Εξόρυξη δεδομένων με χρήση τεχνικών μηχανικής μάθησης

Τμ. Μηχανικών Πληροφορικής και Υπολογιστών

Δέντρα Αποφάσεων – Παράδειγμα

HEALTH DECISION TREE

Παράδειγμα- *HEALTH DECISION TREE*

Ο στόχος της άσκησης είναι η διερεύνηση και η εξοικείωση με την τεχνική της κατηγοριοποίησης (classification) μέσω δέντρων αποφάσεων.

Το dataset που θα χρησιμοποιήσουμε περιέχει μετρήσεις συγκεκριμένων δεικτών του αίματος και την εμφάνιση ή όχι καρδιολογικών προβλημάτων.

Παράδειγμα- *HEALTH DECISION TREE*

Το dataset (HealthDataSet_Training) περιέχει τις παρακάτω στήλες

- Age
- Marital_Status: (Single=0, Married=1, Divorced=2, Widowed=3)
- Gender
- Weight_Category
- Cholesterol
- Stress_Management
- Trait_Anxiety
- Heart_Problems

Παράδειγμα- *HEALTH DECISION TREE*

Το βασικό ζητούμενο είναι να δημιουργήσετε το δέντρο αποφάσεων σχετικά με το αν κάποιος εμφανίζει καρδιολογικά προβλήματα ή όχι. Στην συνέχεια θα πρέπει να χρησιμοποιήσετε το αρχείο «HealthDataSet_Scoring» προκειμένου να προβλέψετε αν οι συγκεκριμένοι ασθενείς θα εμφανίσουν κάποιο καρδιολογικό πρόβλημα.

Παράδειγμα- *HEALTH DECISION TREE*

<u>Φάση 1.</u>

Στην πρώτη φάση της εργασίες θα θέλαμε να μας παρουσιάσετε τα βήματα που πραγματοποιήσατε στο στάδιο της προετοιμασίας των δεδομένων καθώς επίσης και την αρχική (πριν την προετοιμασία) και την τελική μορφή των δεδομένων (μετά το στάδιο της προετοιμασίας).

Παράδειγμα- *HEALTH DECISION TREE*

<u>Φάση 2.</u>

Στην δεύτερη φάση θα θέλαμε να εφαρμόσετε τον αλγόριθμο του δέντρου αποφάσεων αλλάζοντας κάθε φορά το κριτήριο (gain_ration, gini_index και accuracy) και τις παραμέτρους leaf size, size of split,... Σε αυτό το στάδιο θα θέλαμε να μας παρουσιάσετε πως διαμορφώνονται τα Performance Vectors βάσει των αλλαγών που κάνετε κάθε φορά στα κριτήρια. Τέλος θα θέλαμε να μας παρουσιάσετε τα κριτήρια που επιλέξατε προκειμένου να χτίσετε το τελικό δέντρο αποφάσεων.

Παράδειγμα- *HEALTH DECISION TREE*

<u>Φάση 3.</u>

Στην τρίτη φάση θα θέλαμε να προβλέψετε αν οι ασθενείς θα εμφανίσουν καρδιολογικά προβλήματα χρησιμοποιώντας το αρχείο «HealthDataSet_Scoring» και να αξιολογήσετε την αντίστοιχη απάντηση βάσει του Confidence.

Παράδειγμα- *HEALTH DECISION TREE*

Παρουσίαση Προτεινόμενης Λύσης

Παράδειγμα- *HEALTH DECISION TREE*

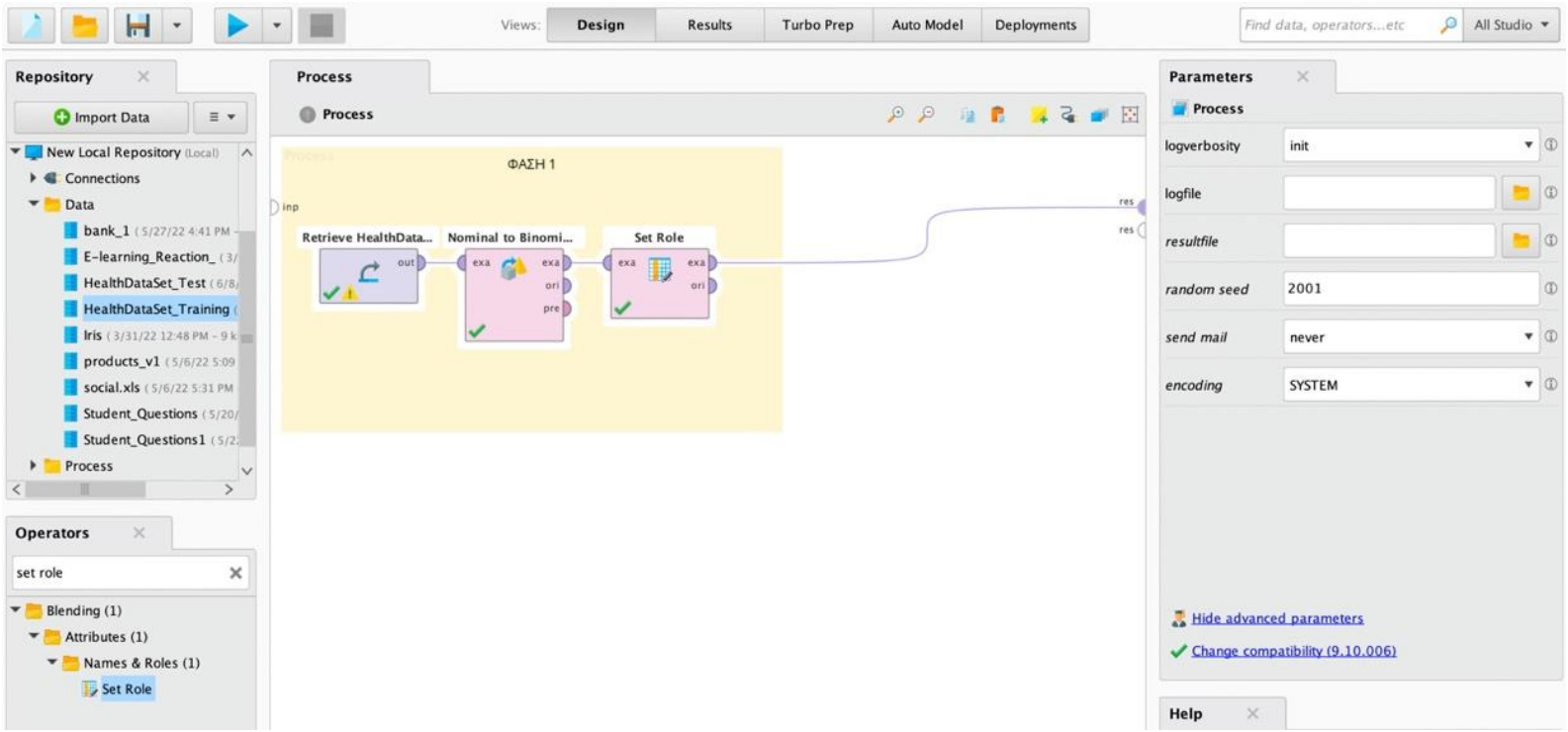
- **ΦΑΣΗ 1**

Στη Φάση 1 χρησιμοποιούμε τον oretator

- **Nominal to Binominal**, για να μετατρέψουμε τις τιμές της στήλης «Heart_Problems» από νούμερα σε τιμές τύπου true/false και
- **«Set Role»**, για να δείξουμε ότι η στήλη «Heart_Problems» έχει μία ιδιότητα βάζοντας ως target role το label που δείχνει το true/false.

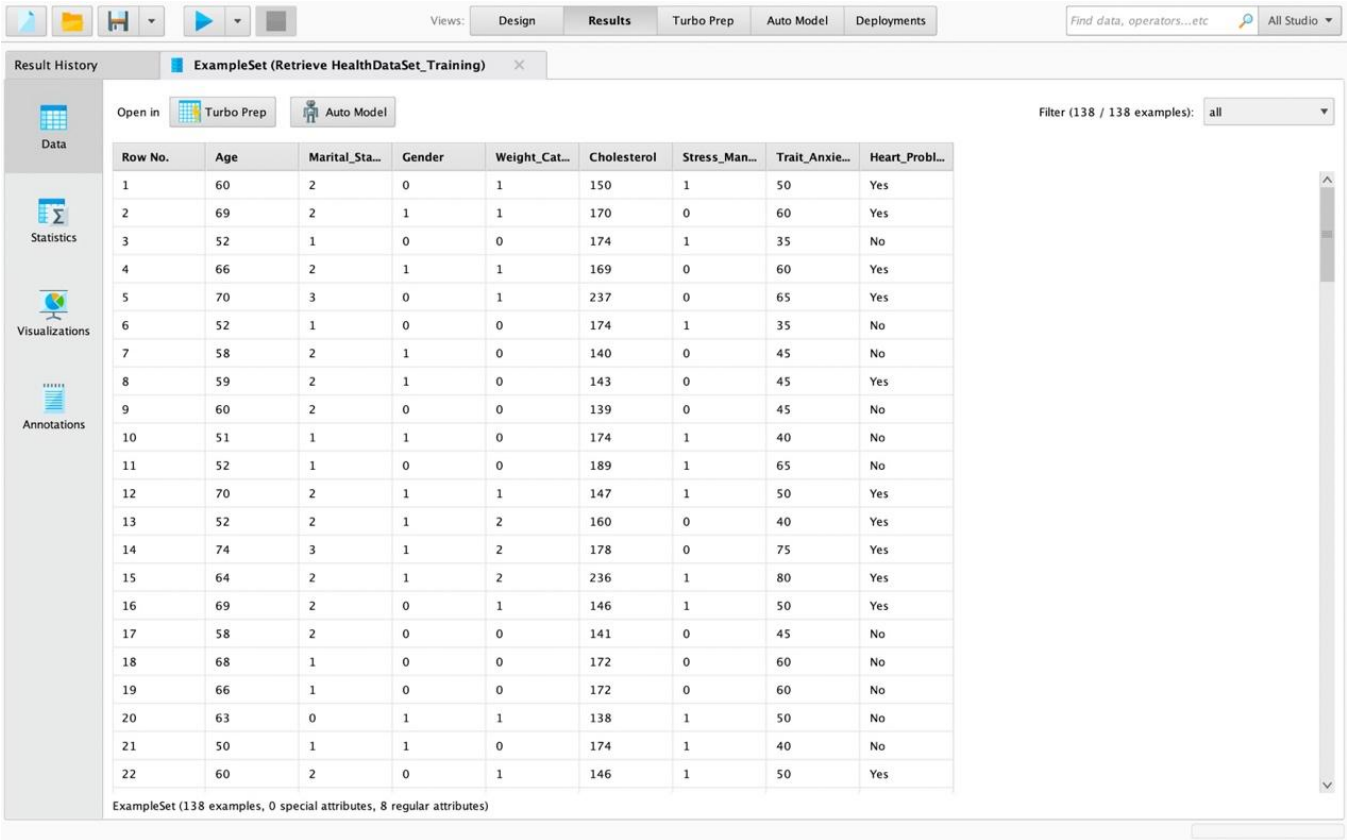
Παράδειγμα- *HEALTH DECISION TREE*

Στιγμιότυπο οθόνης με τα βήματα για **την προετοιμασία των δεδομένων (Data Preparation)**



Παράδειγμα- *HEALTH DECISION TREE*

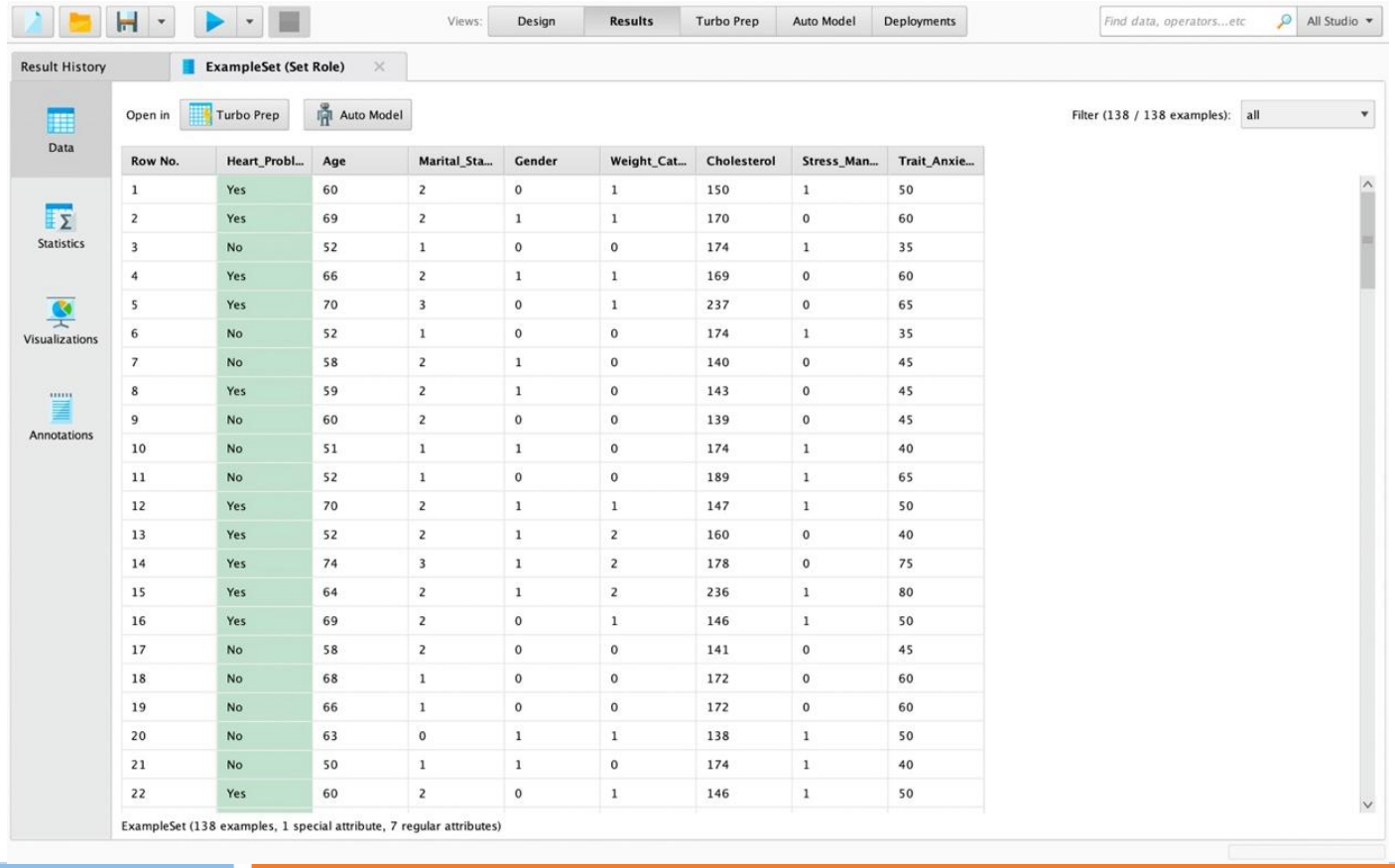
Αρχική μορφή των δεδομένων (πριν την προετοιμασία)



Row No.	Age	Marital_Sta...	Gender	Weight_Cat...	Cholesterol	Stress_Man...	Trait_Anxie...	Heart_Probl...
1	60	2	0	1	150	1	50	Yes
2	69	2	1	1	170	0	60	Yes
3	52	1	0	0	174	1	35	No
4	66	2	1	1	169	0	60	Yes
5	70	3	0	1	237	0	65	Yes
6	52	1	0	0	174	1	35	No
7	58	2	1	0	140	0	45	No
8	59	2	1	0	143	0	45	Yes
9	60	2	0	0	139	0	45	No
10	51	1	1	0	174	1	40	No
11	52	1	0	0	189	1	65	No
12	70	2	1	1	147	1	50	Yes
13	52	2	1	2	160	0	40	Yes
14	74	3	1	2	178	0	75	Yes
15	64	2	1	2	236	1	80	Yes
16	69	2	0	1	146	1	50	Yes
17	58	2	0	0	141	0	45	No
18	68	1	0	0	172	0	60	No
19	66	1	0	0	172	0	60	No
20	63	0	1	1	138	1	50	No
21	50	1	1	0	174	1	40	No
22	60	2	0	1	146	1	50	Yes

Παράδειγμα- *HEALTH DECISION TREE*

Τελική **μορφή των δεδομένων (μετά την προετοιμασία)**



Row No.	Heart_Probl...	Age	Marital_Sta...	Gender	Weight_Cat...	Cholesterol	Stress_Man...	Trait_Anxie...
1	Yes	60	2	0	1	150	1	50
2	Yes	69	2	1	1	170	0	60
3	No	52	1	0	0	174	1	35
4	Yes	66	2	1	1	169	0	60
5	Yes	70	3	0	1	237	0	65
6	No	52	1	0	0	174	1	35
7	No	58	2	1	0	140	0	45
8	Yes	59	2	1	0	143	0	45
9	No	60	2	0	0	139	0	45
10	No	51	1	1	0	174	1	40
11	No	52	1	0	0	189	1	65
12	Yes	70	2	1	1	147	1	50
13	Yes	52	2	1	2	160	0	40
14	Yes	74	3	1	2	178	0	75
15	Yes	64	2	1	2	236	1	80
16	Yes	69	2	0	1	146	1	50
17	No	58	2	0	0	141	0	45
18	No	68	1	0	0	172	0	60
19	No	66	1	0	0	172	0	60
20	No	63	0	1	1	138	1	50
21	No	50	1	1	0	174	1	40
22	Yes	60	2	0	1	146	1	50

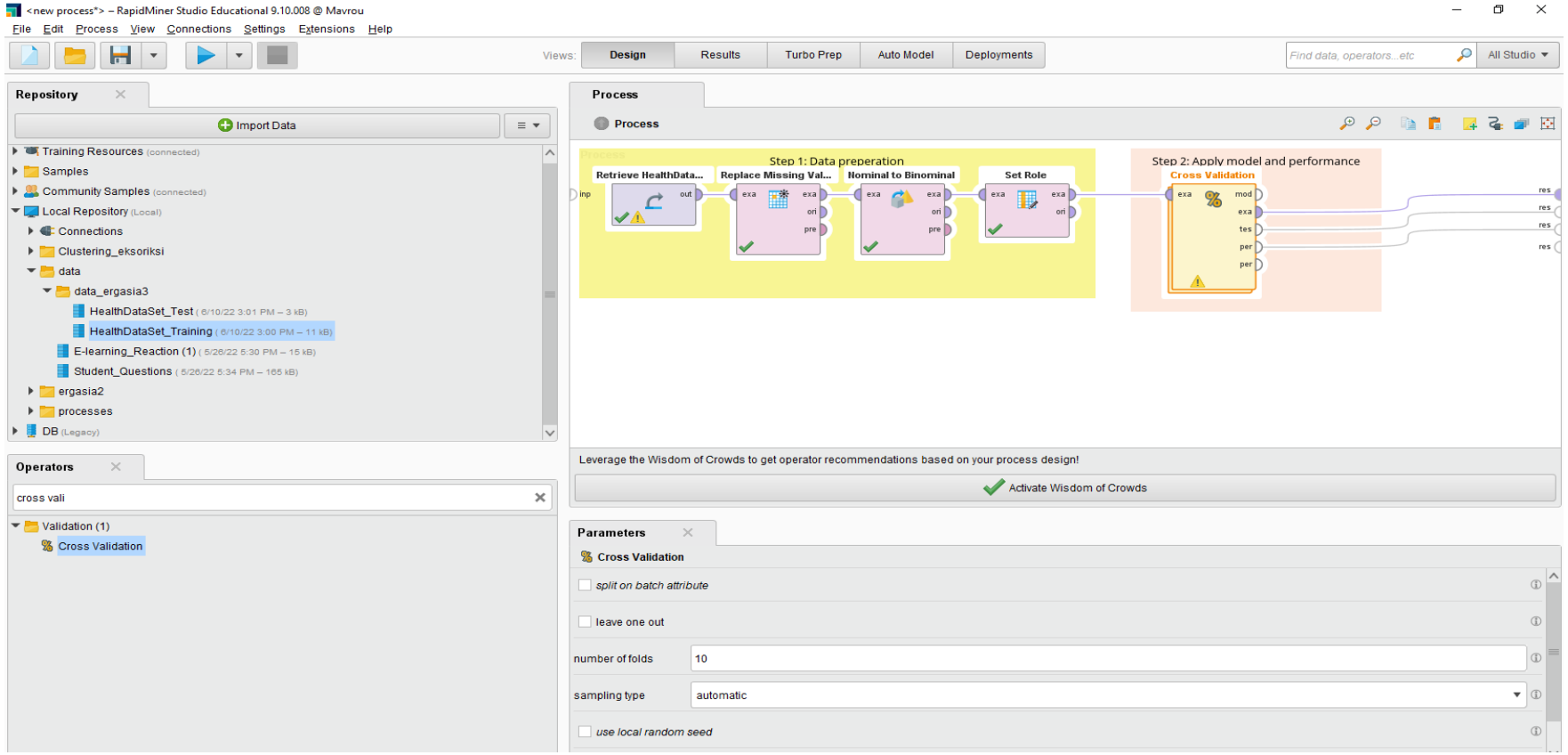
Παράδειγμα- *HEALTH DECISION TREE*

Φάση 2: Εφαρμογή αλγορίθμου

Χρησιμοποιήθηκε ο operator Cross validation και στη συνέχεια δημιουργήθηκε το δένδρο αποφάσεων

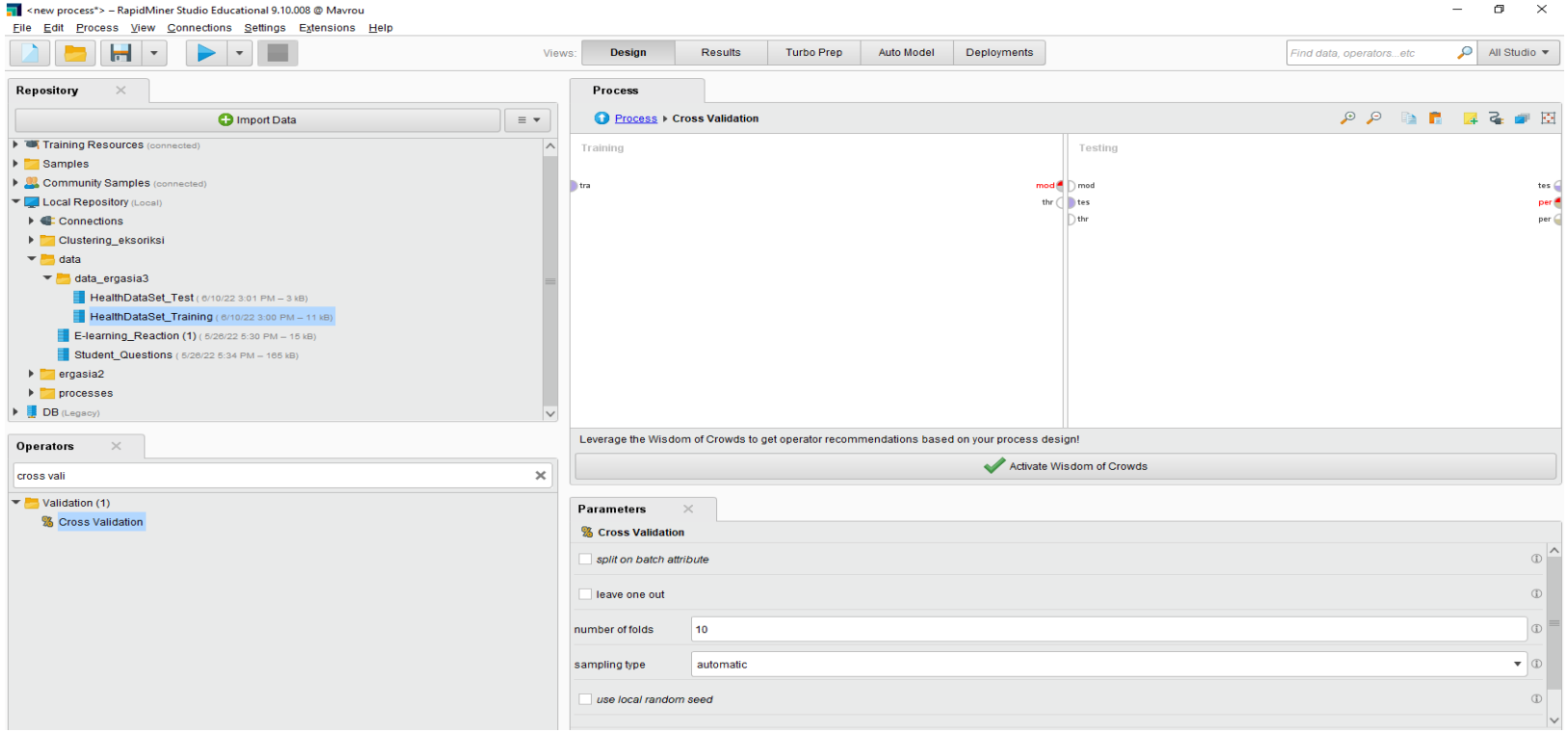
Παράδειγμα- *HEALTH DECISION TREE*

Φάση 2: Εφαρμογή αλγορίθμου



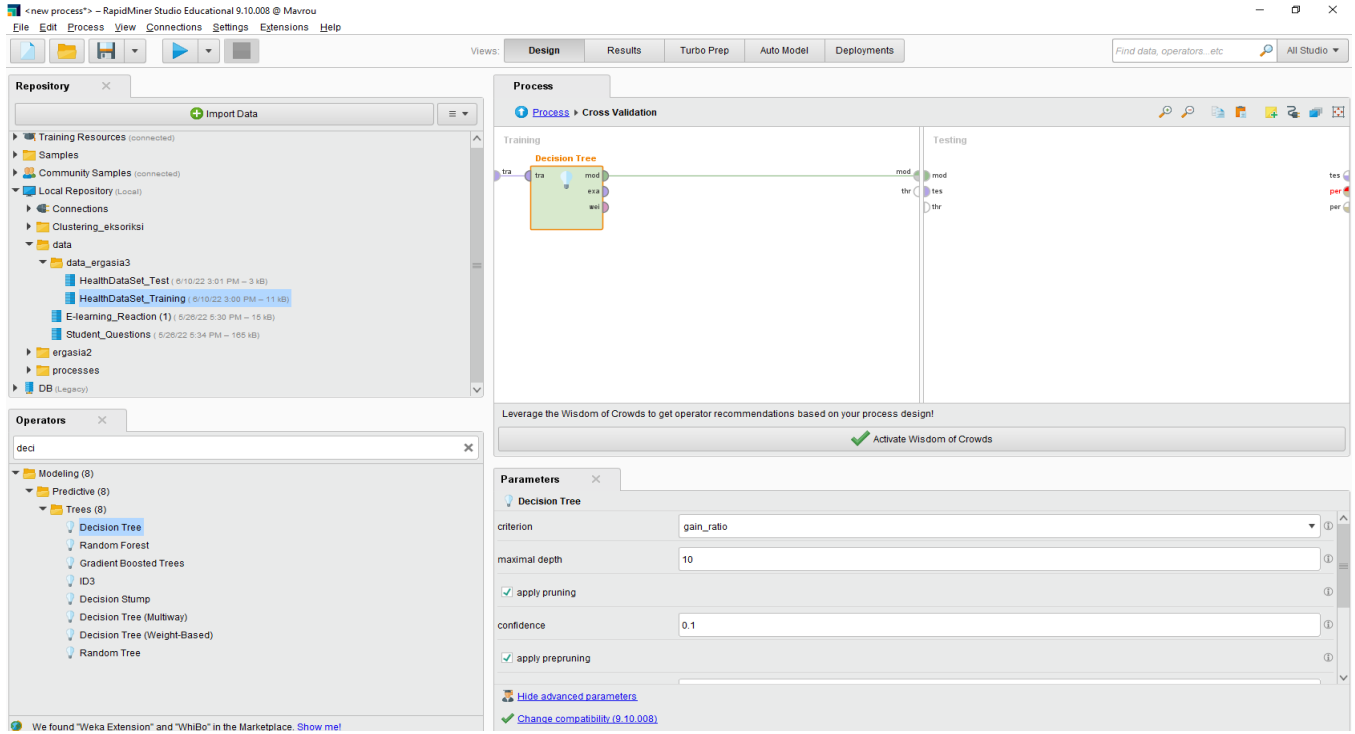
Παράδειγμα- *HEALTH DECISION TREE*

Αφού κάνουμε την εισαγωγή του operator και τις κατάλληλες συνδέσεις, εισάγουμε τον operator του δένδρου



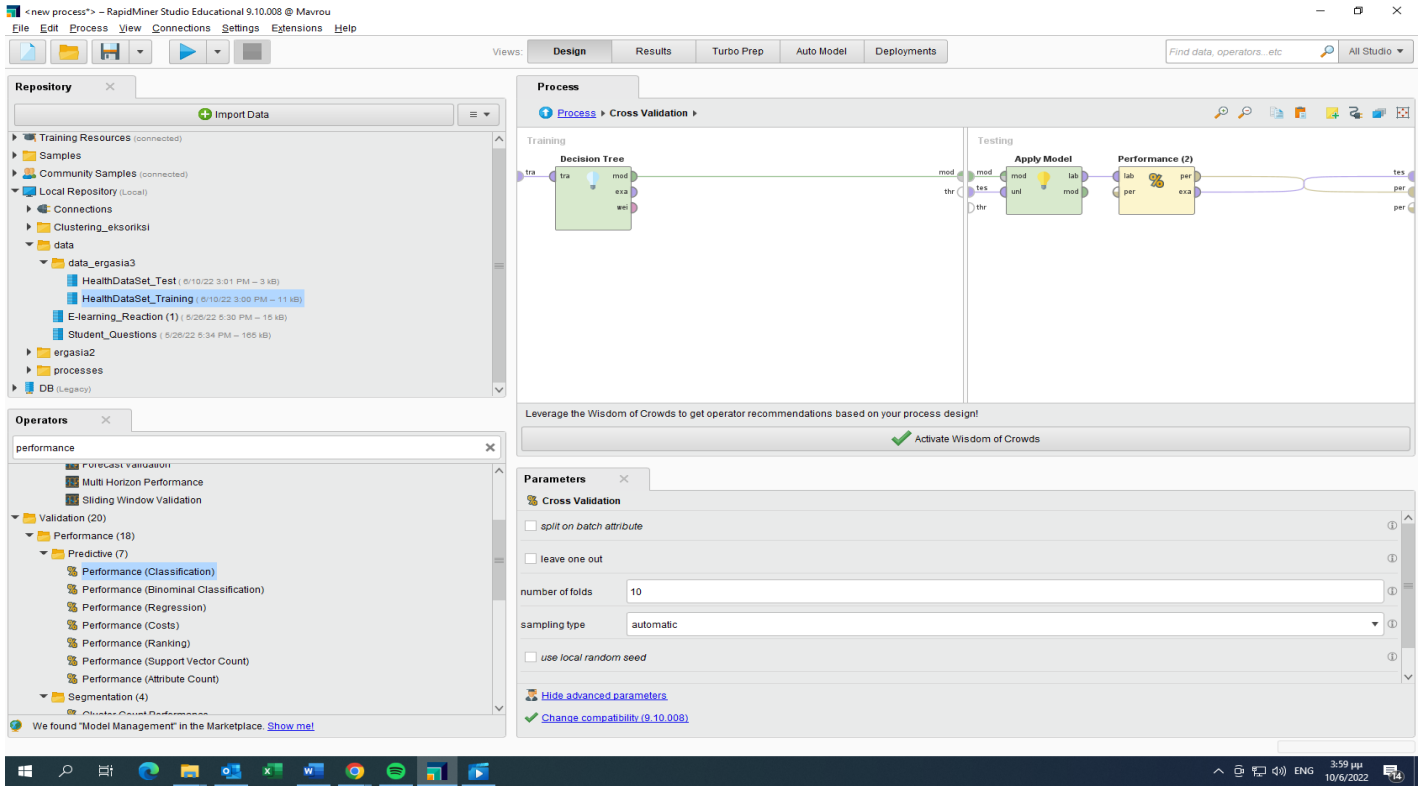
Παράδειγμα- *HEALTH DECISION TREE*

Αφού κάνουμε την εισαγωγή του operator και τις κατάλληλες συνδέσεις, εισάγουμε τον operator του δένδρου



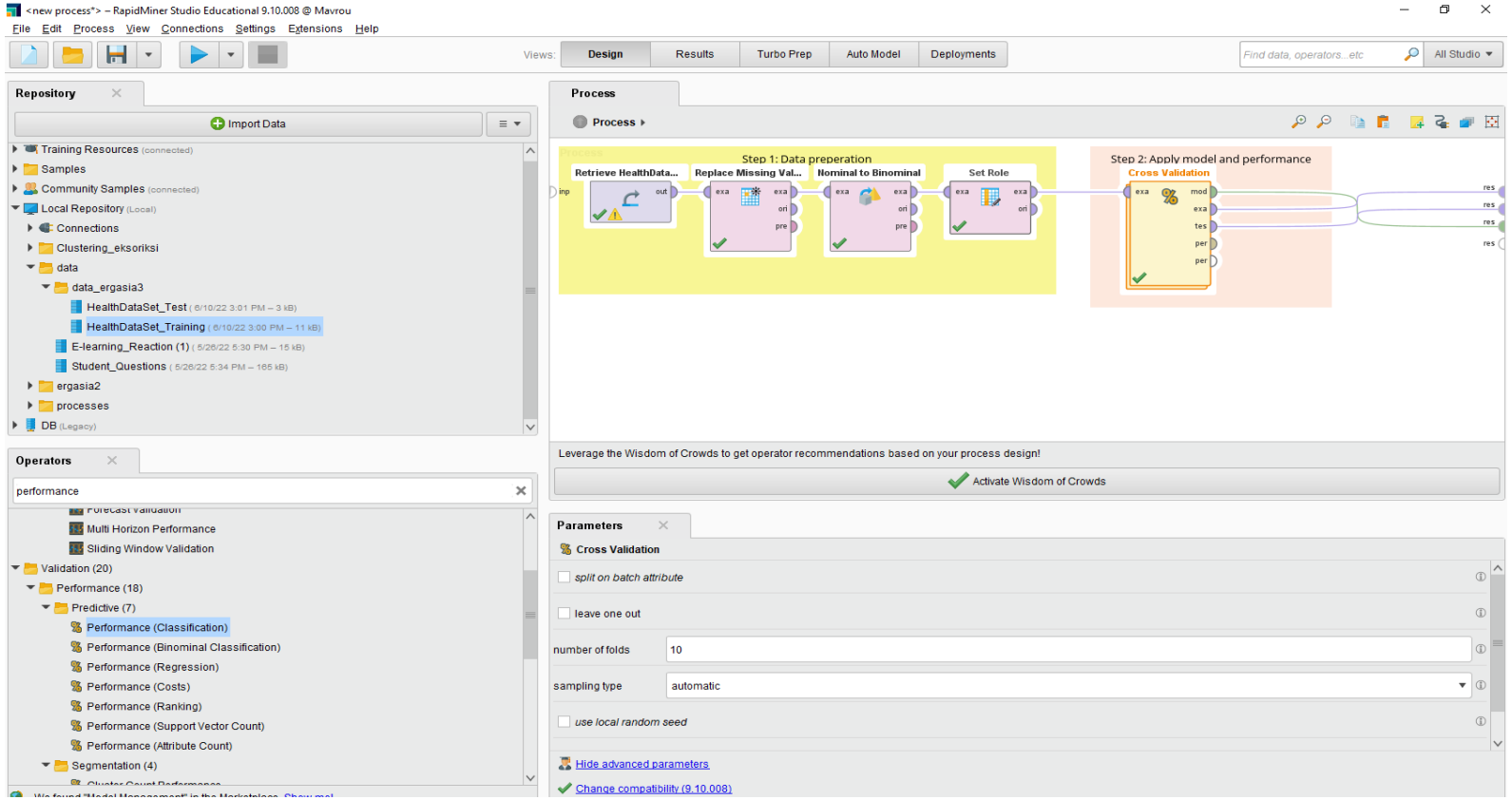
Παράδειγμα- *HEALTH DECISION TREE*

Αρχικά έχουμε επιλέξει 10 «ρίζες» και το είδος δειγματοληψίας στο αυτόματο.



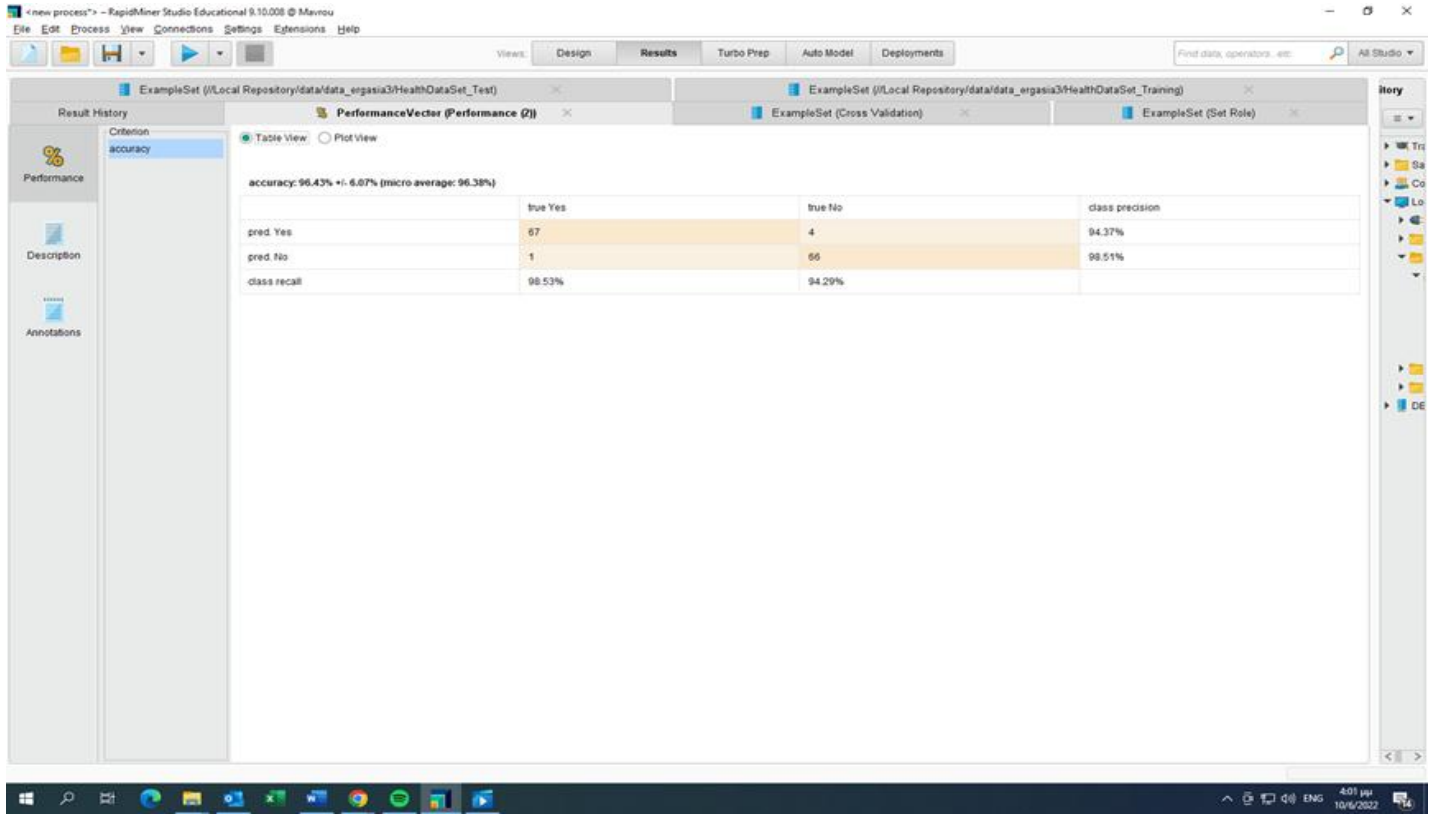
Παράδειγμα- *HEALTH DECISION TREE*

Αρχικά έχουμε επιλέξει 10 «ρίζες» και το είδος δειγματοληψίας στο αυτόματο.



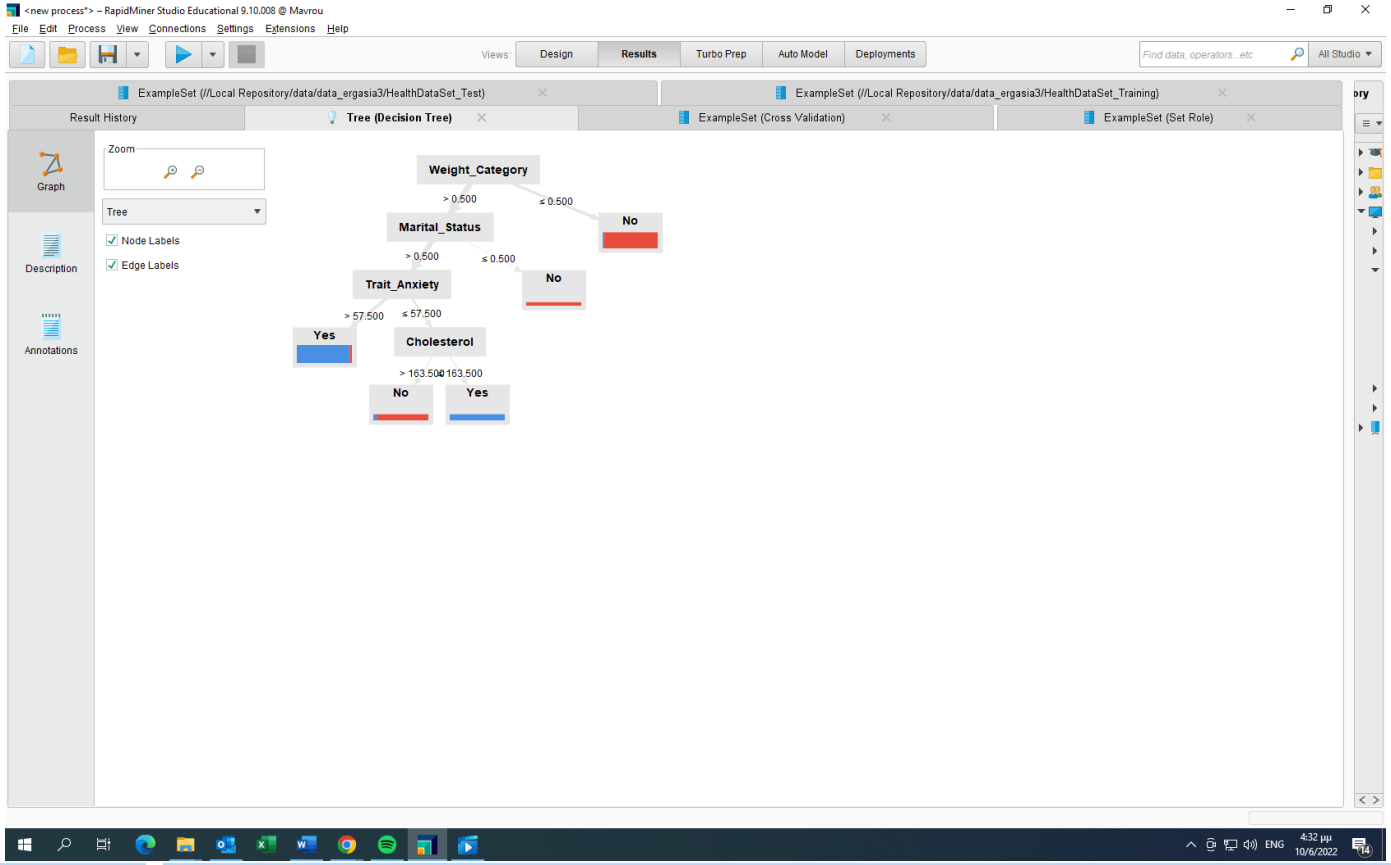
Παράδειγμα- *HEALTH DECISION TREE*

Εδώ φαίνονται τα πρώτα αποτελέσματα .



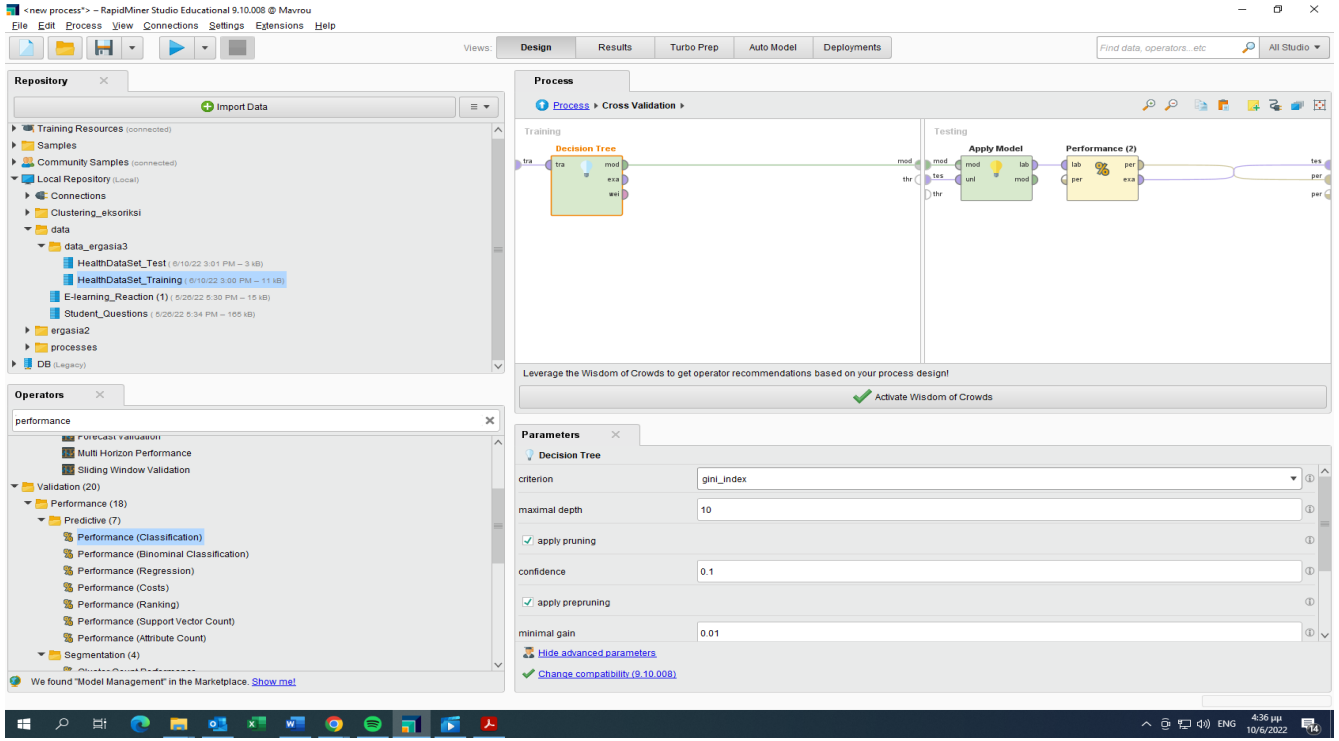
Παράδειγμα- *HEALTH DECISION TREE*

Εδώ φαίνονται τα πρώτα αποτελέσματα .



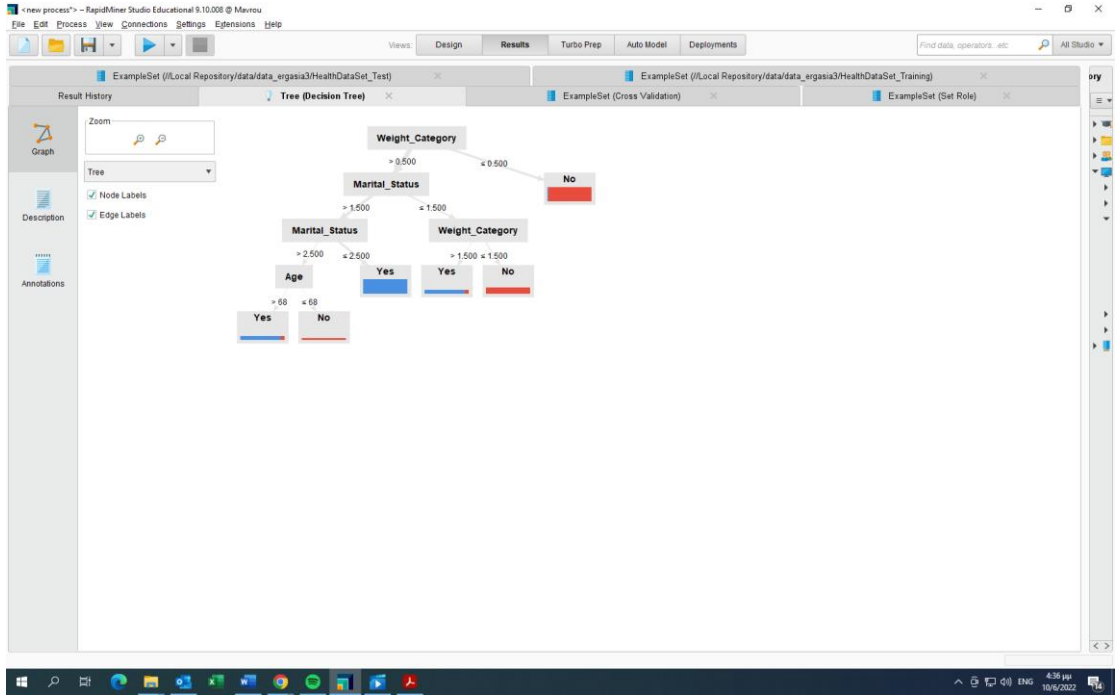
Παράδειγμα- *HEALTH DECISION TREE*

Στη συνέχεια αλλάζουμε τις παραμέτρους για να δούμε τις διαφορές και να επιλέξουμε Gini index (maximal depth 10)



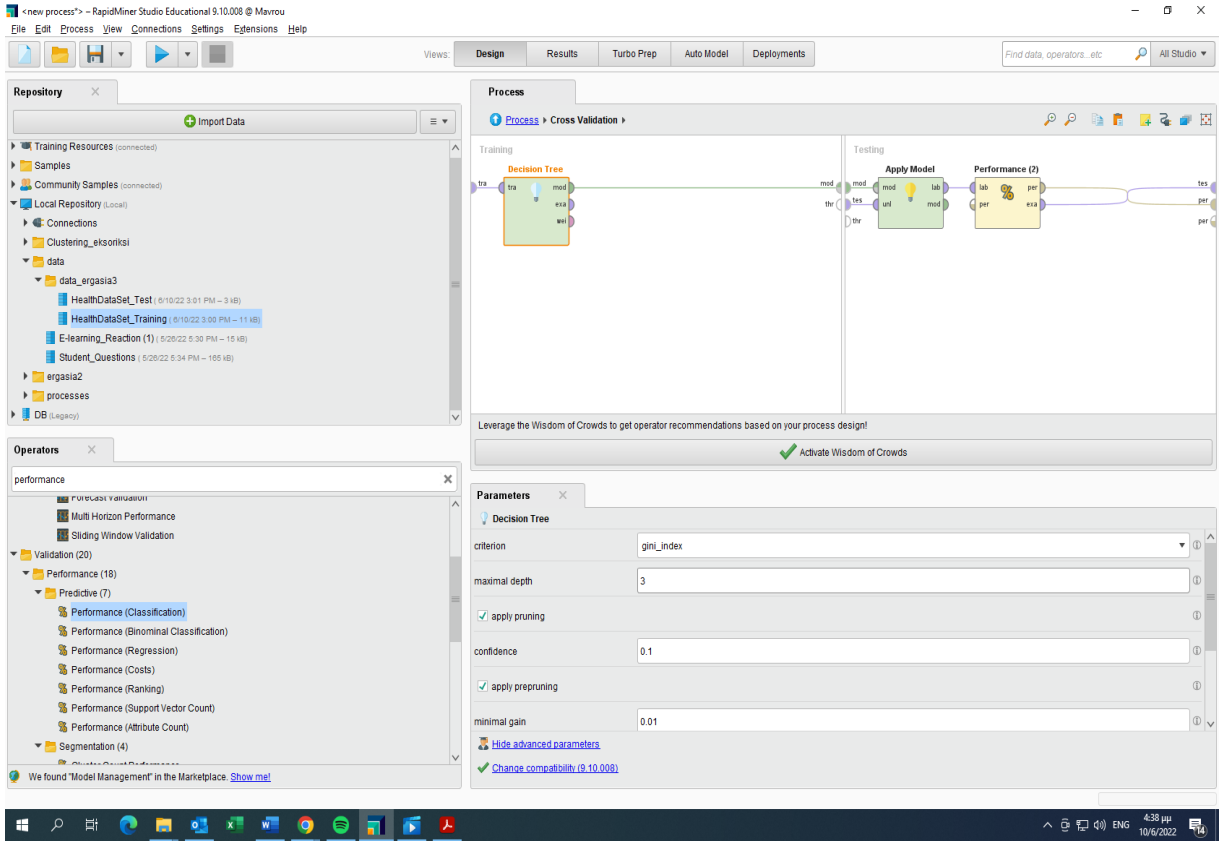
Παράδειγμα- *HEALTH DECISION TREE*

Στη συνέχεια αλλάζουμε τις παραμέτρους για να δούμε τις διαφορές και να επιλέξουμε Gini index (maximal depth 10)



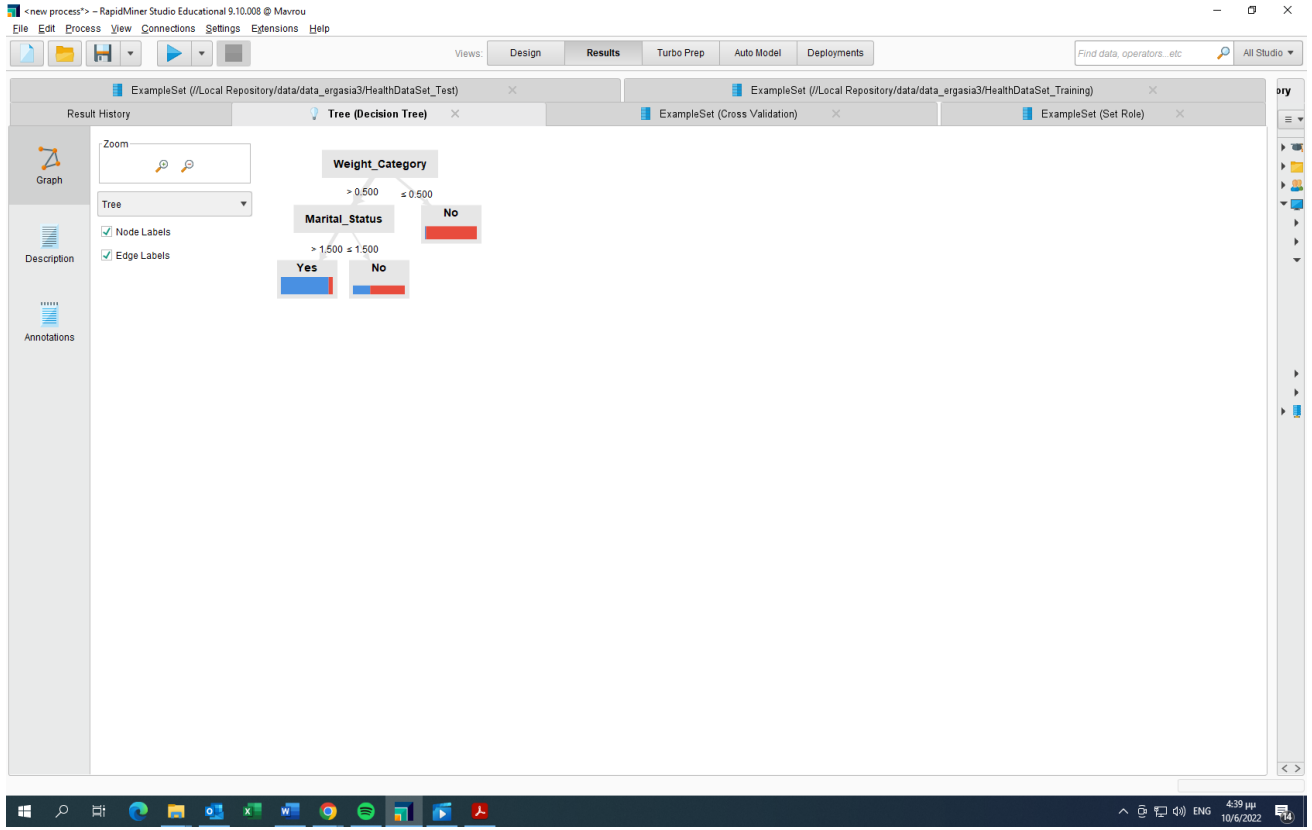
Παράδειγμα- *HEALTH DECISION TREE*

maximal depth 3



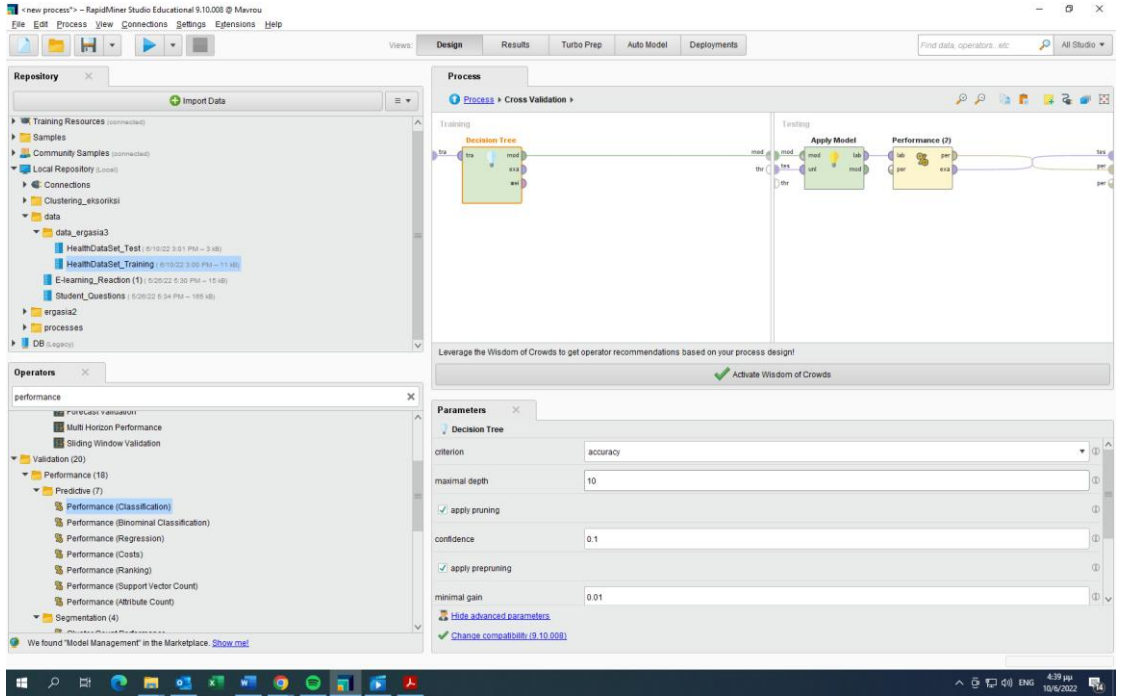
Παράδειγμα- *HEALTH DECISION TREE*

maximal depth 3



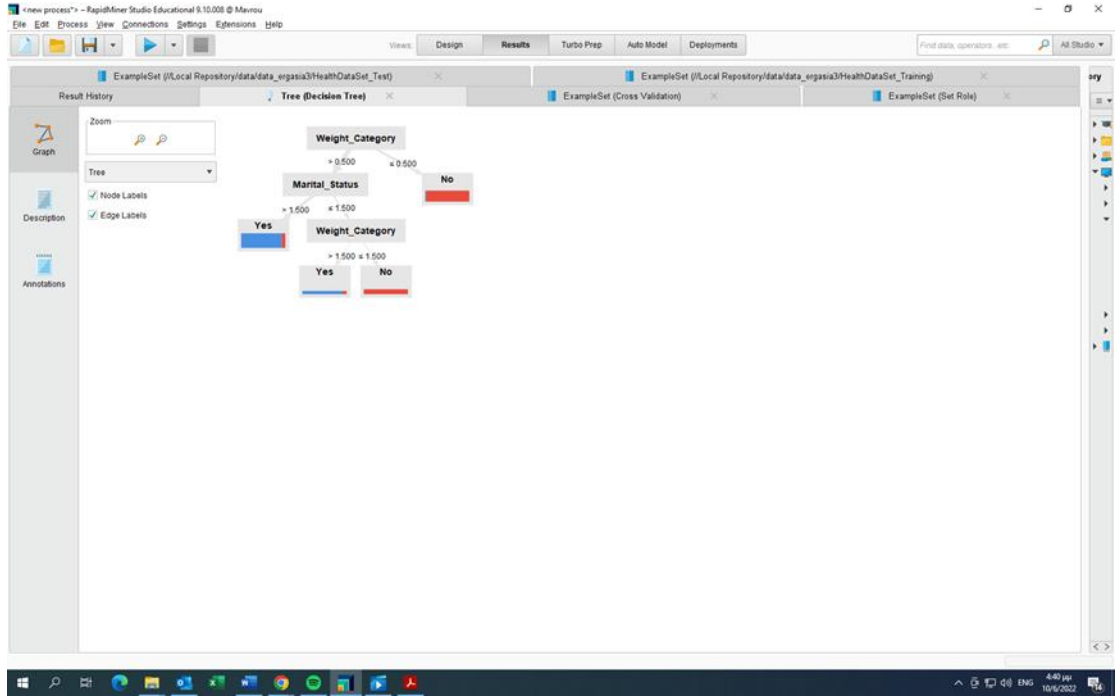
Παράδειγμα- *HEALTH DECISION TREE*

Accurancy (maximal depth 10)



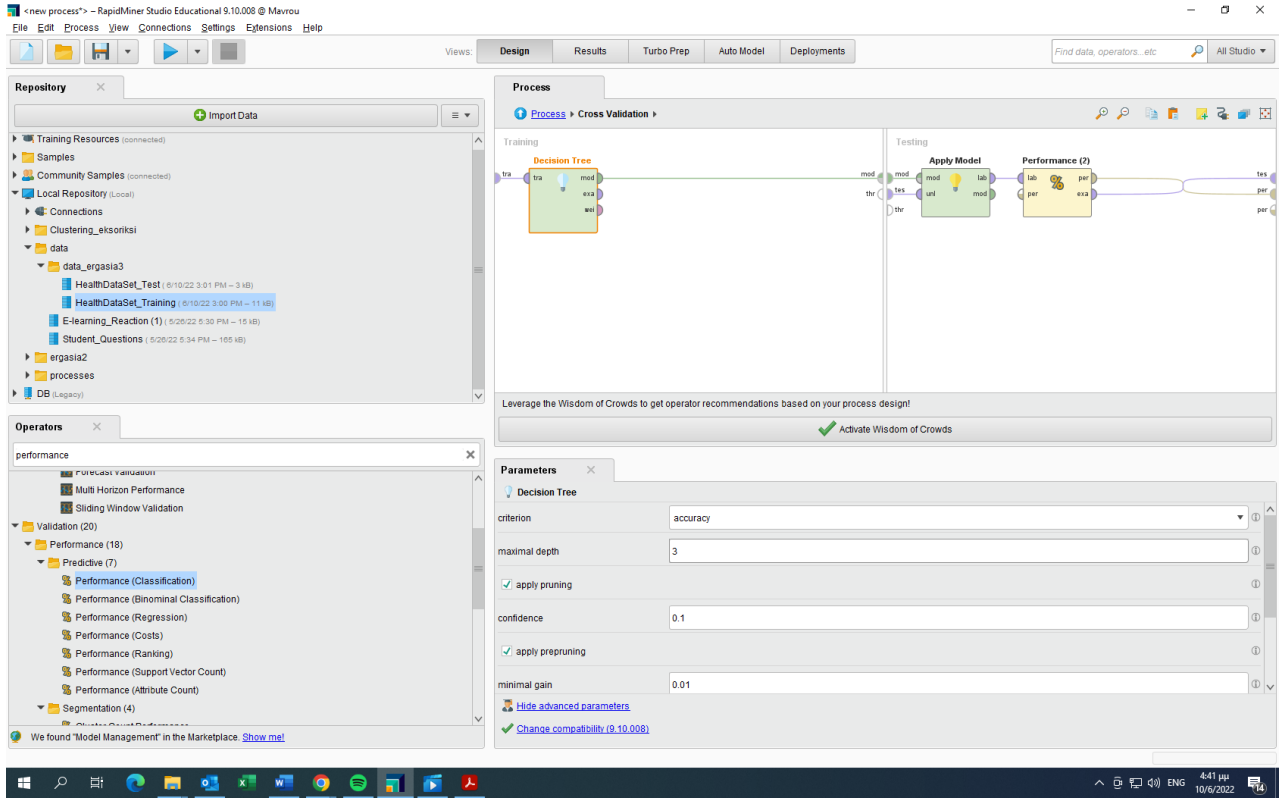
Παράδειγμα- *HEALTH DECISION TREE*

Accurancy (maximal depth 10)



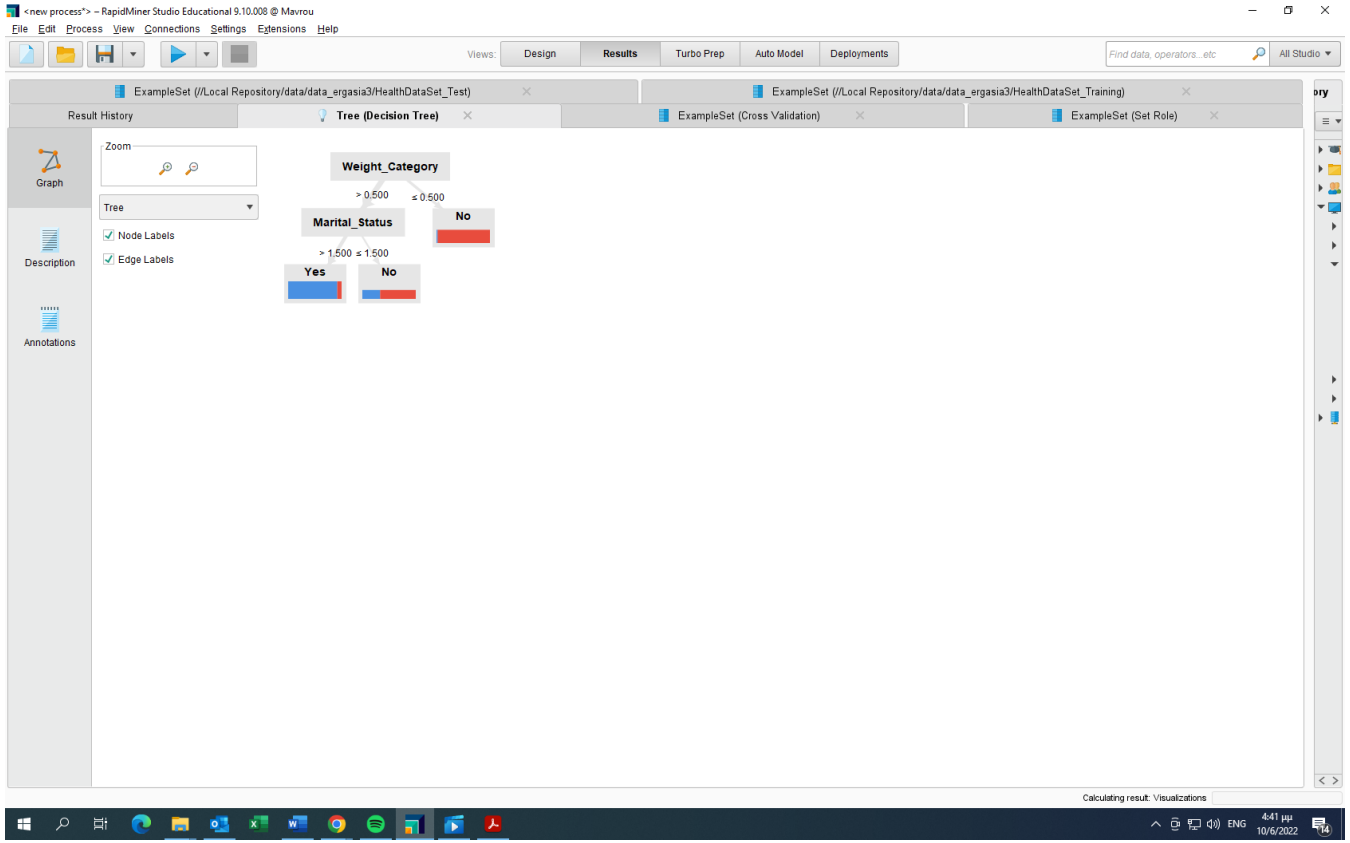
Παράδειγμα- *HEALTH DECISION TREE*

Accurancy (maximal depth 3)



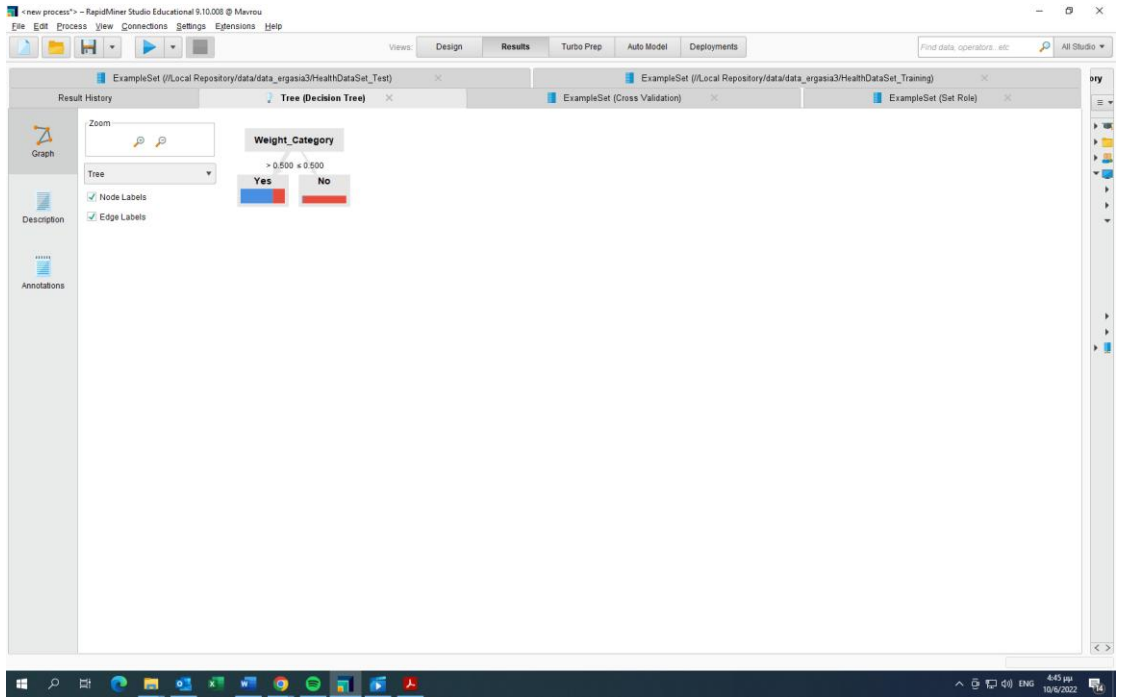
Παράδειγμα- *HEALTH DECISION TREE*

Accurancy (maximal depth 3)



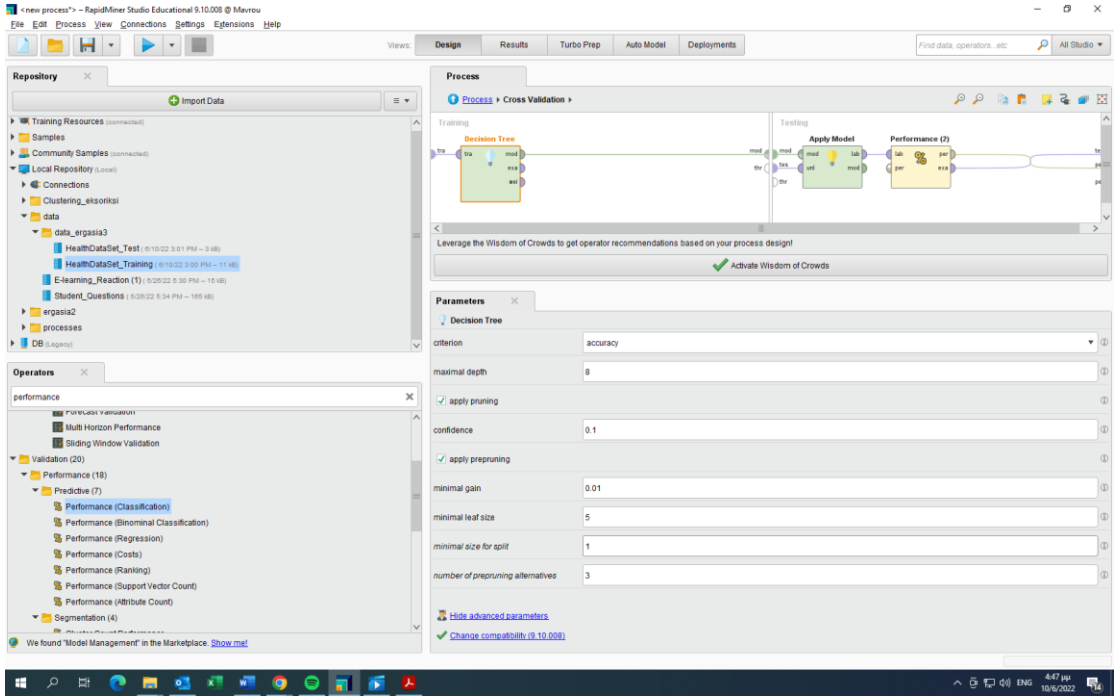
Παράδειγμα- *HEALTH DECISION TREE*

Accurancy (maximal depth 2)



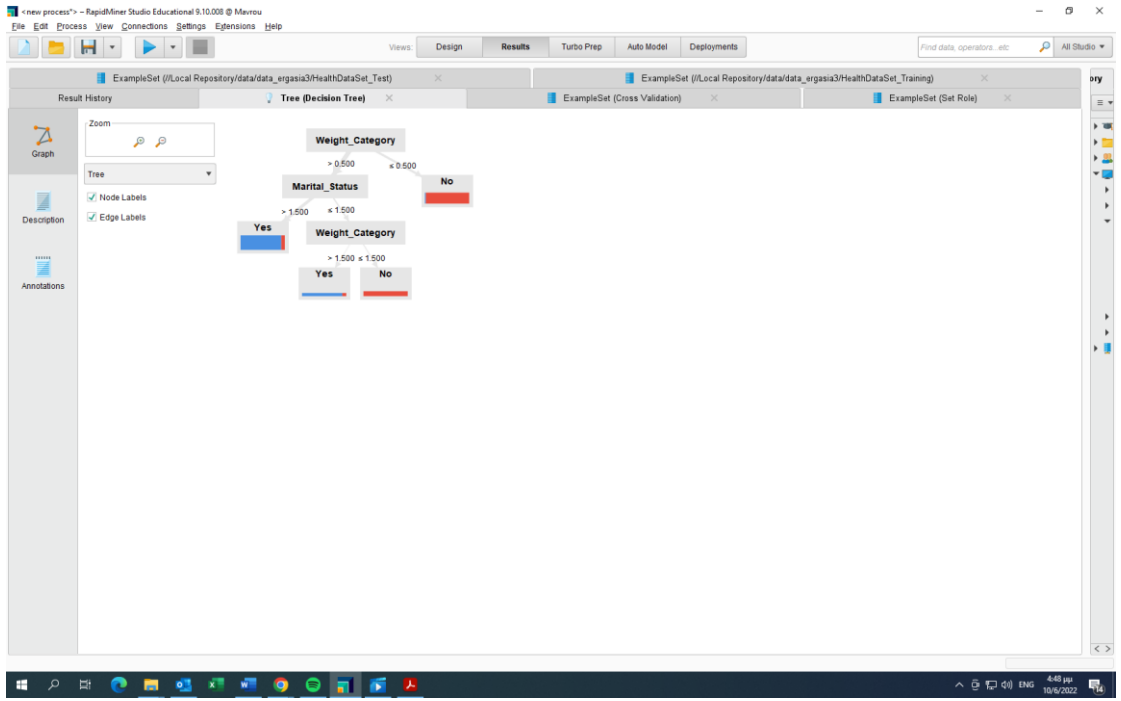
Παράδειγμα- *HEALTH DECISION TREE*

Accurancy 8 και αλλαγές στα minimal gain and leaf size



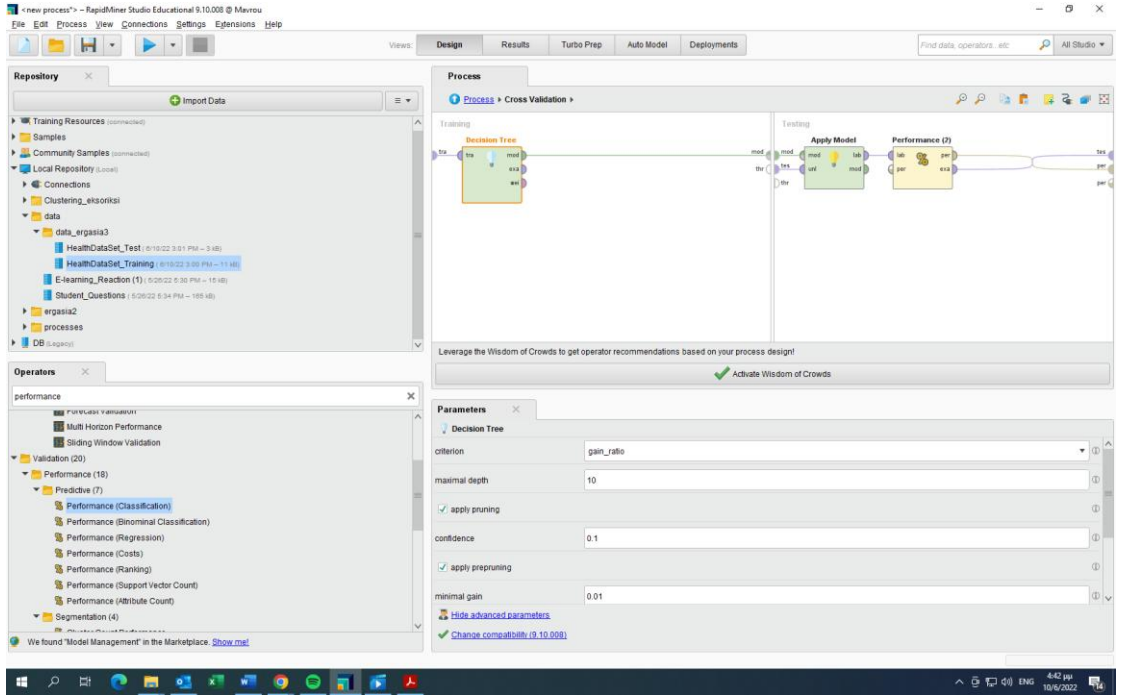
Παράδειγμα- *HEALTH DECISION TREE*

Accurancy 8 και αλλαγές στα minimal gain and leaf size



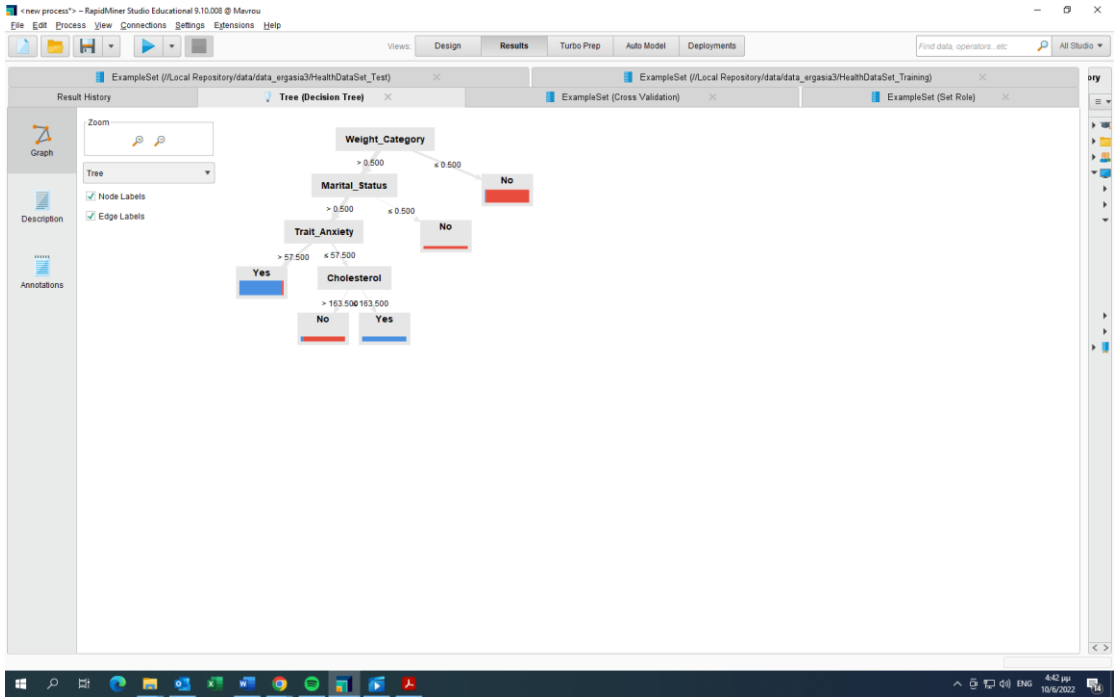
Παράδειγμα- *HEALTH DECISION TREE*

Gain ratio (maximal depth 10)



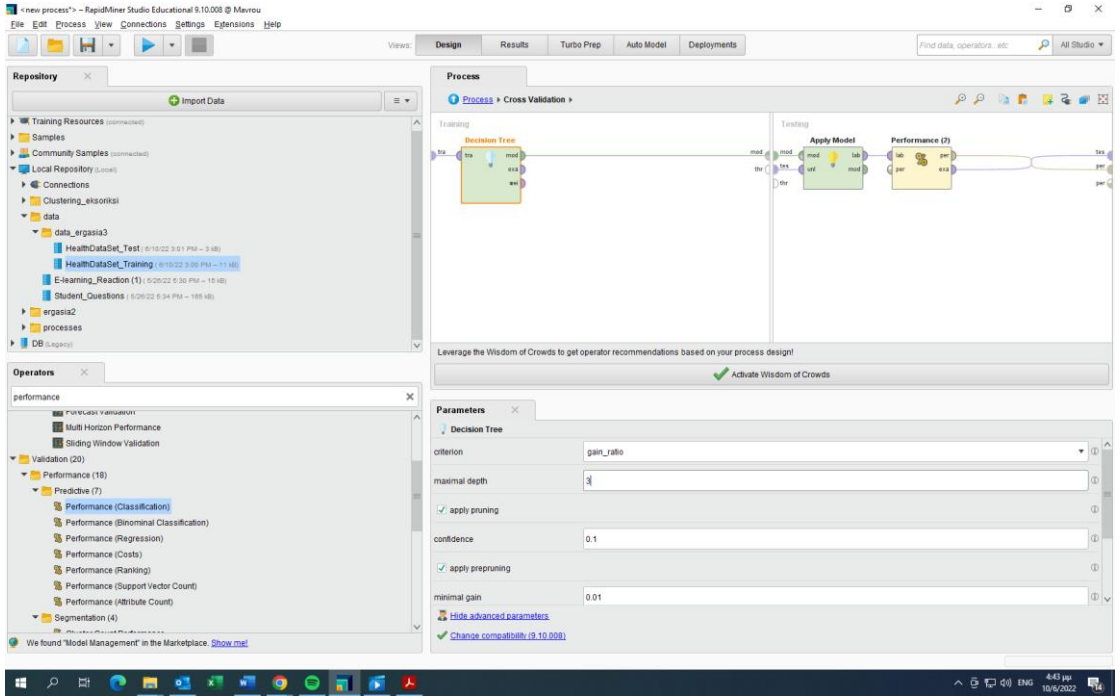
Παράδειγμα- *HEALTH DECISION TREE*

Gain ratio (maximal depth 10)



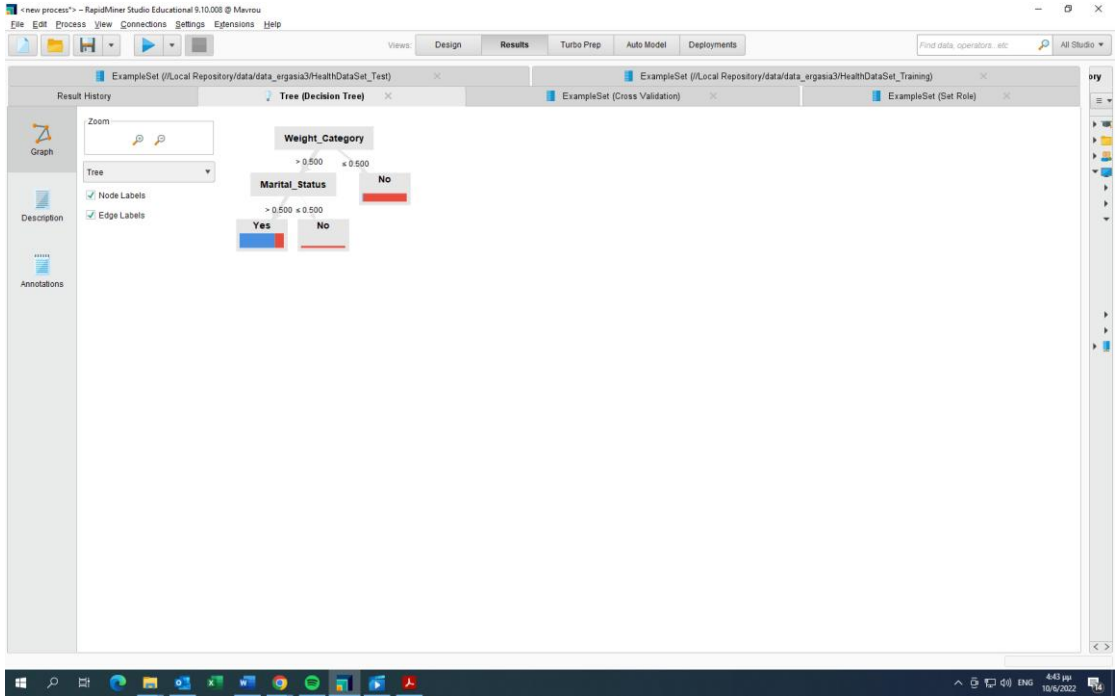
Παράδειγμα- *HEALTH DECISION TREE*

Gain ratio (maximal depth 3)



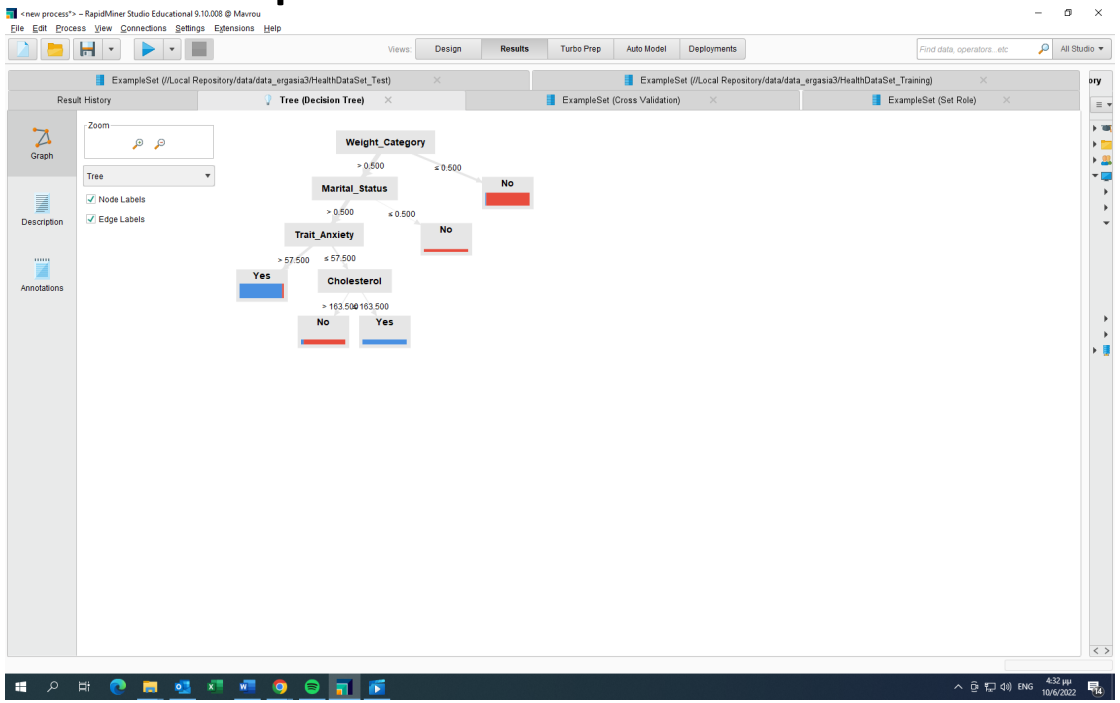
Παράδειγμα- *HEALTH DECISION TREE*

Gain ratio (maximal depth 3)



Παράδειγμα- *HEALTH DECISION TREE*

Για το τελικό δέντρο αποφάσεων επιλέγουμε την αρχική περίπτωση, όπου maximal depth 10

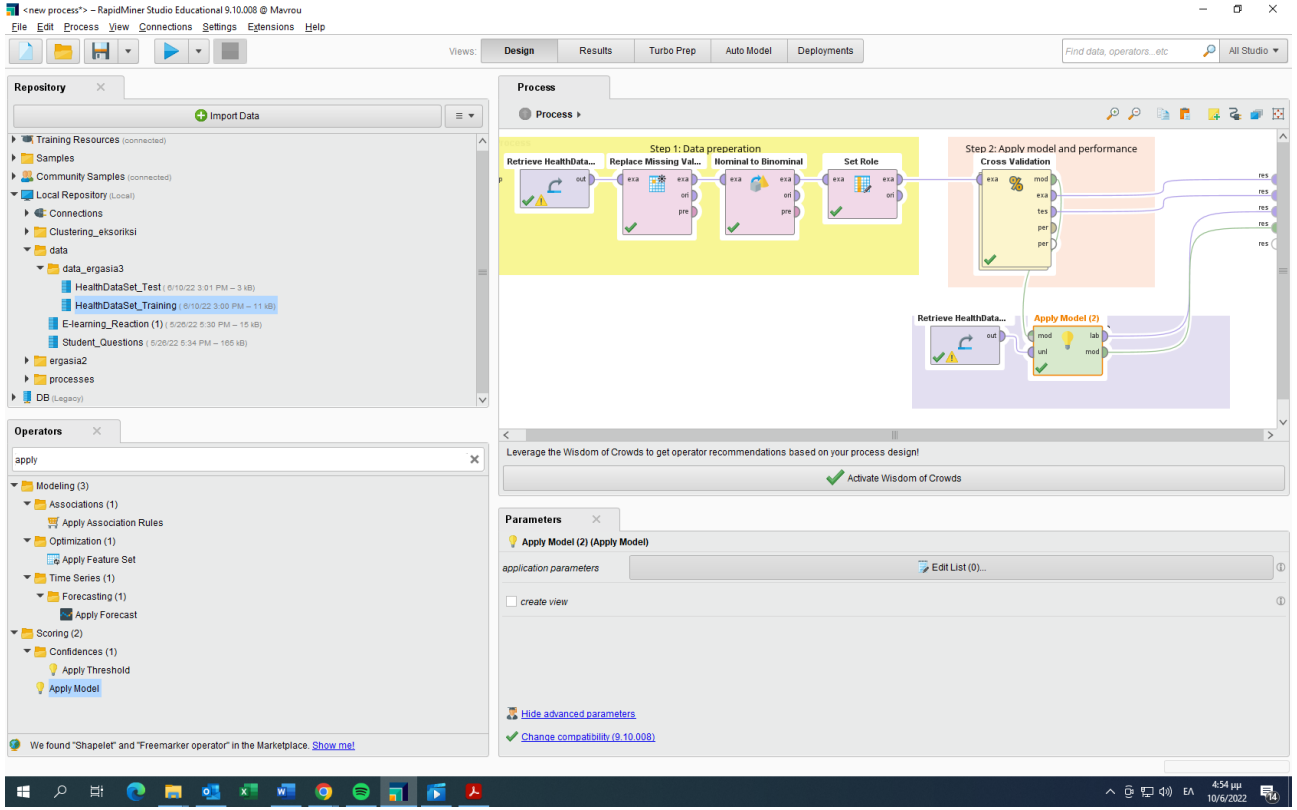


Παράδειγμα- *HEALTH DECISION TREE*

Φάση 3: Αξιολόγηση προβλέψεων (confidence)

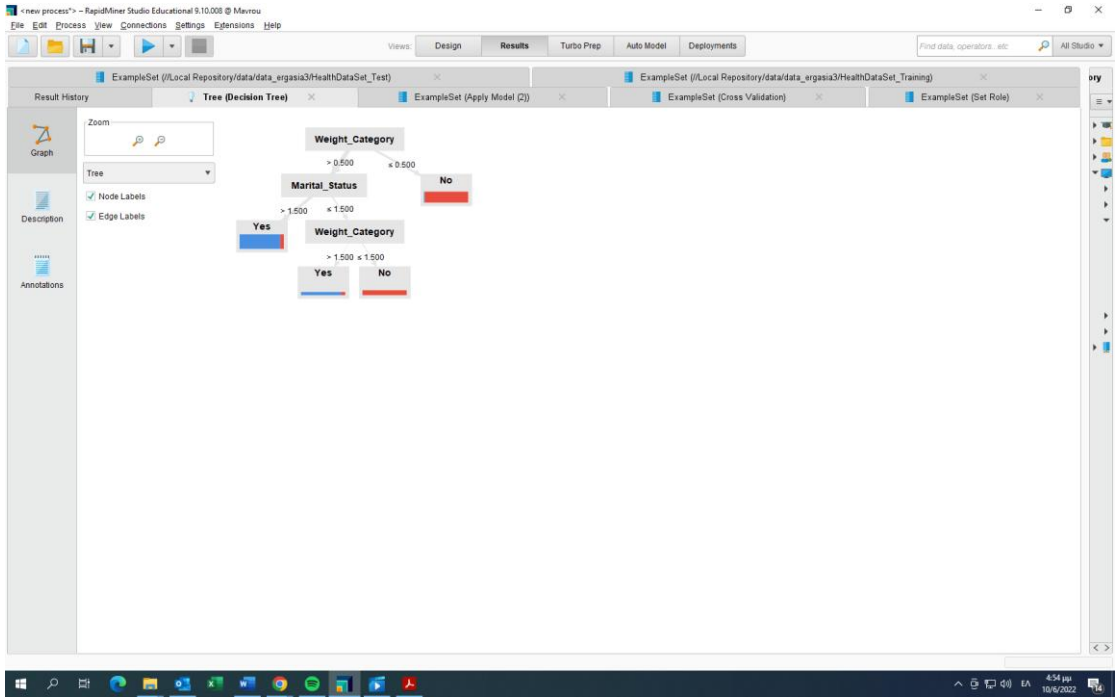
Παράδειγμα- *HEALTH DECISION TREE*

Χρησιμοποιήθηκε το αρχείο test για να δοκιμαστεί το Confidence του μοντέλου.



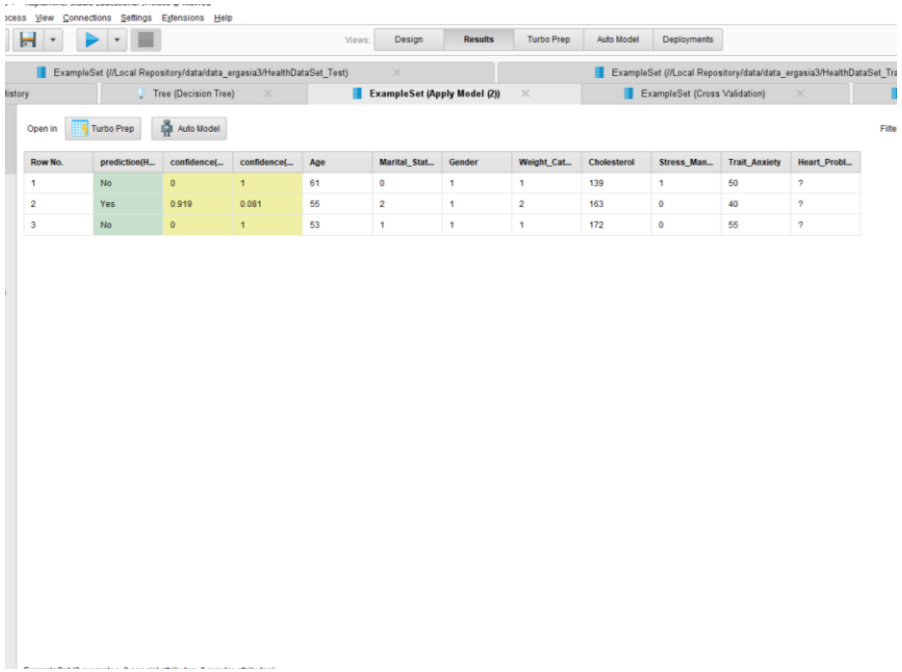
Παράδειγμα- *HEALTH DECISION TREE*

Χρησιμοποιήθηκε το αρχείο test για να δοκιμαστεί το Confidence του μοντέλου.



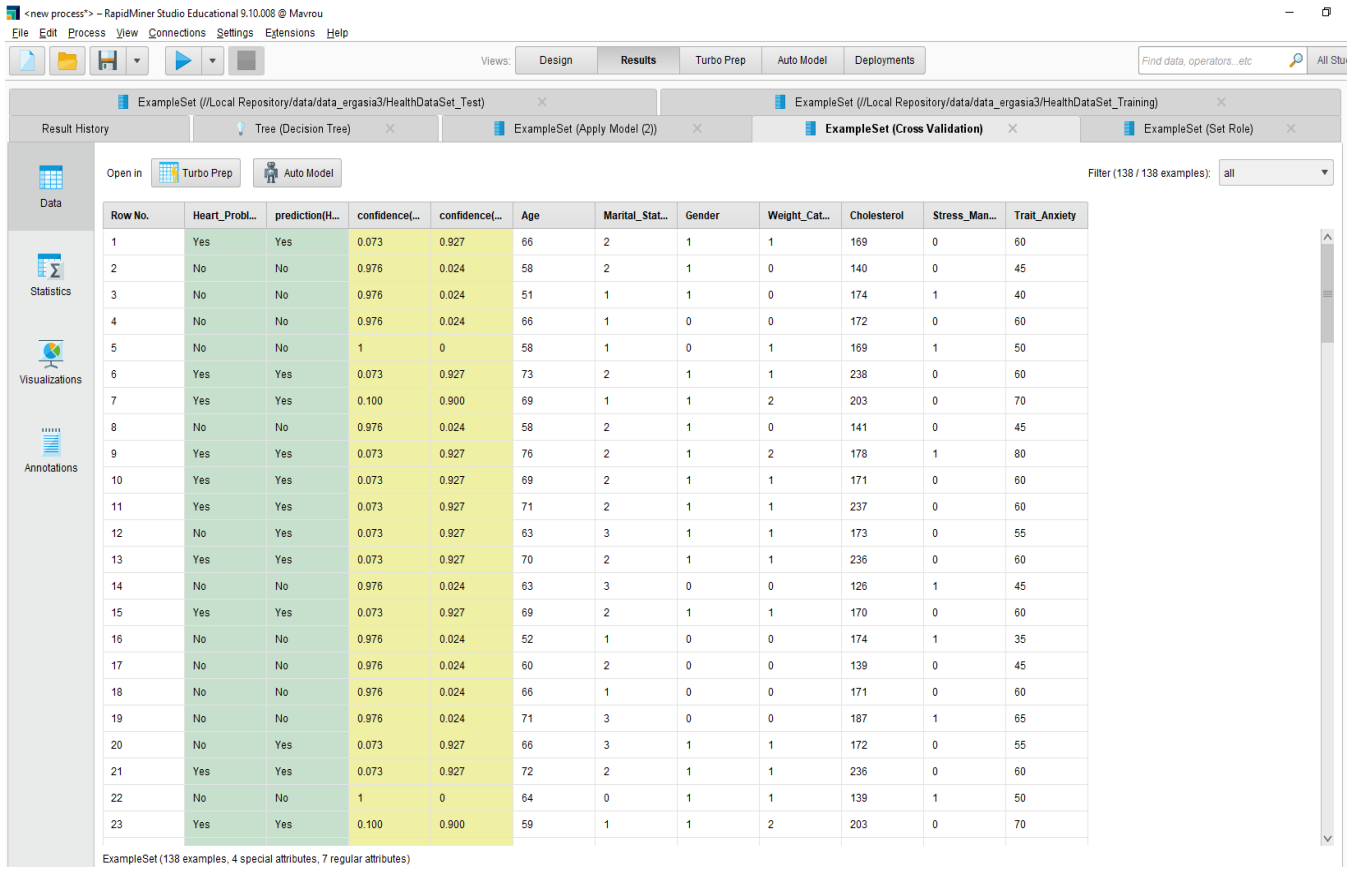
Παράδειγμα- *HEALTH DECISION TREE*

Στην παρακάτω εικόνα βλέπουμε τους δείκτες σχετικά με την εμφάνιση ή όχι καρδιολογικών προβλημάτων.
Στην πρώτη σειρά φαίνεται ότι ο δείκτης confidence είναι 1, που ερμηνεύεται ότι οι περιπτώσεις που ανήκουν σε αυτή την κατηγορία δεν θα εμφανίσουν (no στο prediction) με ποσοστό 100%.
Ομοίως στη δεύτερη σειρά, το confidence είναι 0.919, που αντιστοιχεί σε ποσοστό 92% ότι είναι σωστή η πρόβλεψη



Row No.	prediction(H...	confidence(...	confidence(...	Age	Marital_Stat...	Gender	Weight_Cat...	Cholesterol	Stress_Man...	Trait_Anxiety	Heart_Probl...
1	No	0	1	61	0	1	1	139	1	50	?
2	Yes	0.919	0.081	55	2	1	2	163	0	40	?
3	No	0	1	53	1	1	1	172	0	55	?

Παράδειγμα- *HEALTH DECISION TREE*



Row No.	Heart_Probl...	prediction(H...	confidence(...	confidence(...	Age	Marital_Stat...	Gender	Weight_Cat...	Cholesterol	Stress_Man...	Trait_Anxiety
1	Yes	Yes	0.073	0.927	66	2	1	1	169	0	60
2	No	No	0.976	0.024	58	2	1	0	140	0	45
3	No	No	0.976	0.024	51	1	1	0	174	1	40
4	No	No	0.976	0.024	66	1	0	0	172	0	60
5	No	No	1	0	58	1	0	1	169	1	50
6	Yes	Yes	0.073	0.927	73	2	1	1	238	0	60
7	Yes	Yes	0.100	0.900	69	1	1	2	203	0	70
8	No	No	0.976	0.024	58	2	1	0	141	0	45
9	Yes	Yes	0.073	0.927	76	2	1	2	178	1	80
10	Yes	Yes	0.073	0.927	69	2	1	1	171	0	60
11	Yes	Yes	0.073	0.927	71	2	1	1	237	0	60
12	No	Yes	0.073	0.927	63	3	1	1	173	0	55
13	Yes	Yes	0.073	0.927	70	2	1	1	236	0	60
14	No	No	0.976	0.024	63	3	0	0	126	1	45
15	Yes	Yes	0.073	0.927	69	2	1	1	170	0	60
16	No	No	0.976	0.024	52	1	0	0	174	1	35
17	No	No	0.976	0.024	60	2	0	0	139	0	45
18	No	No	0.976	0.024	66	1	0	0	171	0	60
19	No	No	0.976	0.024	71	3	0	0	187	1	65
20	No	Yes	0.073	0.927	66	3	1	1	172	0	55
21	Yes	Yes	0.073	0.927	72	2	1	1	236	0	60
22	No	No	1	0	64	0	1	1	139	1	50
23	Yes	Yes	0.100	0.900	59	1	1	2	203	0	70

ΤΕΛΟΣ ΕΝΟΤΗΤΑΣ