

Τμήμα Ψηφιακών Συστημάτων
Ακαδημαϊκό Έτος 2021 – 2022
Τεχνητή Νοημοσύνη

ΑΤΟΜΙΚΗ ΕΡΓΑΣΙΑ #2

Κωνσταντίνος Κωνσταντίνος
ΑΜ : 2026201900036

Για την εργασία μας θα χρησιμοποιήσω τα σύνολα δεδομένων για την αξιολόγηση αυτοκινήτων.

Η εργασία θα πραγματοποιηθεί στο πρόγραμμα [WEKA](#) μια σουίτα, η οποία περιέχει μια συλλογή από εργαλεία οπτικοποίησης και αλγορίθμους για την ανάλυση δεδομένων την προγνωστική μοντελοποίηση και μηχανικής μάθησης, μαζί με γραφικές διεπαφές χρήστη για εύκολη πρόσβαση σε αυτές τις λειτουργίες

Κατά αρχάς το datasheet θα πάρει μια διαμόρφωση ειδικά στα numeric χαρακτηριστικά όπου 5+ more → 5 και όπου 4+ more → 5

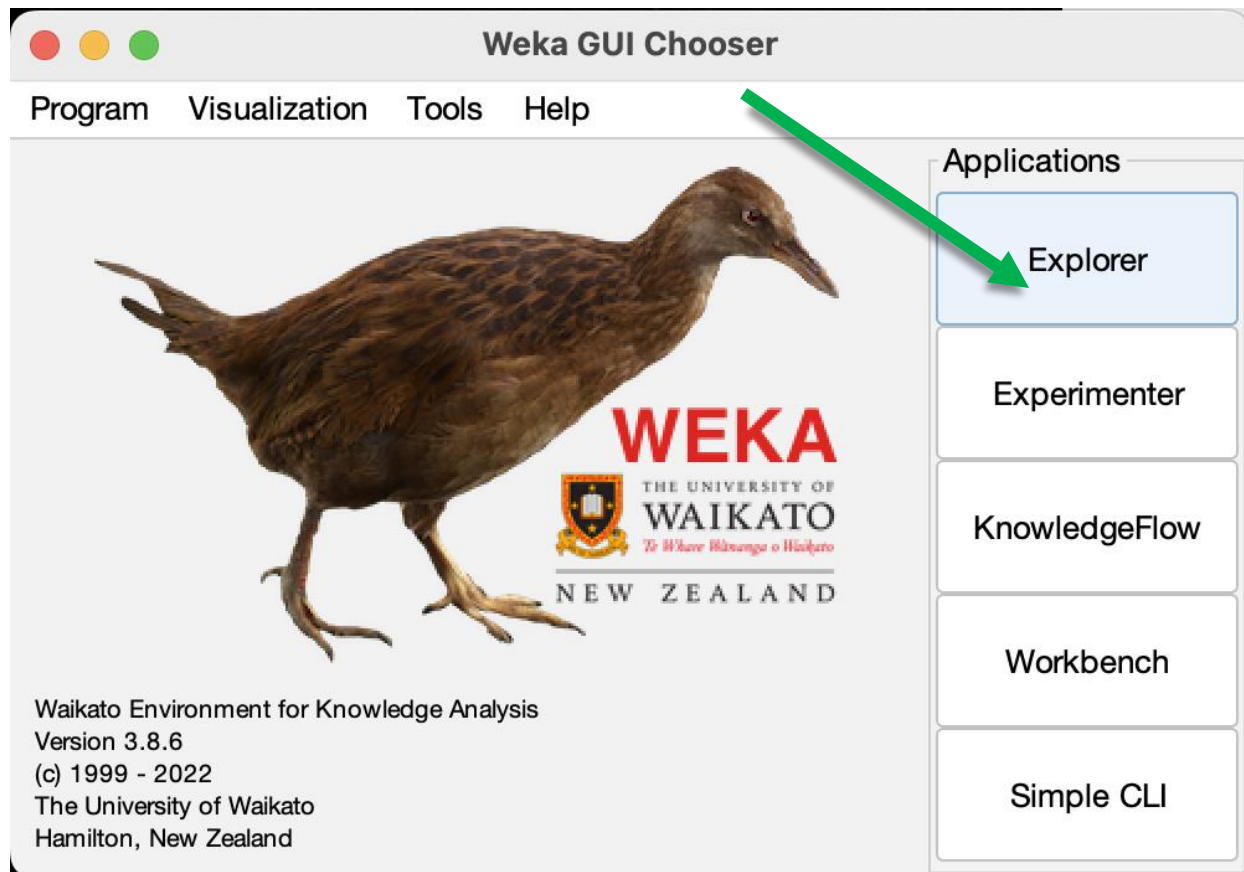
ΑΞΙΟΛΟΓΗΣΗ ΑΥΤΟΚΙΝΗΤΩΝ (CAR EVALUATION)	
Χαρακτηριστικά	Κλάση (κατηγοριοποίηση)
<ol style="list-style-type: none">Κόστος αγοράς (buying):<ol style="list-style-type: none">Πολύ υψηλό (vhigh)Υψηλό (high)Μέτριο (med)Χαμηλό (low)Κόστος συντήρησης (maint):<ol style="list-style-type: none">Πολύ υψηλό (vhigh)Υψηλό (high)Μέτριο (med)Χαμηλό (low)Αριθμός θυρών (doors):<ol style="list-style-type: none">2345 5+ (5more)Αριθμός ατόμων (persons):<ol style="list-style-type: none">245 4+ (more)Χώρος αποσκευών (lug_boot):<ol style="list-style-type: none">Μικρός (small)Μεσαίος (med)Μεγάλος (big)Ασφάλεια (safety):<ol style="list-style-type: none">Χαμηλή (low)Μέτρια (med)Υψηλή (high)	<ul style="list-style-type: none">Κατάσταση<ul style="list-style-type: none">Μη ικανοποιητική (unacc)Ικανοποιητική (acc)Καλή (good)Άριστη (very good)

Πιο συγκεκριμένα...

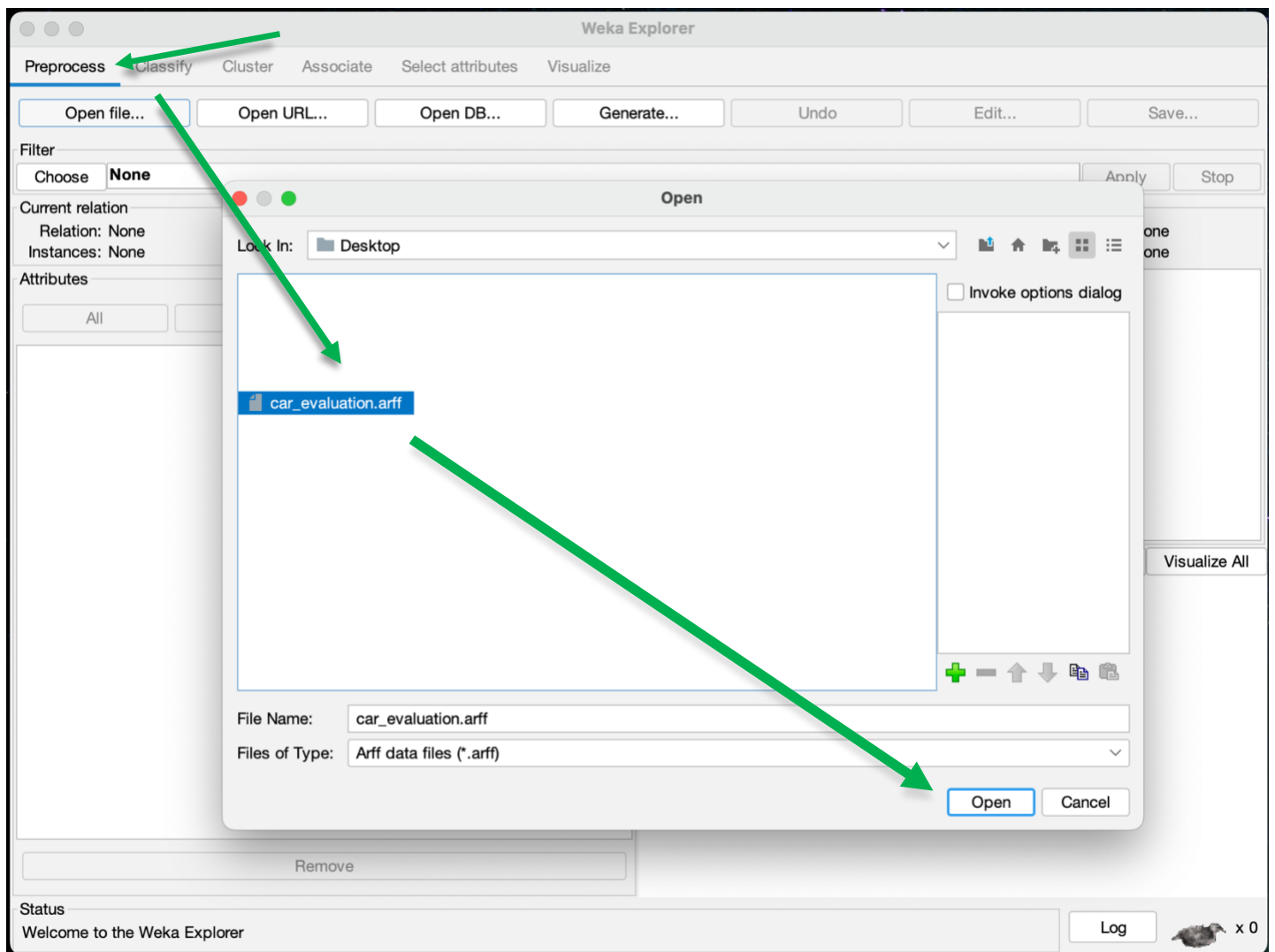
Ξεχωρίζουμε τα attributes και τα data κατά αυτόν τον τρόπο και σώζουμε το datasheet με κατάληξη .arff

```
car_evaluation.arff x
1  @relation car_evaluation
2
3  @attribute buying {vhigh,high,med,low}
4  @attribute maint {vhigh,high,med,low}
5  @attribute doors numeric
6  @attribute persons numeric
7  @attribute lug_boot {small,med,big}
8  @attribute safety {low,med,high}
9  @attribute Class {unacc,acc,good,vgood}
10
11 @data
12 vhigh,vhigh,2,2,small,low,unacc
13 vhigh,vhigh,2,2,small,med,unacc
14 vhigh,vhigh,2,2,small,high,unacc
15 vhigh,vhigh,2,2,med,low,unacc
16 vhigh,vhigh,2,2,med,med,unacc
17 vhigh,vhigh,2,2,med,high,unacc
18 vhigh,vhigh,2,2,big,low,unacc
19 vhigh,vhigh,2,2,big,med,unacc
20 vhigh,vhigh,2,2,big,high,unacc
21 vhigh,vhigh,2,4,small,low,unacc
22 vhigh,vhigh,2,4,small,med,unacc
23 vhigh,vhigh,2,4,small,high,unacc
24 vhigh,vhigh,2,4,med,low,unacc
25 vhigh,vhigh,2,4,med,med,unacc
26 vhigh,vhigh,2,4,med,high,unacc
27 vhigh,vhigh,2,4,big,low,unacc
28 vhigh,vhigh,2,4,big,med,unacc
29 vhigh,vhigh,2,4,big,high,unacc
30 vhigh,vhigh,2,more,small,low,unacc
31 vhigh,vhigh,2,more,small,med,unacc
32 vhigh,vhigh,2,more,small,high,unacc
33 vhigh,vhigh,2,more,med,low,unacc
34 vhigh,vhigh,2,more,med,med,unacc
35 vhigh,vhigh,2,more,med,high,unacc
36 vhigh,vhigh,2,more,big,low,unacc
37 vhigh,vhigh,2,more,big,med,unacc
38 vhigh,vhigh,2,more,big,high,unacc
39 vhigh,vhigh,3,2,small,low,unacc
40 vhigh,vhigh,3,2,small,med,unacc
41 vhigh,vhigh,3,2,small,high,unacc
42 vhigh,vhigh,3,2,med,low,unacc
43 vhigh,vhigh,3,2,med,med,unacc
44 vhigh,vhigh,3,2,med,high,unacc
```

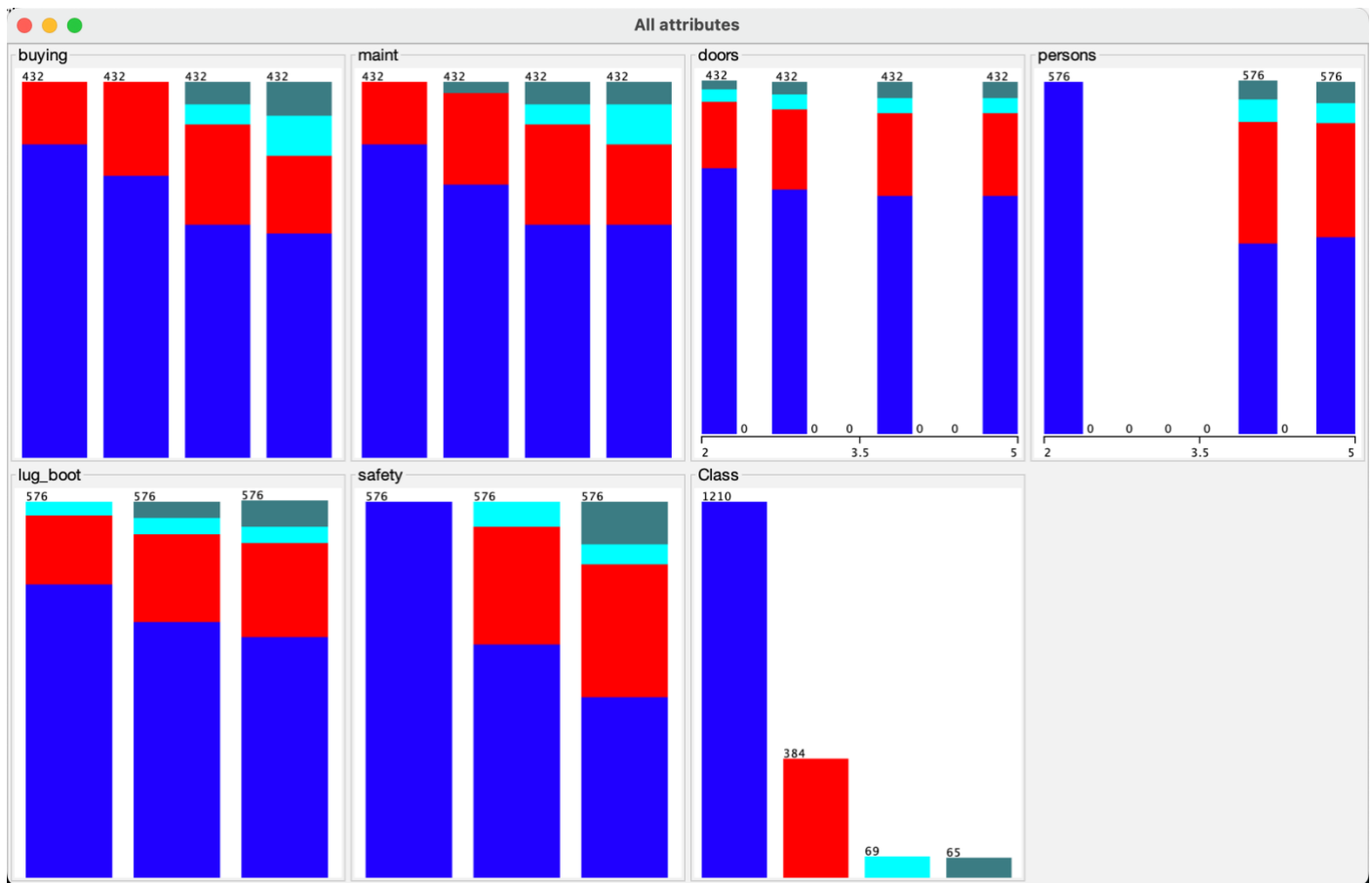
Στην συνέχεια ανοίγουμε το WEKA ➔ Explore



Φορτώνουμε το αρχείο μας...



Αρχικά μας δίνει μια πρώτη εικόνα με στατιστικά όλων των χαρακτηριστικών του datasheet μας.



Η πρώτη Παρατήρηση μας είναι στην Κλάση Κατάσταση όπου βλέπουμε μεγάλο ποσοστό Μη Ικανοποιητική. Πιο συγκεκριμένα..

Μη Ικανοποιητική : 1210

Ικανοποιητική : 384

Καλή : 69

Άριστη : 65

Selected attribute

Name: Class
Missing: 0 (0%)

Distinct: 4

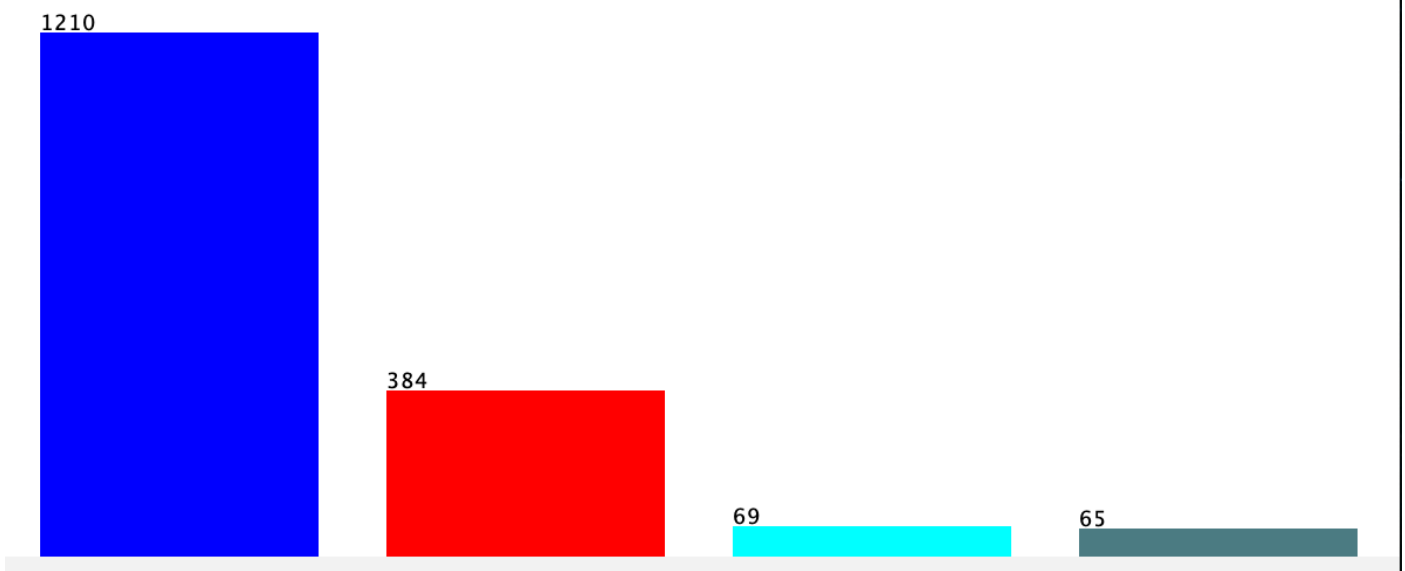
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	unacc	1210	1210
2	acc	384	384
3	good	69	69
4	vgood	65	65

Class: Class (Nom)



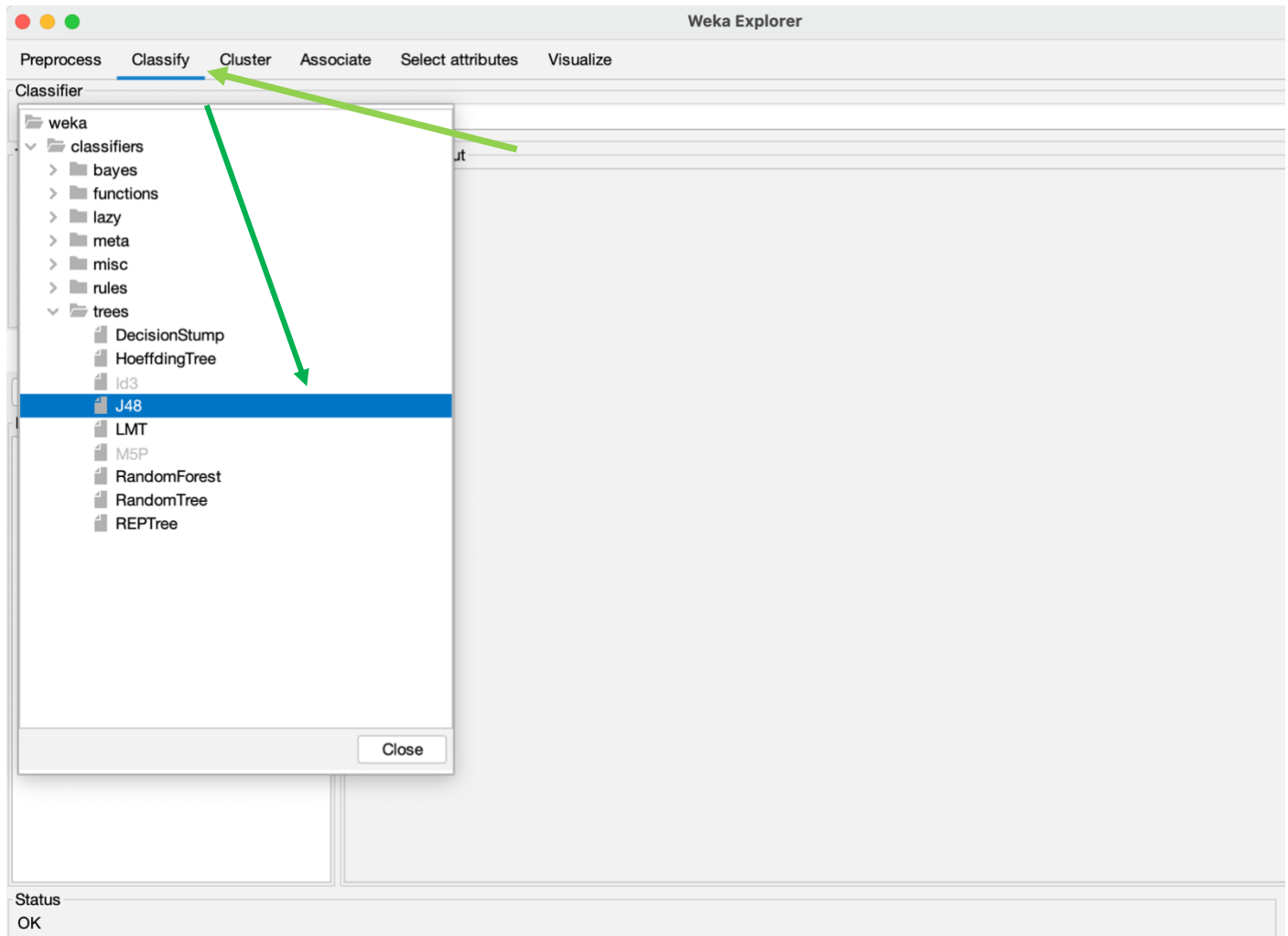
Visualize All



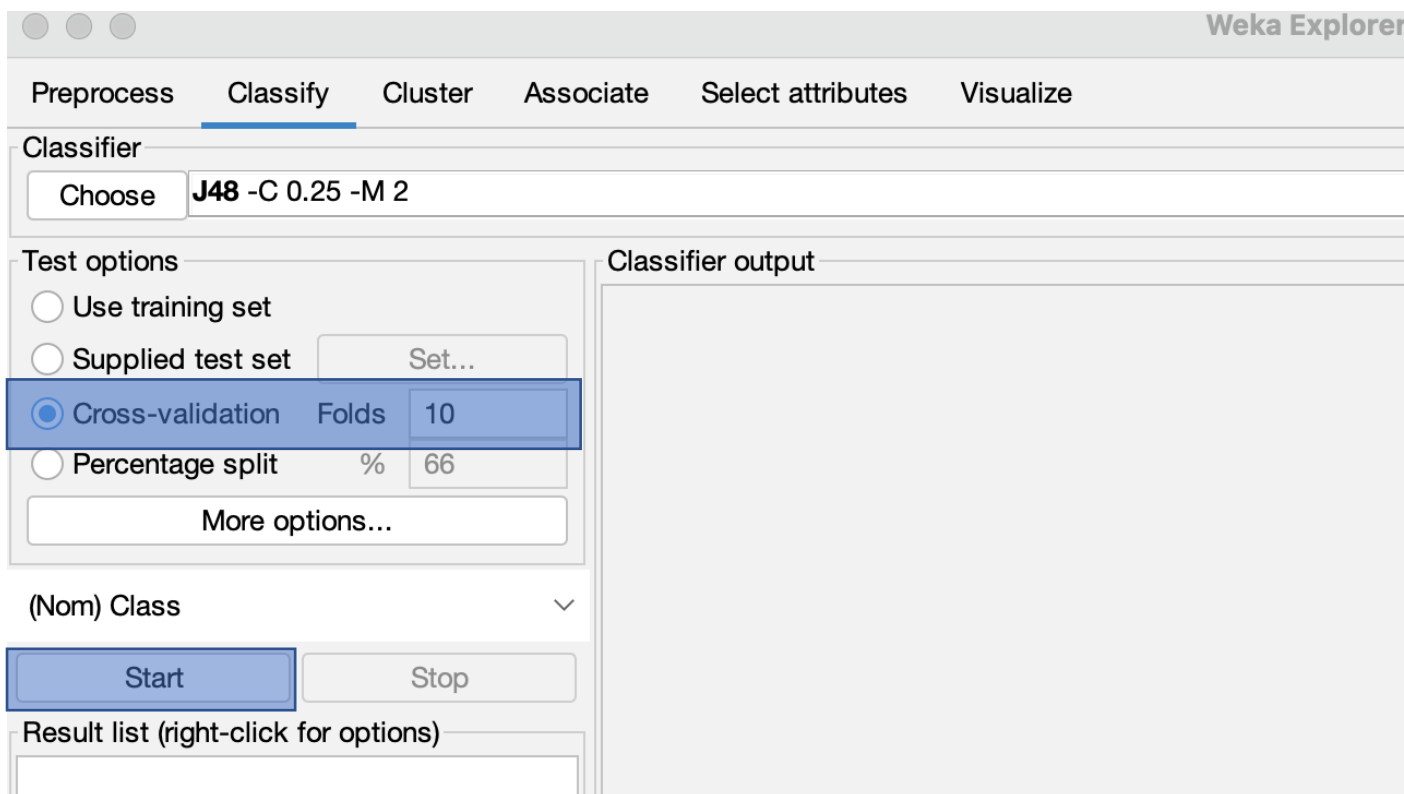
Εν συνεχεία θα τρέξουμε τον Αλγόριθμο ID3 να βγάλουμε συμπεράσματα σχετικά με τα στατιστικά του Datasheet μας.

Ως Εξής...

Πάμε Classify και επιλέγουμε από το Μενού Classifier στα Δένδρα τον Αλγόριθμο **J48** μια προχωρημένη έκδοση που χρησιμοποιεί την τεχνική του **ID3**.



Επιλέγουμε Cross Validation : 10 με Attribute Class : Κατάσταση Και **Start**



Και παίρνουμε αποτελέσματα...

Number of Leaves :	79
Size of the tree :	120

Time taken to build model: 0.04 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1665	96.3542 %
Incorrectly Classified Instances	63	3.6458 %
Kappa statistic	0.9215	
Mean absolute error	0.0254	
Root mean squared error	0.1319	
Relative absolute error	11.0993 %	
Root relative squared error	39.0029 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.974	0.014	0.994	0.974	0.984	0.949	0.980	0.990	unacc
	0.958	0.028	0.909	0.958	0.933	0.914	0.975	0.864	acc
	0.826	0.007	0.838	0.826	0.832	0.825	0.945	0.750	good
	0.938	0.005	0.884	0.938	0.910	0.907	0.968	0.905	vgood
Weighted Avg.	0.964	0.016	0.965	0.964	0.964	0.934	0.977	0.949	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1179	28	3	0	a = unacc
7	368	5	4	b = acc
0	8	57	4	c = good
0	1	3	61	d = vgood

Κατά Αρχάς βλέπουμε ένα καλό ποσοστό εγκυρότητας με 96,3542
Επίσης πολύ καλό Precision(Ακρίβεια) Recall και F-Measure
Το νούμερο των διακλαδώσεων
Και το μέγεθος του δένδρου.

Ας δούμε και μια απεικόνιση του δένδρου

Δεξι πλήκτο ➔ Visualize Tree

[illegible]

Τα αποτελέσματα που παίρνουμε ξεκαθαρίζουν όντως λίγο την εικόνα ως εξής...

Attribute Evaluator

Choose **GainRatioAttributeEval**

Search Method

Choose **Ranker -T -1.7976931348623157E308 -N -1**

Attribute Selection Mode

☒ Use full training set

☐ Cross-validation Folds 10 Seed 1

No class

Start Stop

Result list (right-click for options)

11:59:17 - Ranker + InfoGainAttributeEval

11:59:36 - Ranker + GainRatioAttributeEval

Attribute selection output

=== Run information ===

Evaluator: weka.attributeSelection.GainRatioAttributeEval
 Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
 Relation: car_evaluation
 Instances: 1728
 Attributes: 7
 buying
 maint
 doors
 persons
 lug_boot
 safety
 Class

Evaluation mode: evaluate on all training data

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

Search Method:
 Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 7 Class):
 Gain Ratio feature evaluator

Ranked attributes:

0.2387	4	persons
0.1654	6	safety
0.0482	1	buying
0.0369	2	maint
0.0189	5	lug_boot
0	3	doors

Selected attributes: 4,6,1,2,5,3 : 6

Attribute Evaluator

Choose **InfoGainAttributeEval**

Search Method

Choose **Ranker -T -1.7976931348623157E308 -N -1**

Attribute Selection Mode

☒ Use full training set

☐ Cross-validation Folds 10 Seed 1

No class

Start Stop

Result list (right-click for options)

11:59:17 - Ranker + InfoGainAttributeEval

Attribute selection output

=== Run information ===

Evaluator: weka.attributeSelection.InfoGainAttributeEval
 Search: weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
 Relation: car_evaluation
 Instances: 1728
 Attributes: 7
 buying
 maint
 doors
 persons
 lug_boot
 safety
 Class

Evaluation mode: evaluate on all training data

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

Search Method:
 Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 7 Class):
 Information Gain Ranking Filter

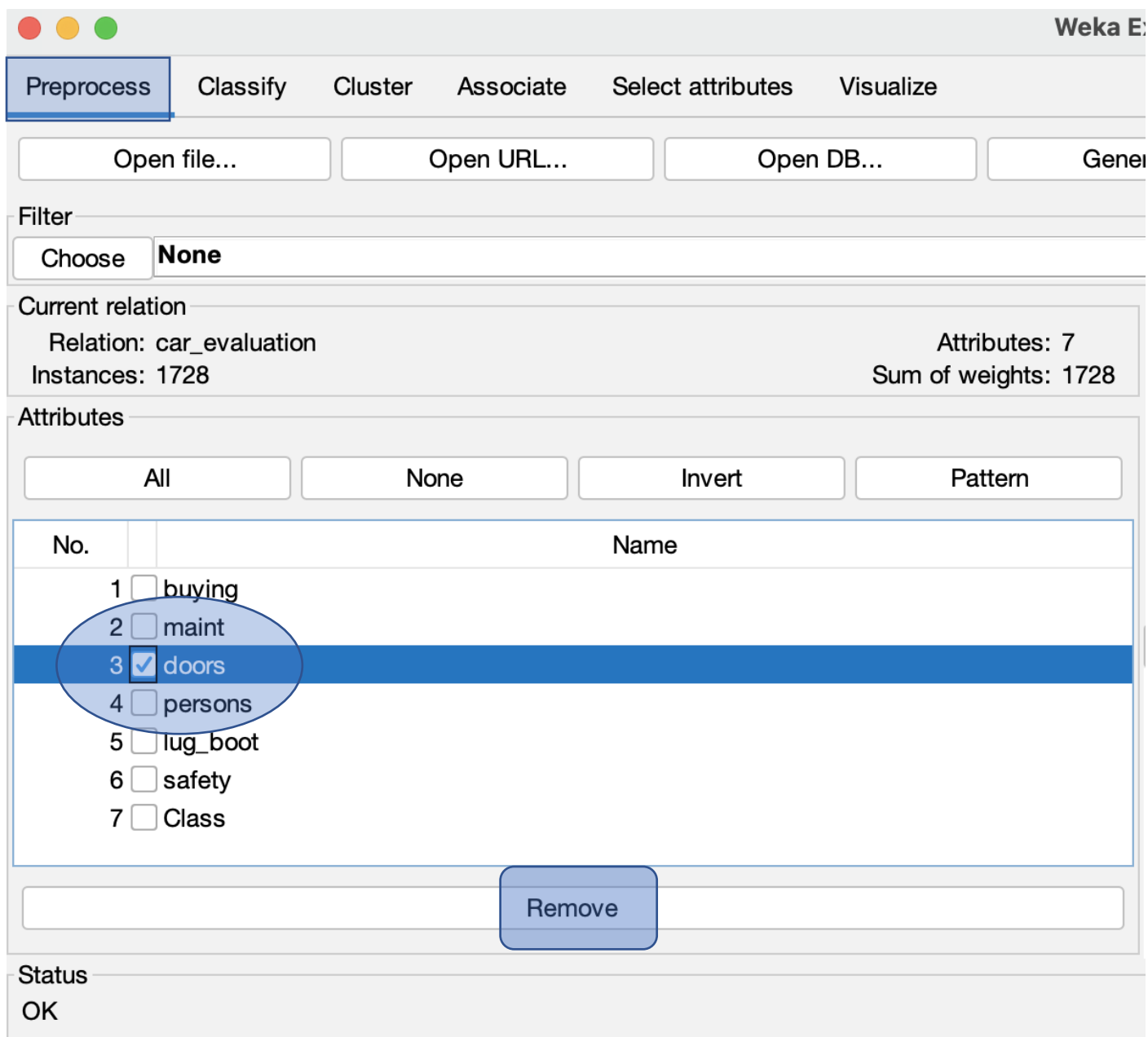
Ranked attributes:

0.2622	6	safety
0.2192	4	persons
0.0964	1	buying
0.0737	2	maint
0.03	5	lug_boot
0	3	doors

Selected attributes: 6,4,1,2,5,3 : 6

Βλέπουμε το γνώρισμα αριθμός θυρών μας δίνει μηδενική εντροπία, θα το αφαιρέσουμε και θα ξανα ξεκινήσουμε τον Αλγόριθμο μας.

Πάμε πάλι Preprocess ➔ Επιλέγουμε το Γνώρισμα ➔ Remove



The screenshot shows the Weka GUI with the 'Preprocess' tab selected. The 'Filter' dropdown is set to 'None'. The 'Current relation' section shows 'Relation: car_evaluation' and 'Instances: 1728'. The 'Attributes' section shows a list of attributes: buying, maint, doors, persons, lug_boot, safety, and Class. The 'doors' attribute is selected and circled in blue. The 'Remove' button is visible at the bottom.

No.	Name
1	<input type="checkbox"/> buying
2	<input type="checkbox"/> maint
3	<input checked="" type="checkbox"/> doors
4	<input type="checkbox"/> persons
5	<input type="checkbox"/> lug_boot
6	<input type="checkbox"/> safety
7	<input type="checkbox"/> Class

Τα αποτελέσματα μας μετά την αφαίρεση του γνωρίσματος και εκκίνησης πάλι του Αλγορίθμου.

Number of Leaves : 65

Size of the tree : 92

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1633	94.5023 %
Incorrectly Classified Instances	95	5.4977 %
Kappa statistic	0.8818	
Mean absolute error	0.0331	
Root mean squared error	0.1416	
Relative absolute error	14.4714 %	
Root relative squared error	41.879 %	
Total Number of Instances	1728	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.965	0.031	0.986	0.965	0.976	0.922	0.985	0.994	unacc
	0.924	0.037	0.877	0.924	0.900	0.871	0.984	0.889	acc
	0.739	0.008	0.785	0.739	0.761	0.752	0.979	0.811	good
	0.908	0.009	0.797	0.908	0.849	0.845	0.997	0.927	vgood
Weighted Avg.	0.945	0.031	0.947	0.945	0.946	0.901	0.985	0.961	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
1168	39	3	0	a = unacc
16	355	7	6	b = acc
0	9	51	9	c = good
0	2	4	59	d = vgood

Το εργαλείο έχει άπειρες δυνατότητες και πληθώρα αλγορίθμων για δοκιμή

Την εργασία την συνοδεύω και με αρχείο βίντεο .mon με την όλη διαδικασία παραπάνω (εν συντομία) καθώς και το τροποποιημένο .arff αρχείο συμβατό με το WEKA του Datasheet μας.

ΤΕΛΟΣ