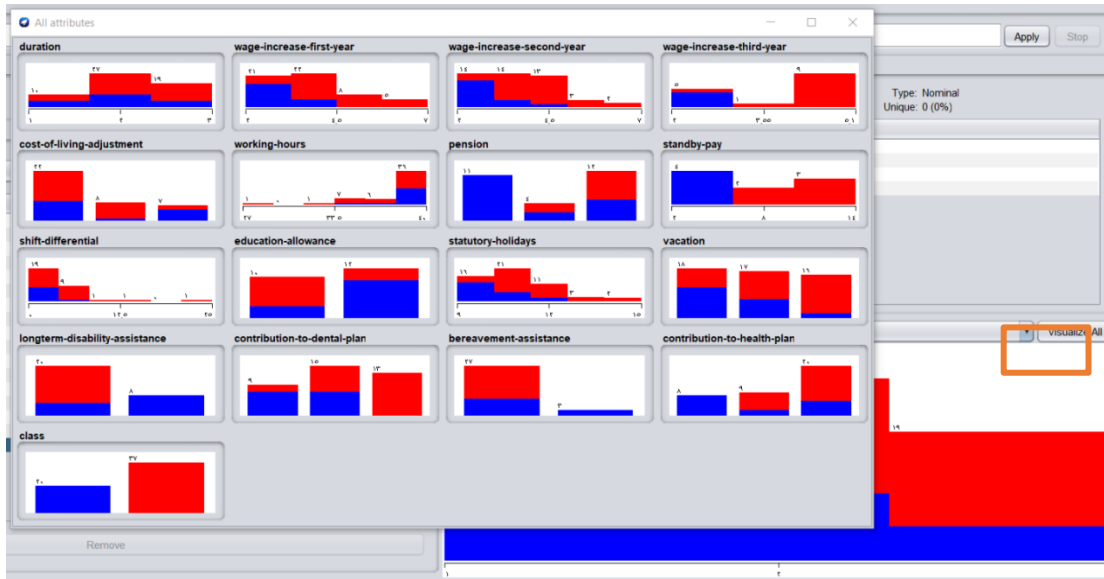


صغير وبالنقر على الرسم الصغير يظهر بشكل رسم مبعثر أكبر

وبالنقر على الرسم الصغير يظهر بشكل رسم مبعثر أكبر

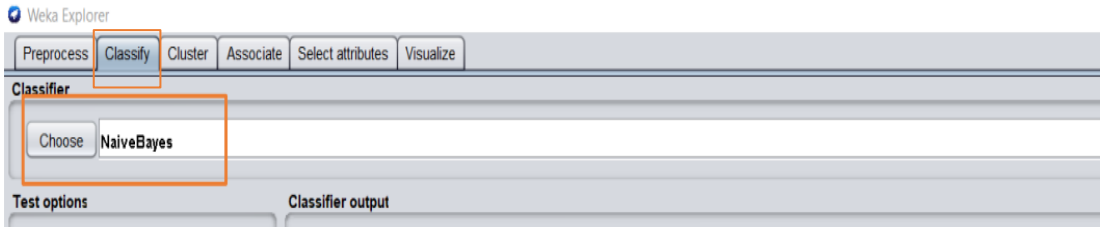


- مكون AttributeSummarizer يمكن أن يظهر على لوحة تحتوي على مصفوفة من مخططات الرسم البياني المضلع histogram واحد لكل من السمات في البيانات المدخلة



## ثالثاً: التصنيف وأداة التنبؤ

باستخدام Naive Bayes



About

A data generator that produces data randomly by producing a decision list. [More](#)

debug ☐ False

maxRuleSize 10

minRuleSize 1

numAttributes 10

**numClasses 2**

numExamples 100

numIrrelevant 0

numNumeric 0

relationName

seed 1

voteFlag ☐ False

Open... Save... OK Cancel

**Classifier output**

false	31.0	23.0
true	37.0	13.0
[total]	68.0	36.0

a5

false	35.0	16.0
true	33.0	20.0
[total]	68.0	36.0

Time taken to build model: . seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	٧٧	٧٧	%
Incorrectly Classified Instances	٢٢	٢٢	%
Kappa statistic	٠.٤٨٣٨		
Mean absolute error	٠.٢١٠٨		
Root mean squared error	٠.٢٩٢١		
Relative absolute error	١٨.١٩٢٧ %		
Root relative squared error	٨٢.١٦٠٩ %		
Total Number of Instances	١٠٠		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	٠.٨٧٠	٠.٨٢٠	٠.٤٨٤	٠.٨٢٧	٠.٨٢٢	٠.٨٢١	٠.٢٥٢	٠.٨٢٢	c0
	٠.١١٠	٠.٨٢٠	٠.٤٨٤	٠.١٥٧	٠.١١٧	٠.١١٧	٠.١١٧	٠.١١٧	c1

=== Confusion Matrix ===

a	b	<-- classified as
١١	٥٥	a = c0
٢٢	١٢	b = c1

النتائج

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

#### Naive Bayes Classifier

Attribute	Class	
	bad (0.36)	good (0.64)
=====		
duration		
mean	2.20	2
std. dev.	0.1821	0.7071
weight sum	36	20
precision	1	1
wage-increase-first-year		
mean	4.3837	2.6063
std. dev.	1.1773	0.8643
weight sum	36	20
precision	0.3120	0.3120
wage-increase-second-year		
mean	4.447	2.9024
std. dev.	0.9800	0.8193
weight sum	36	10
precision	0.3071	0.3071
wage-increase-third-year		
mean	4.0790	2.0344
std. dev.	0.7893	0.1678
weight sum	11	4
precision	0.3870	0.3870
cost-of-living-adjustment		

النتائج

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'ZeroR'. The 'Test options' section shows 'Cross-validation' with 'Folds' set to 10. The 'Classifier output' pane displays the following results:

```

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

ZeroR predicts class value: good

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      11.9112      27 %
Incorrectly Classified Instances    30.0887      70 %
Kappa statistic                     0
Mean absolute error                  0.4071
Root mean squared error              0.4770
Relative absolute error              100 %
Root relative squared error          100 %
Total Number of Instances           47

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MDC   ROC Area  PRC Area  Cl
          -----  -
          0.229    0.4712    ?          ?    0.0002    0.149    0.000    0.000    bad
          0.129    0.4712    0.7857    0.000    0.0000    0.149    0.000    0.000    good
Weighted Avg.    0.022    0.4712    ?          ?    0.1492    0.149    0.149    0.149

=== Confusion Matrix ===

 a  b  <-- classified as
20  0 | a = bad
27  0 | b = good
  
```

The 'Result list' on the left shows '1A:1:10 - rules ZeroR'.

تم تطبيق مصفوفة الشك

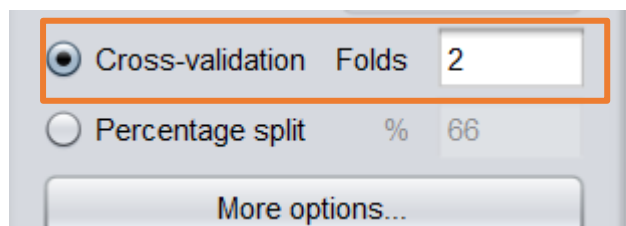
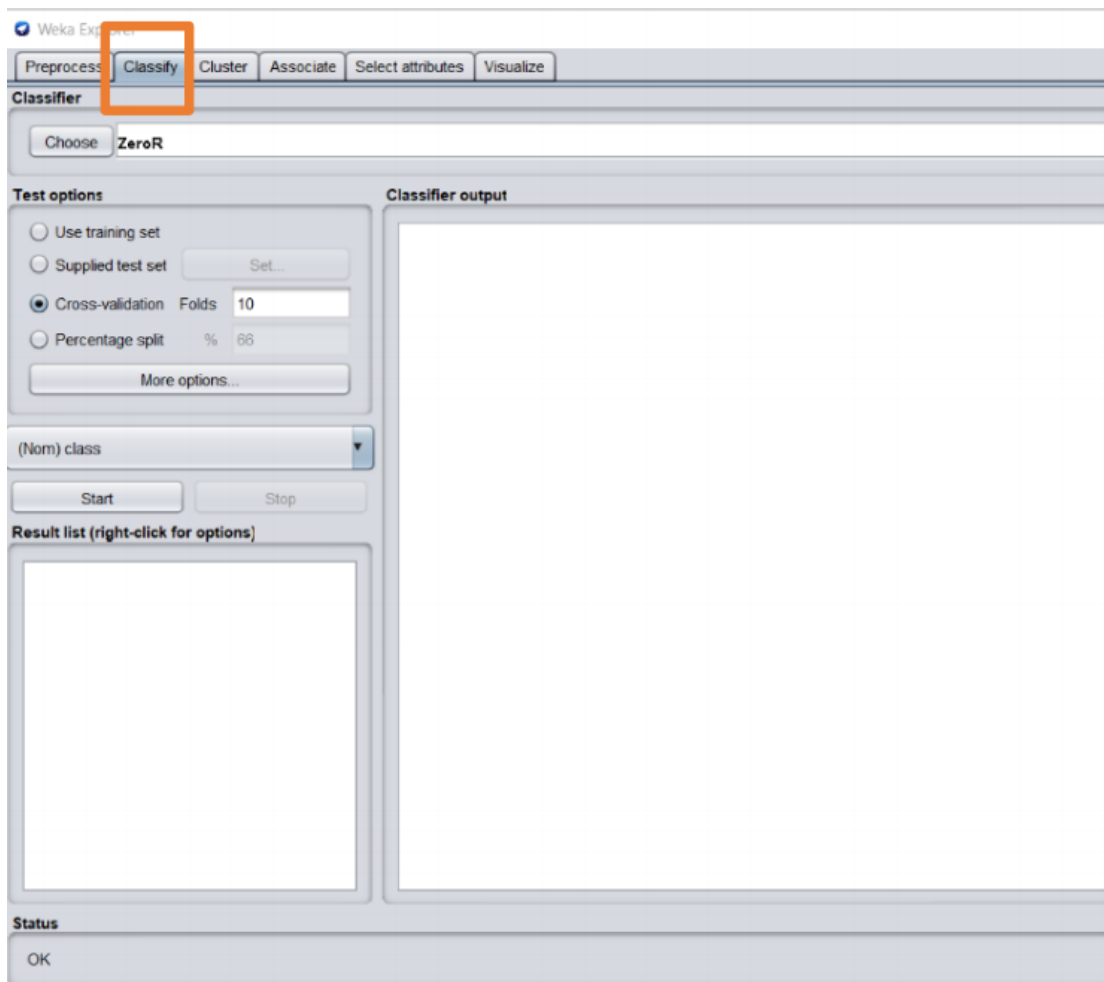
على المشروع للأمام على

كافة الجوانب التي تم

دراستها



## رابعاً: الشبكات العصبية



weka.gui.GenericObjectEditor

weka.classifiers.functions.MultilayerPerceptron

About

A classifier that uses backpropagation to learn a multi-layer perceptron to classify instances.

More

Capabilities

GUI True

autoBuild True

batchSize 100

debug False

decay False

doNotCheckCapabilities False

hiddenLayers 3

learningRate 0.3

momentum 0.2

nominalToBinaryFilter True

normalizeAttributes True

normalizeNumericClass True

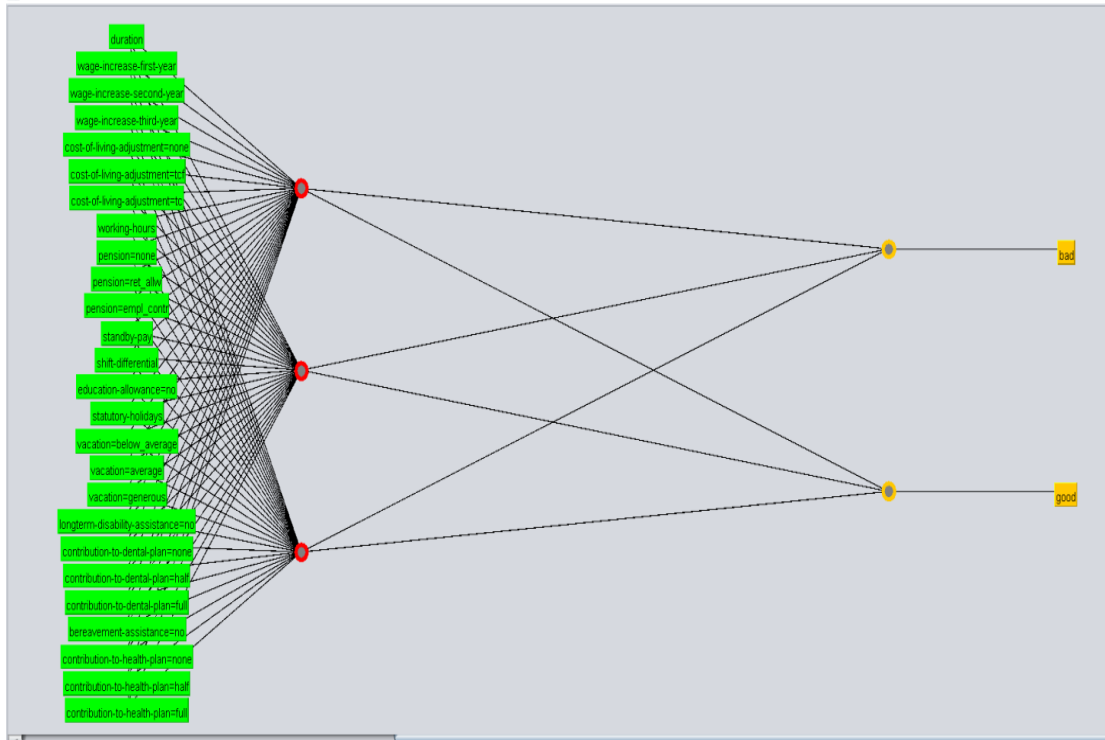
numDecimalPlaces 2

reset False

resume False

seed 0

trainingTime 500



## خامساً: التحليل العنقودي

