Abdhel Muthazz Mashood
Student ID : 21595621

Business Intelligence Technologies: Practical 1: Classification using Decision Trees

Weighting: 4% Due date: 10 Dec 2022

Using the heart.csv data provided, construct an initial 'default' decision tree for classifying whether a patient has less chance, or more chance of having a heart attack, and summarise the performance of the model.

Improve the performance by changing the decision tree parameters. Suggest reasons for why the performance has improved, and explain the results.

Document your processes, showing step by step screens and outputs, including an explanation of the final parameters you used, and what effect they had on your model.

Assessment Criteria:

Correct and fully documented processes: 2 marks
Explanation demonstrating sound understanding of principles of classification using Decision Trees : 2 marks

Abdhel Muthazz Mashood
Student ID : 21595621

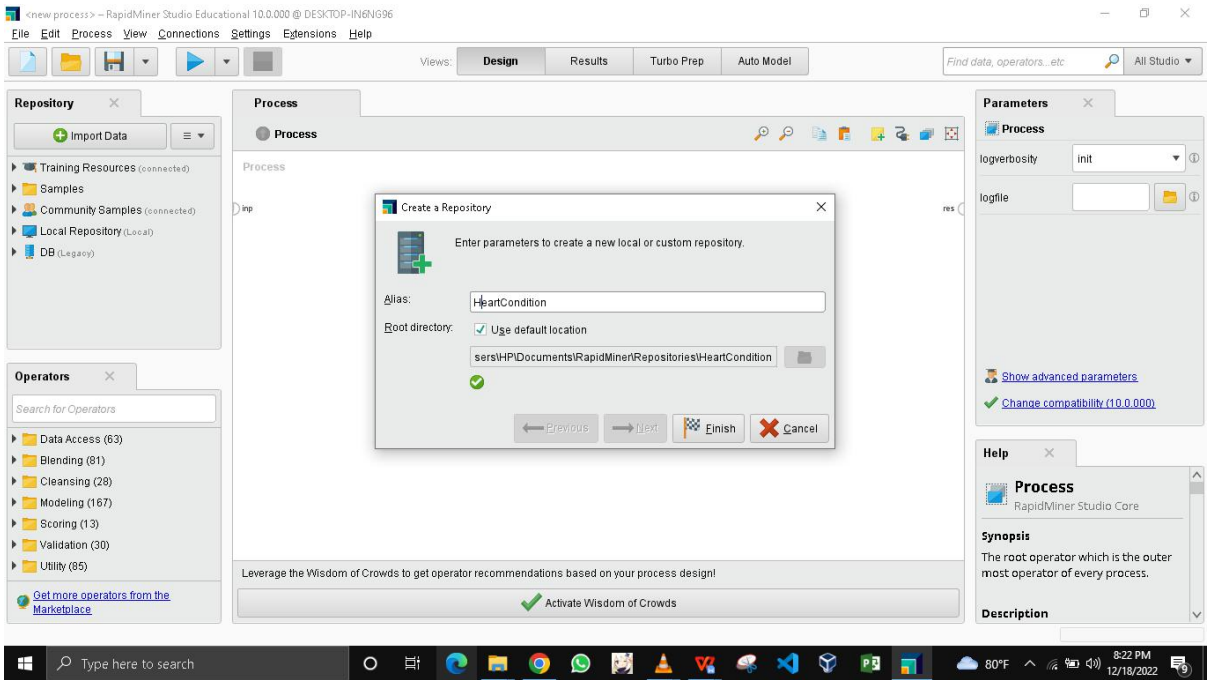
Description of the Heart dataset:

- Age : Age of the patient
- Sex : Sex of the patient
- exng: exercise induced angina (1 = yes; 0 = no)
- caa: number of major vessels coloured by flourosopy (0-3)
- cp : Chest Pain type chest pain type
    - Value 1: typical angina
    - Value 2: atypical angina
    - Value 3: non-anginal pain
    - Value 4: asymptomatic
- trtbps : resting blood pressure (in mm Hg)
- chol : cholestoral in mg/dl fetched via BMI sensor
- fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
- restecg : resting electrocardiographic results
    - Value 0: normal
    - Value 1: having ST-T wave abnormality
    - Value 2: showing probable or definite left ventricular hypertrophy
- thalachh : maximum heart rate achieved
- oldpeak: ST depression induced by exercise relative to rest (-2.6 to 6.2)
- slp: the slope of the peak execise ST ssegment (0=upsloping, 1=flat, 2=Down)
- thall: 0=normal; 1=fixed defect; 2=reversable defect
- output (ie the label) : 0= less chance of heart attack 1= more chance of heart attack

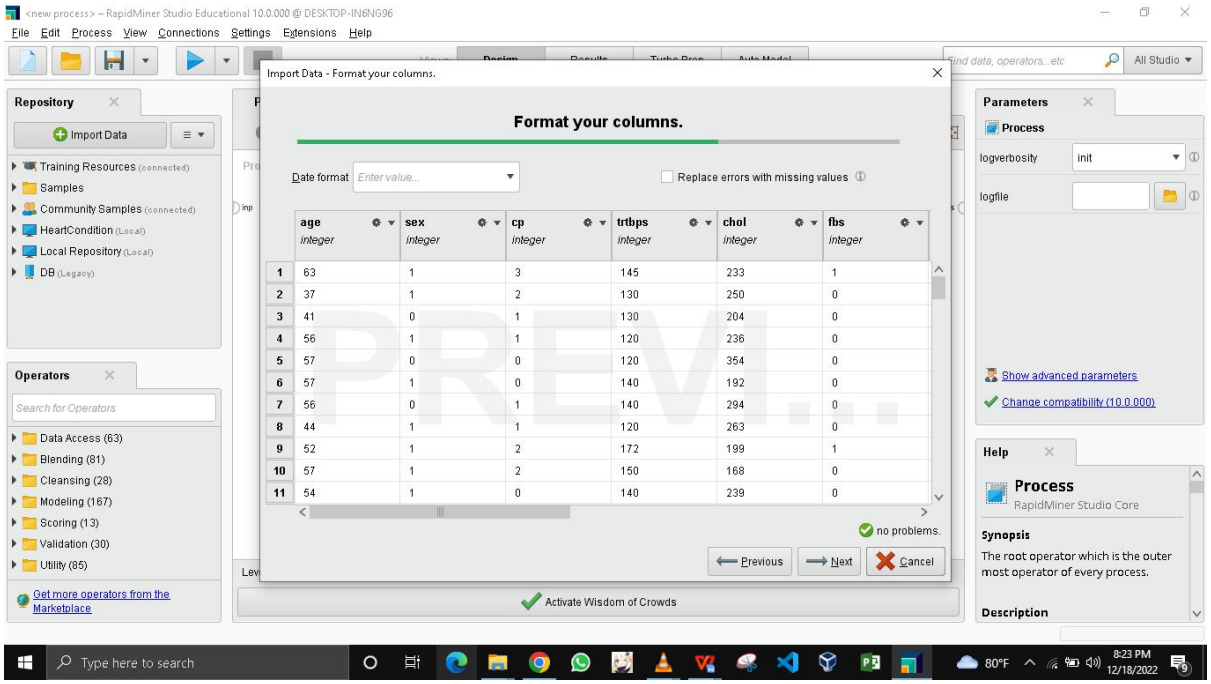
Abdhel Muthazz Mashood
Student ID : 21595621

Practical Documentation

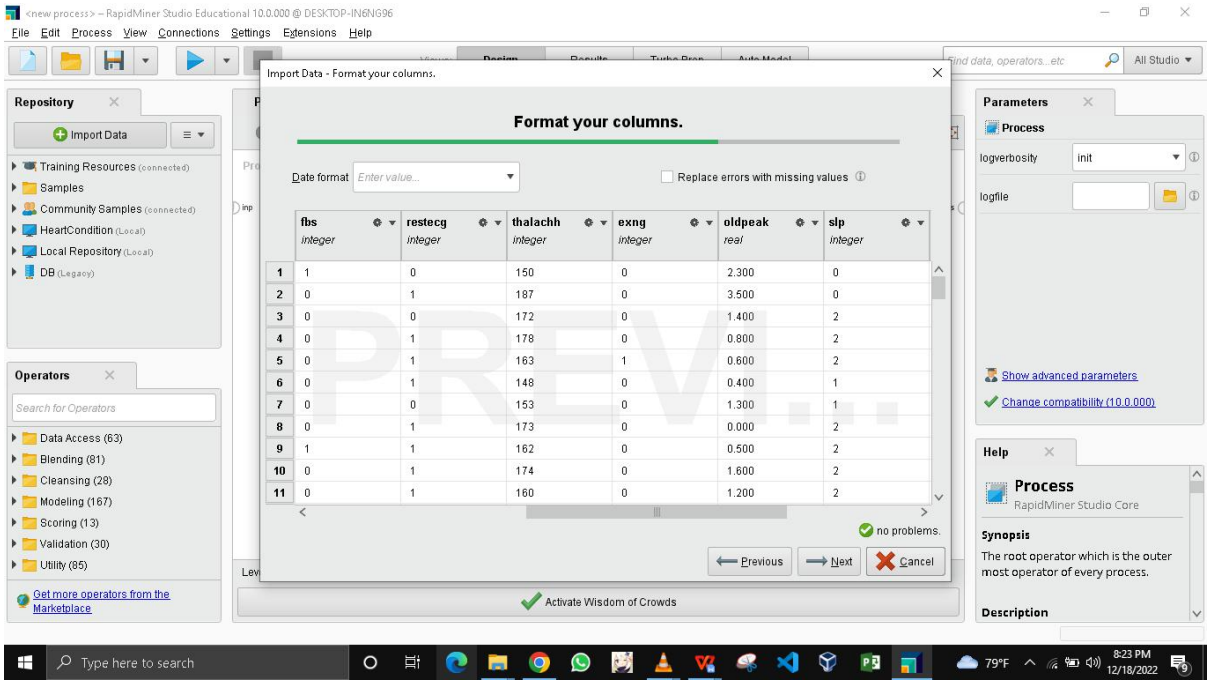
1. Created a Repository labeled “HeartCondition”

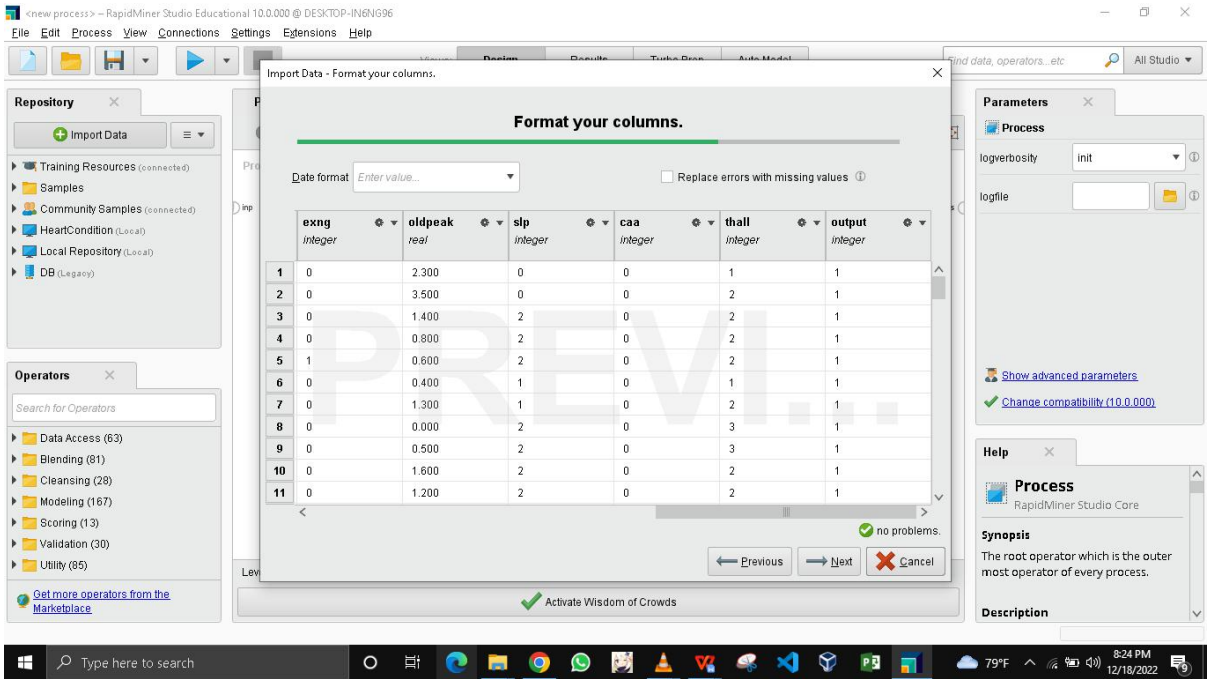


2. Imported the heartData table into the HeartCondition Repository with the table’s default attributes unchanged as shown below



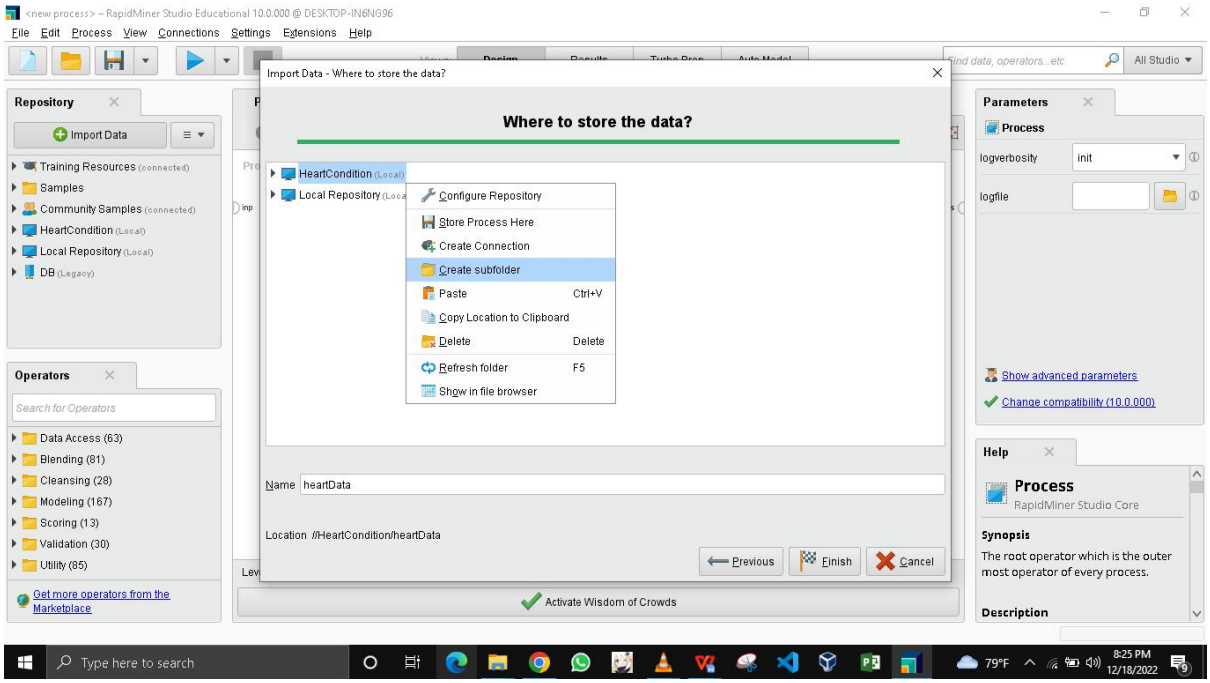
Abdhel Muthazz Mashood
Student ID : 21595621

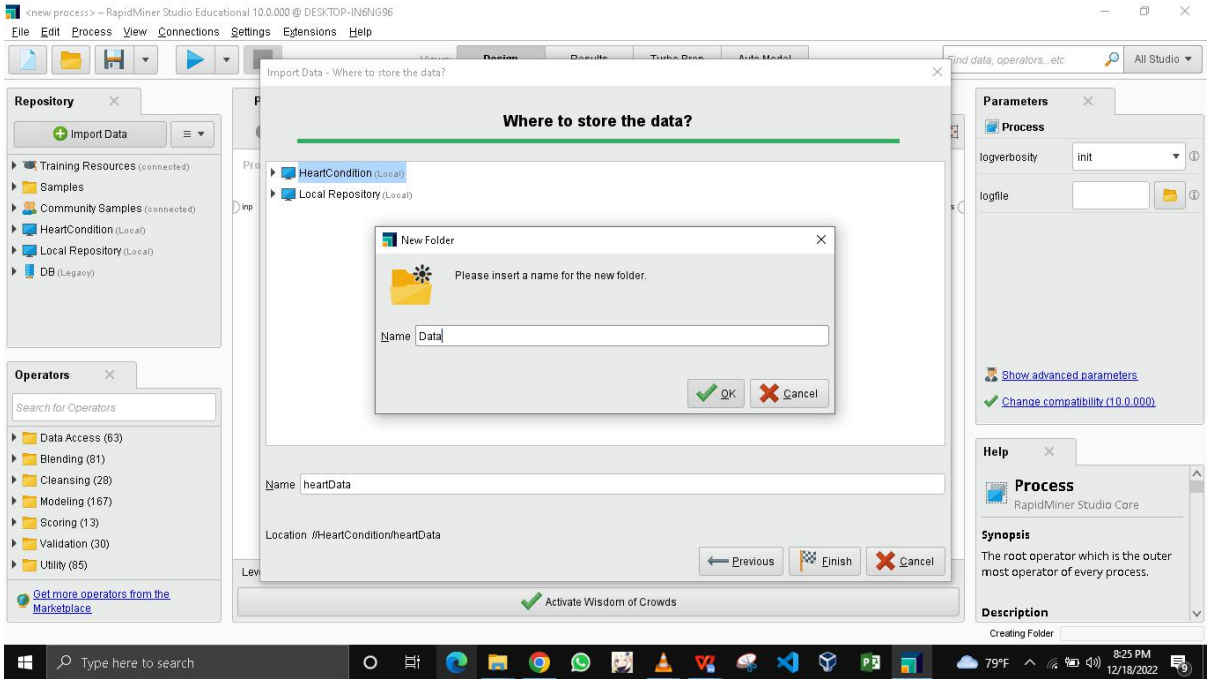




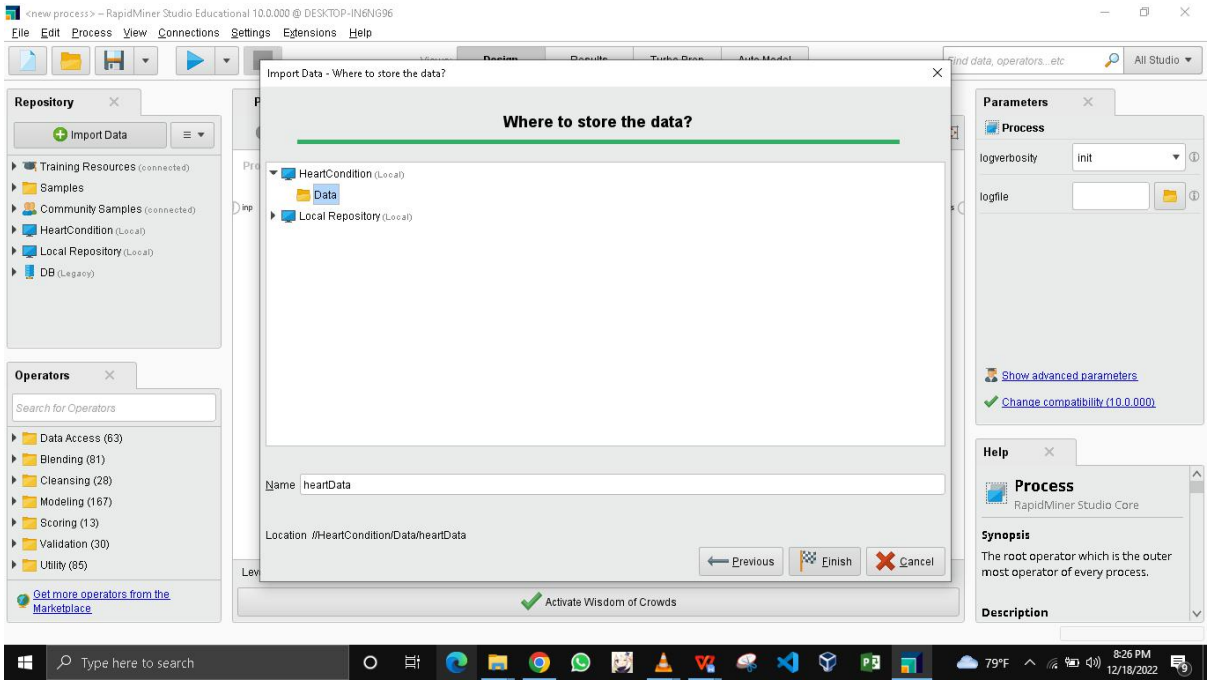
Abdhel Muthazz Mashood
Student ID : 21595621

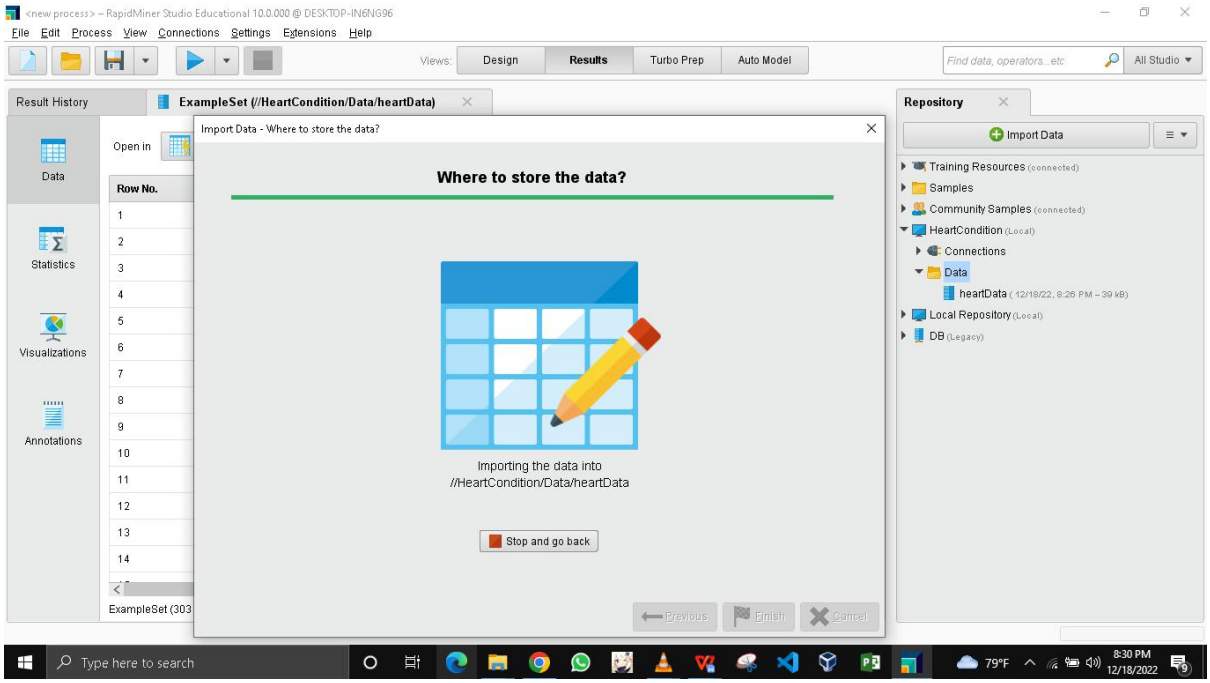
3. Created a “Data” Folder within the Heart Condition Repository in which the heartData table was imported into





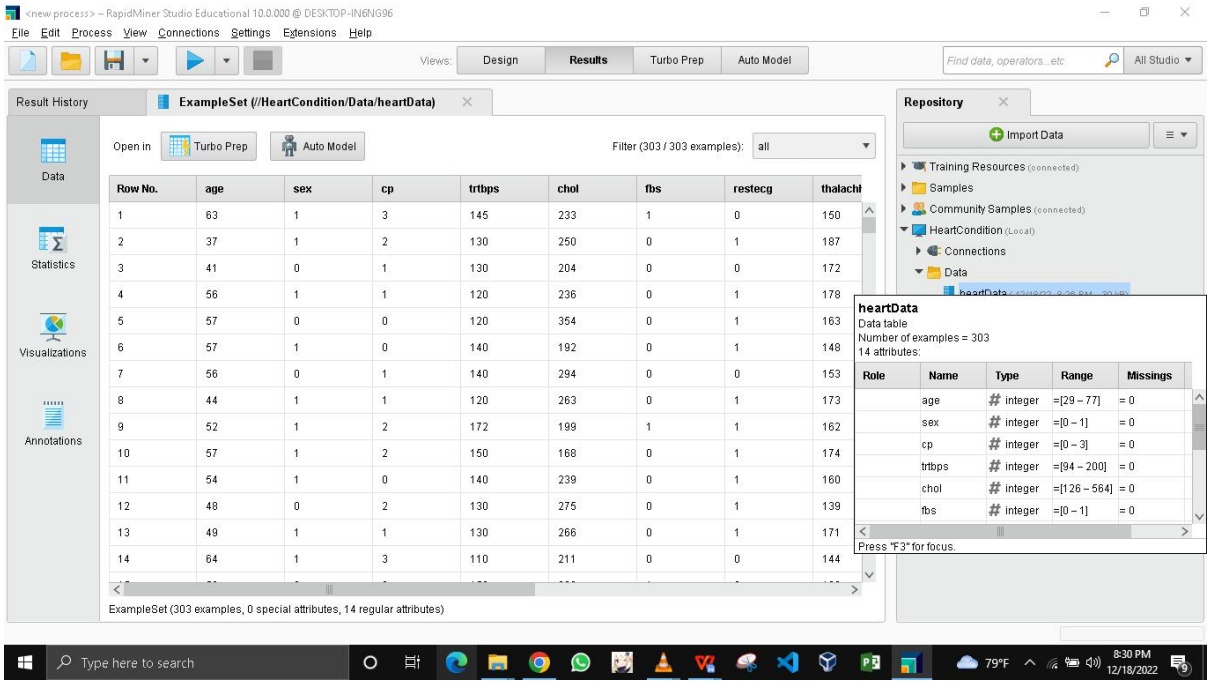
Abdhel Muthazz Mashood
Student ID : 21595621

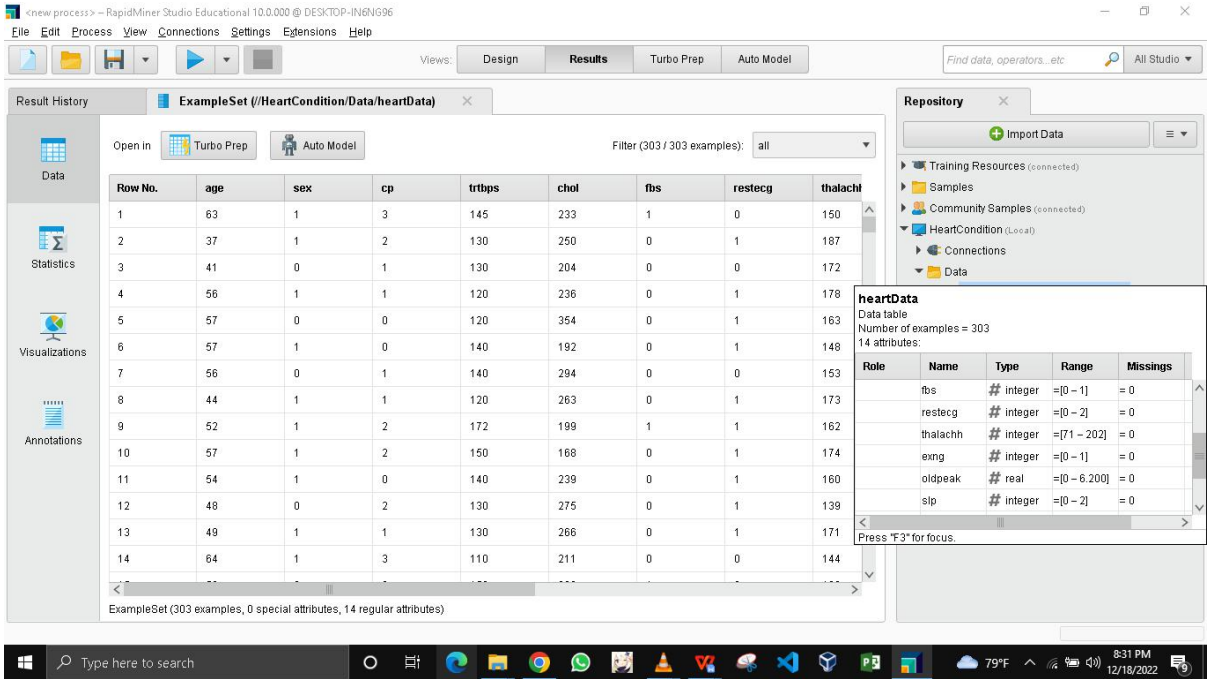




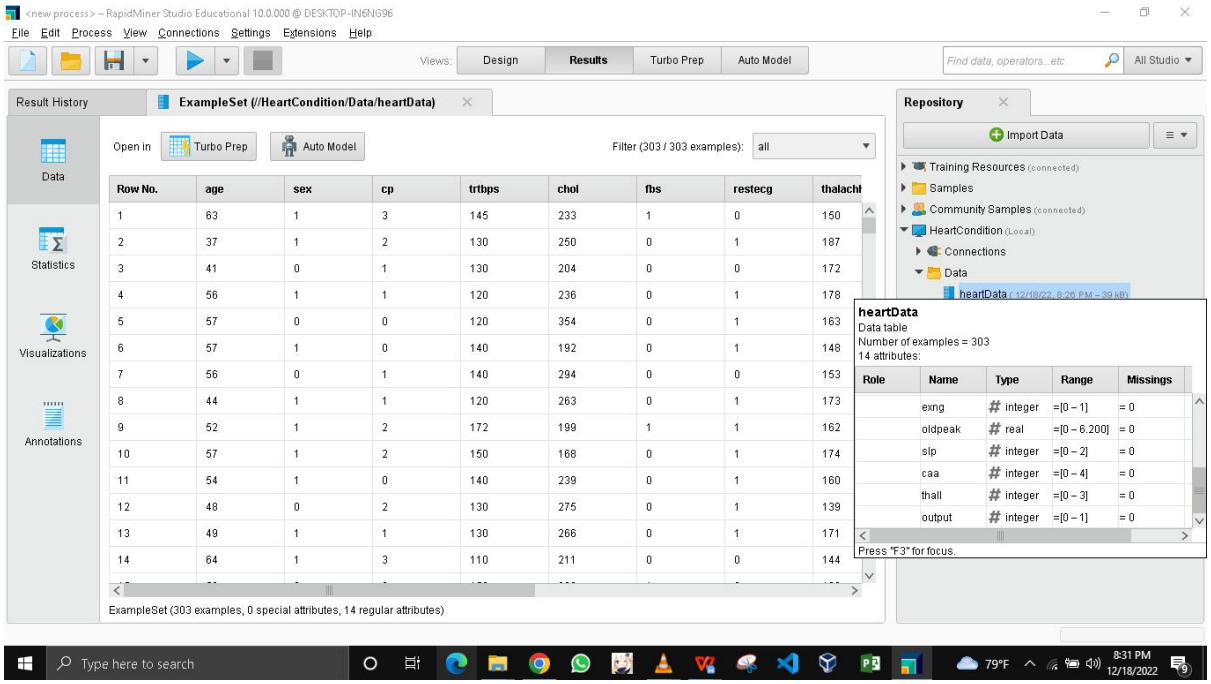
Abdhel Muthazz Mashood
Student ID : 21595621

Below are the results of the completed import with the ranges of the values within the heartData table shown for further detail

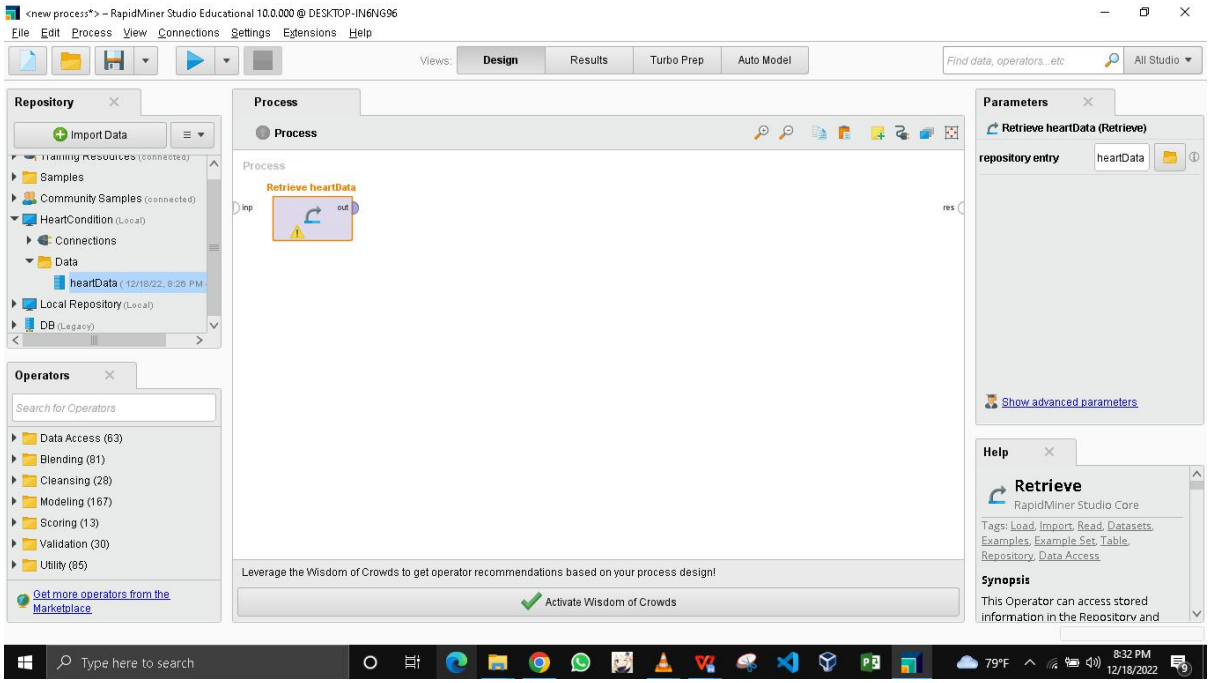




Abdhel Muthazz Mashood
Student ID : 21595621

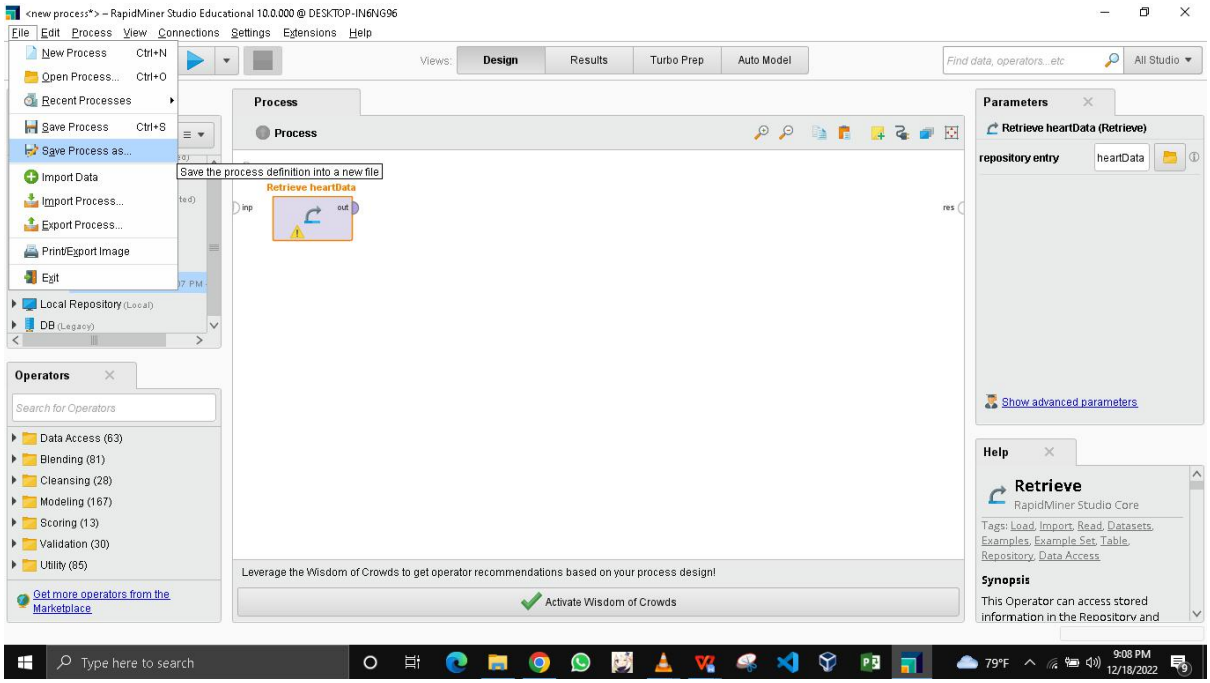


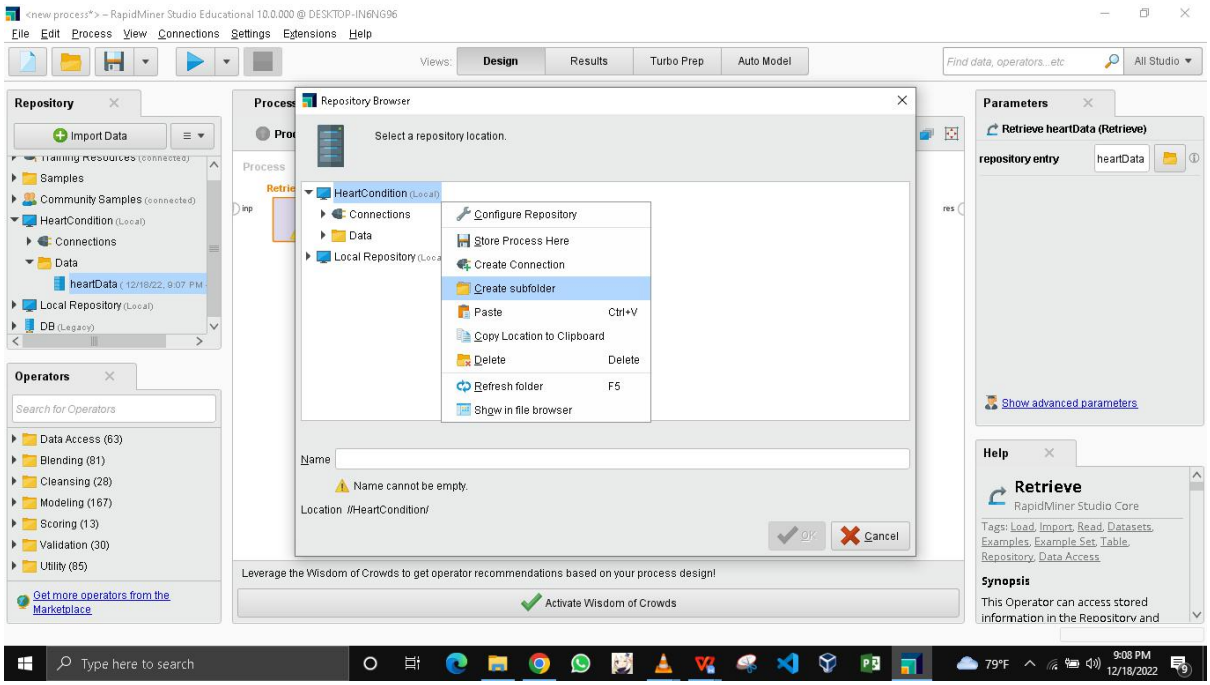
4. Retrieved the heartData table into the Process panel



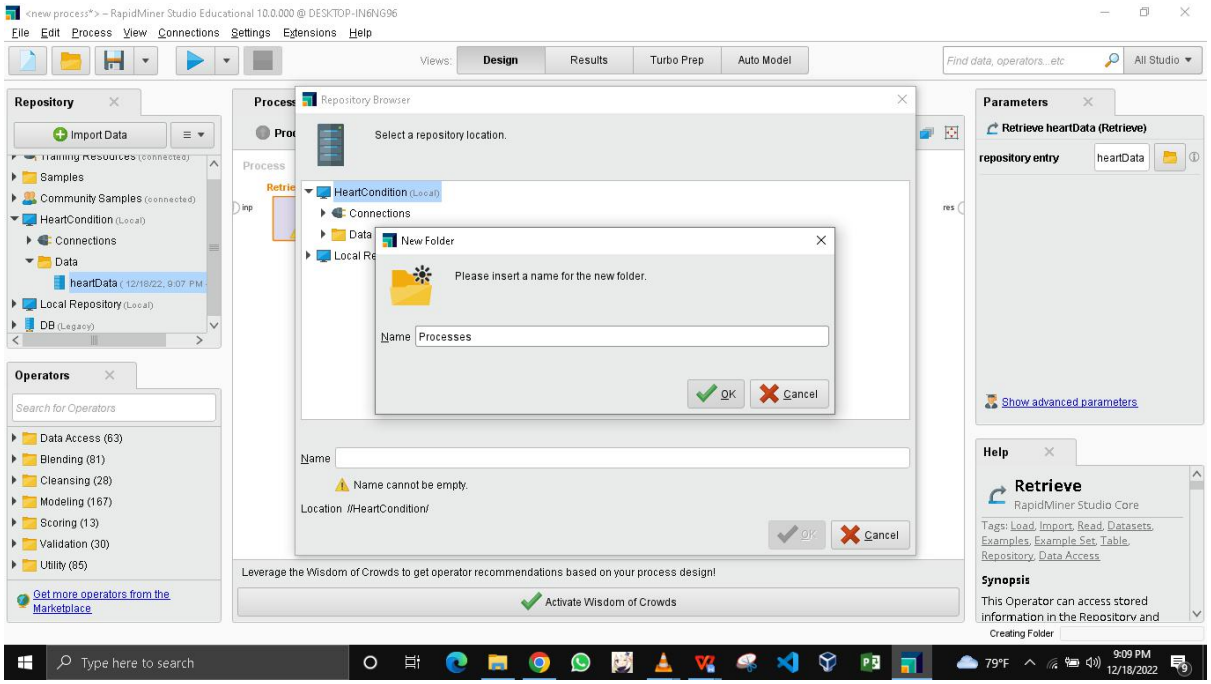
Abdhel Muthazz Mashood
Student ID : 21595621

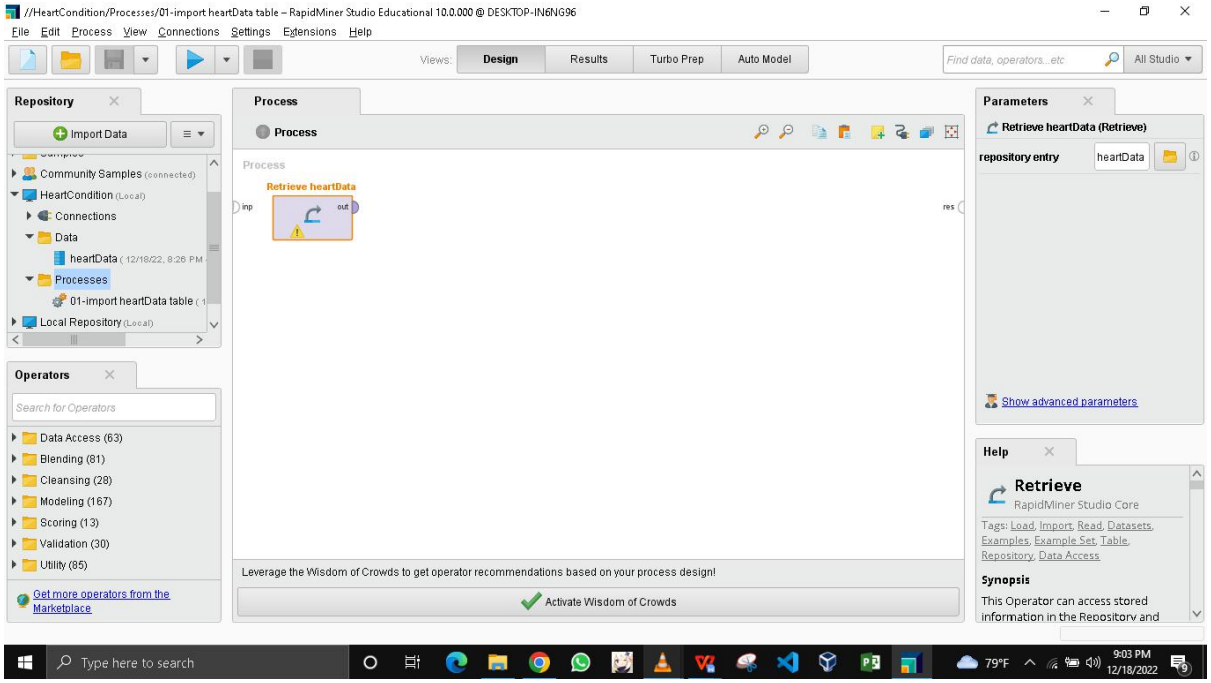
5. Created a “Processes” Folder and saved the process as “01-import heartData table” in the “Processes” Folder





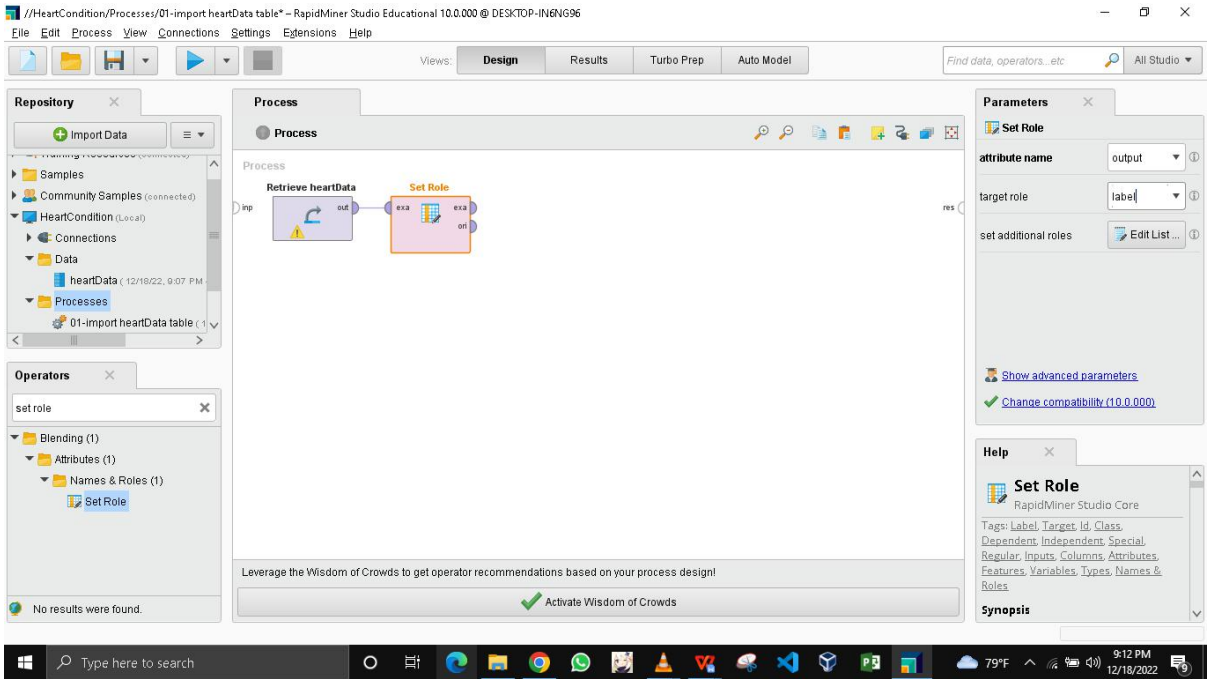
Abdhel Muthazz Mashood
Student ID : 21595621

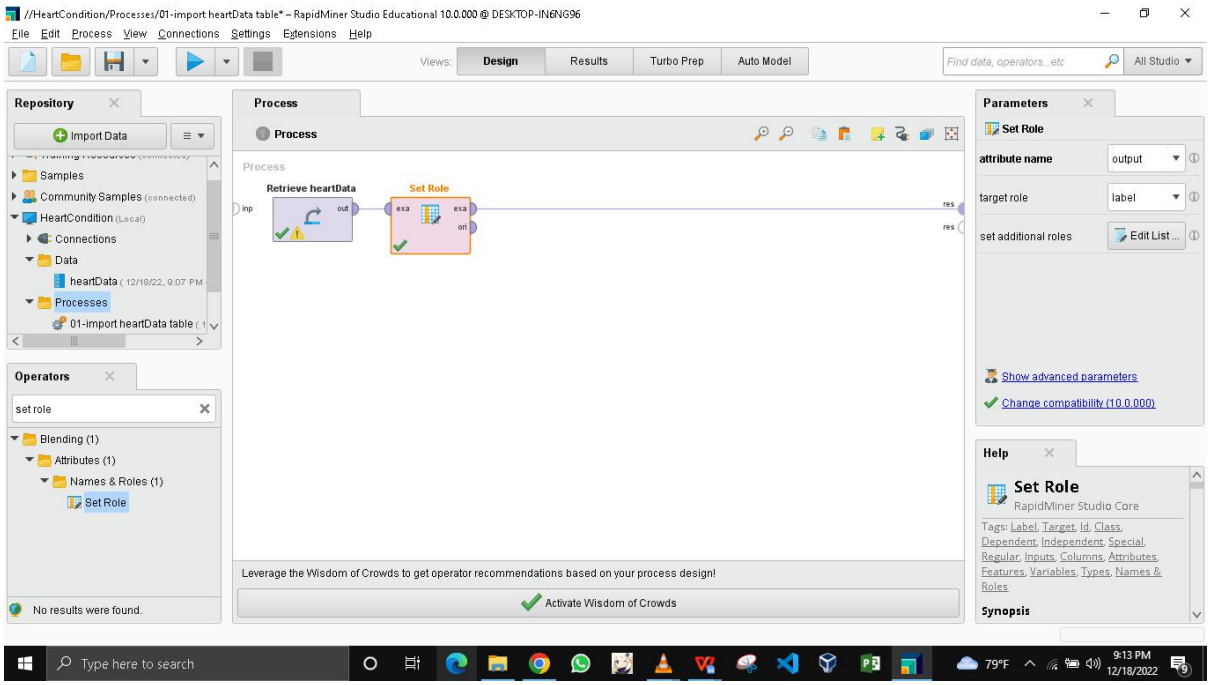




Abdhel Muthazz Mashood
Student ID : 21595621

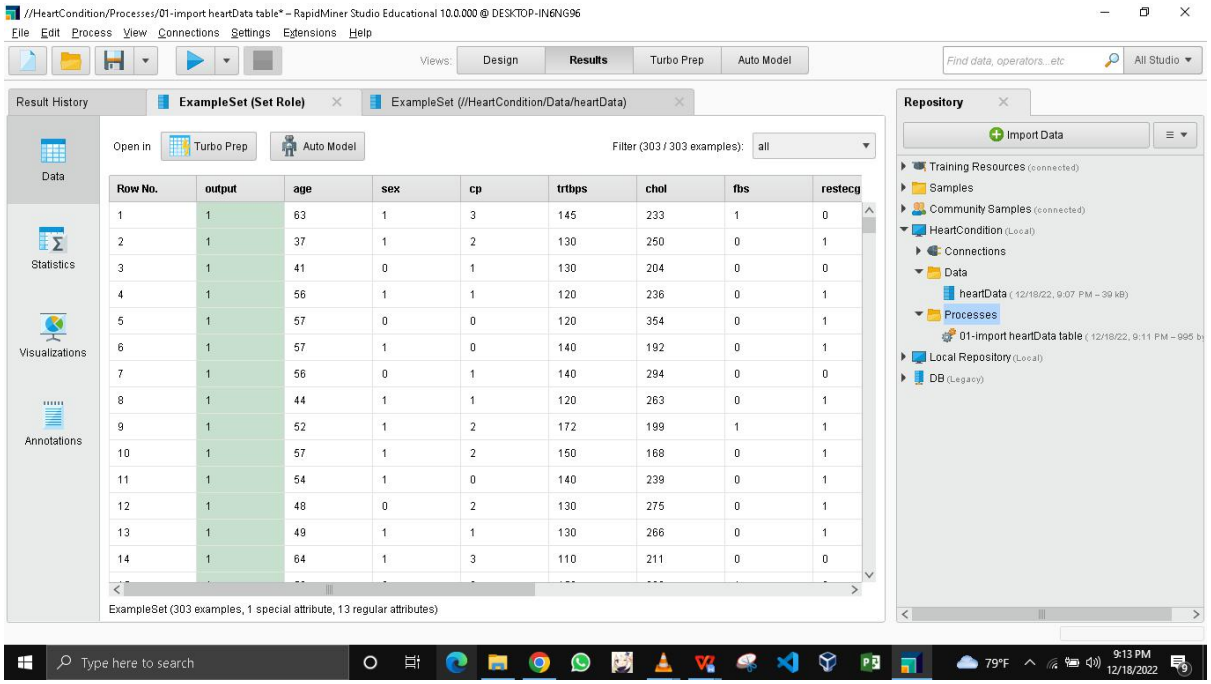
6. Defined the “Output” column within the heartData table as the Label using the “Set Role” operator



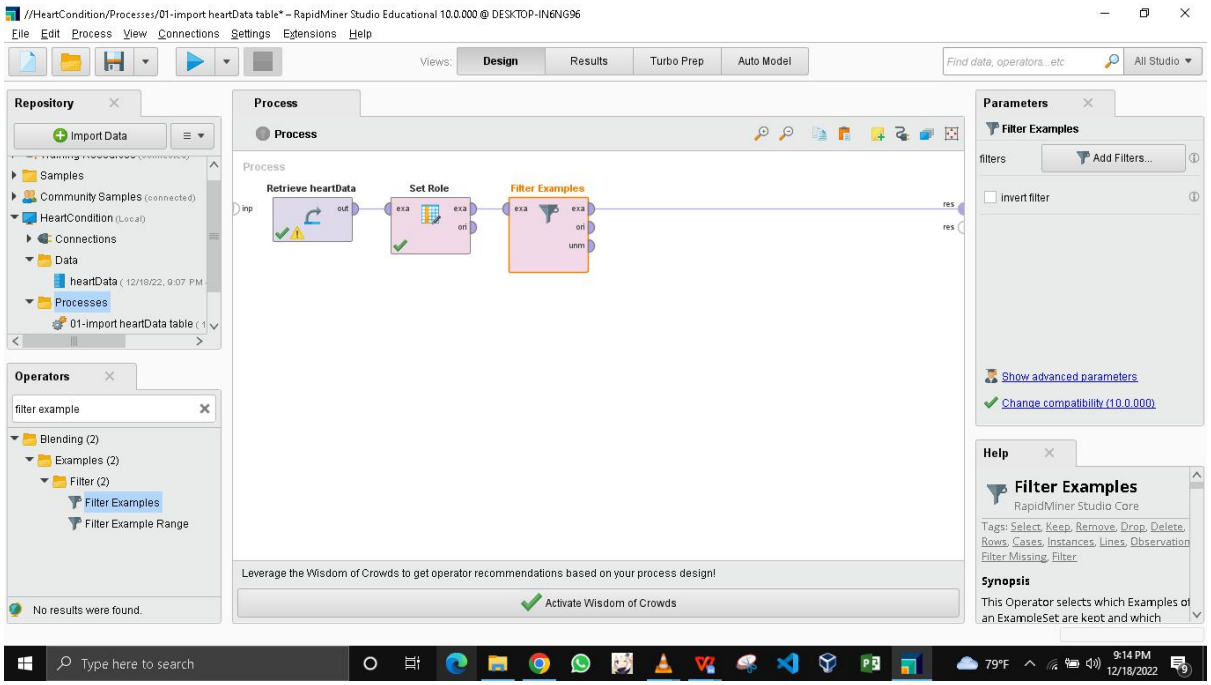


Abdhel Muthazz Mashood
Student ID : 21595621

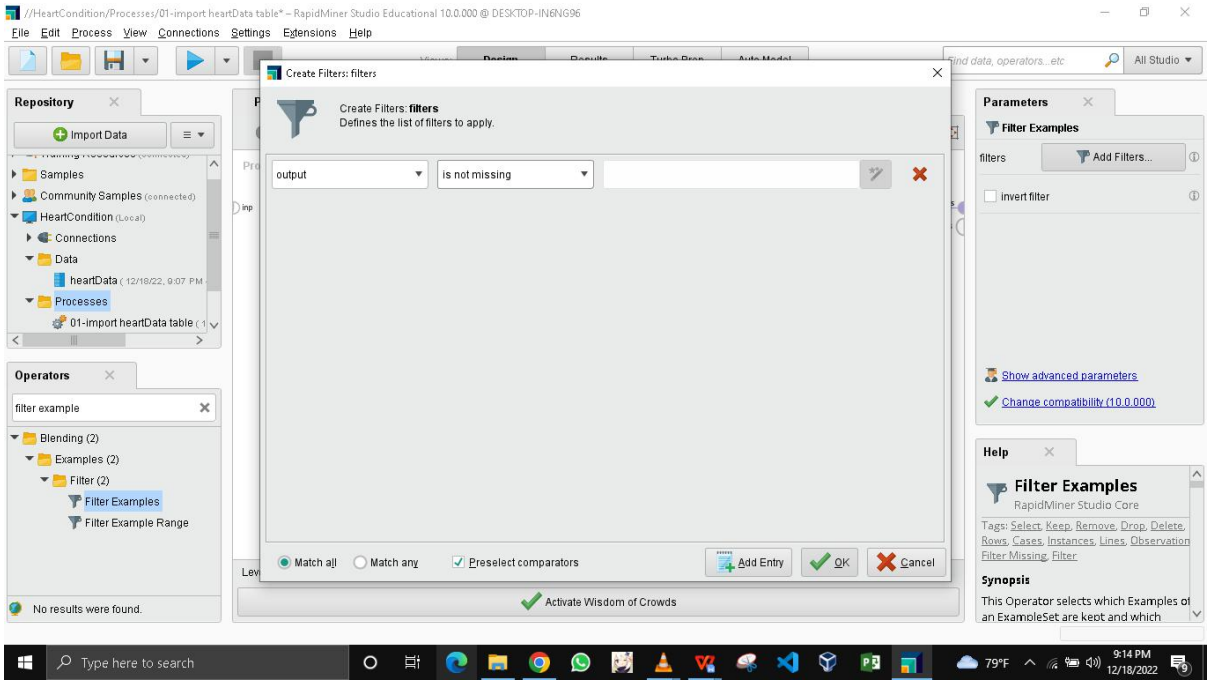
Shown below is a preview of the changes applied to the table after the change

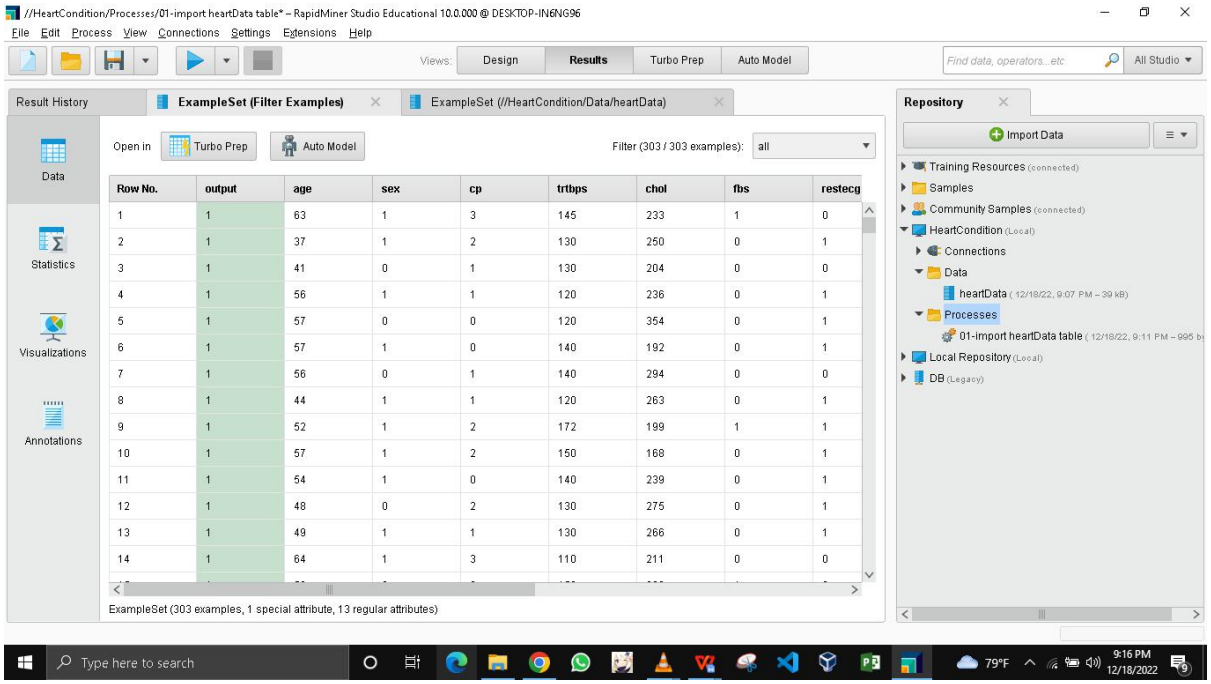


7. Filtered the heartData Table to only show fields under the “Output” column containing defined values using the “Filter Examples” operator



Abdhel Muthazz Mashood
Student ID : 21595621

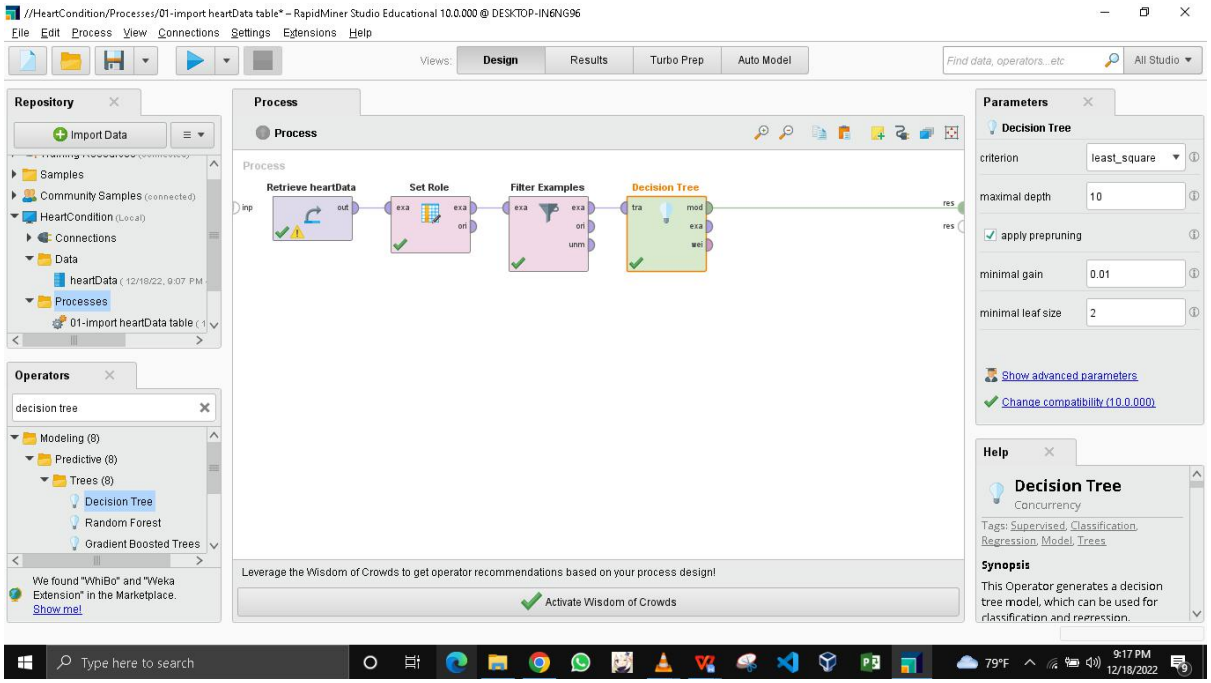


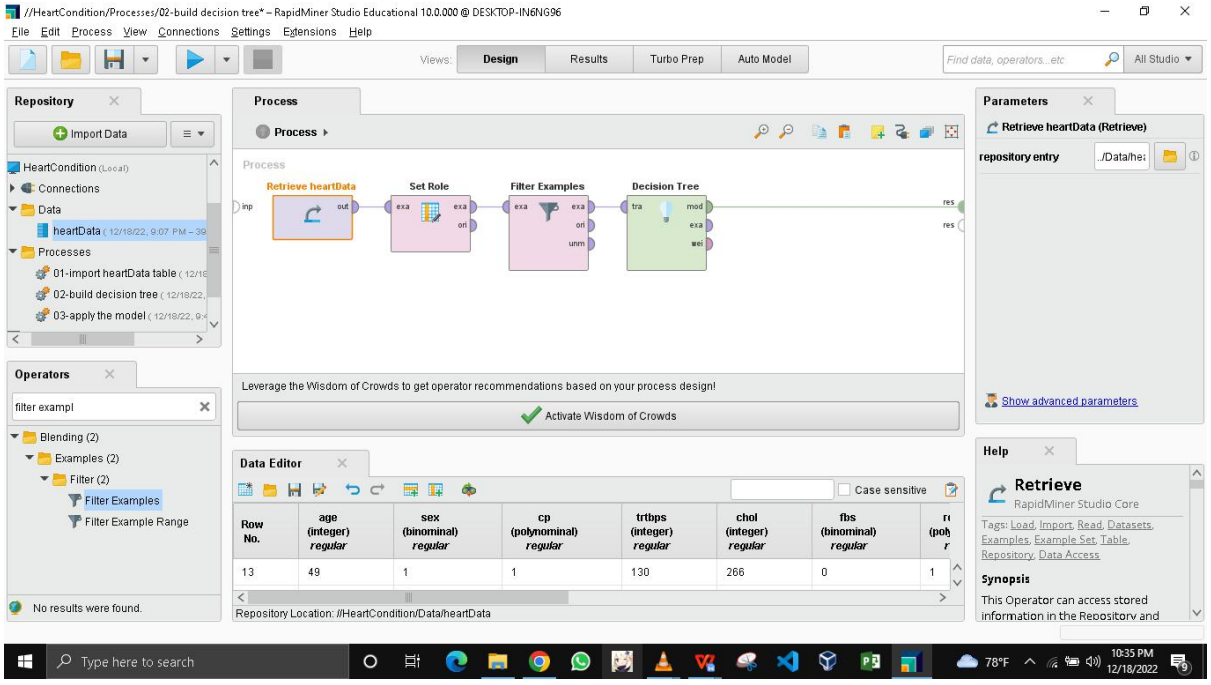


No changes take place as all the fields under the output column consist of defined values

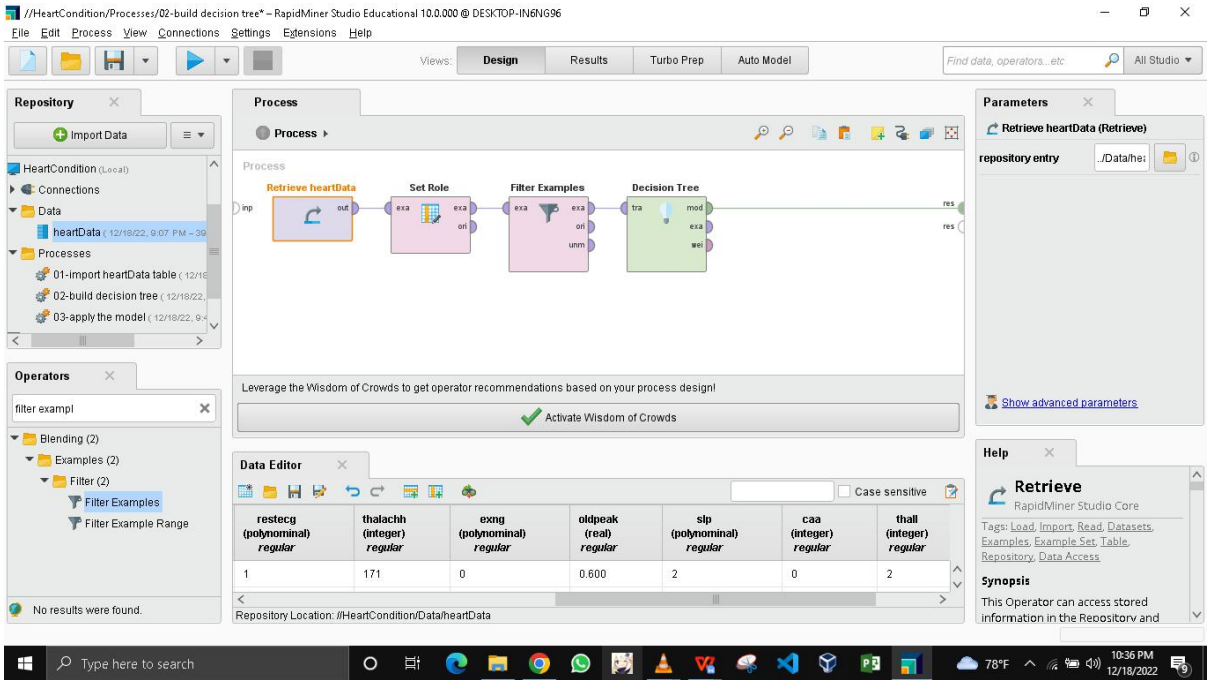
Abdhel Muthazz Mashood
Student ID : 21595621

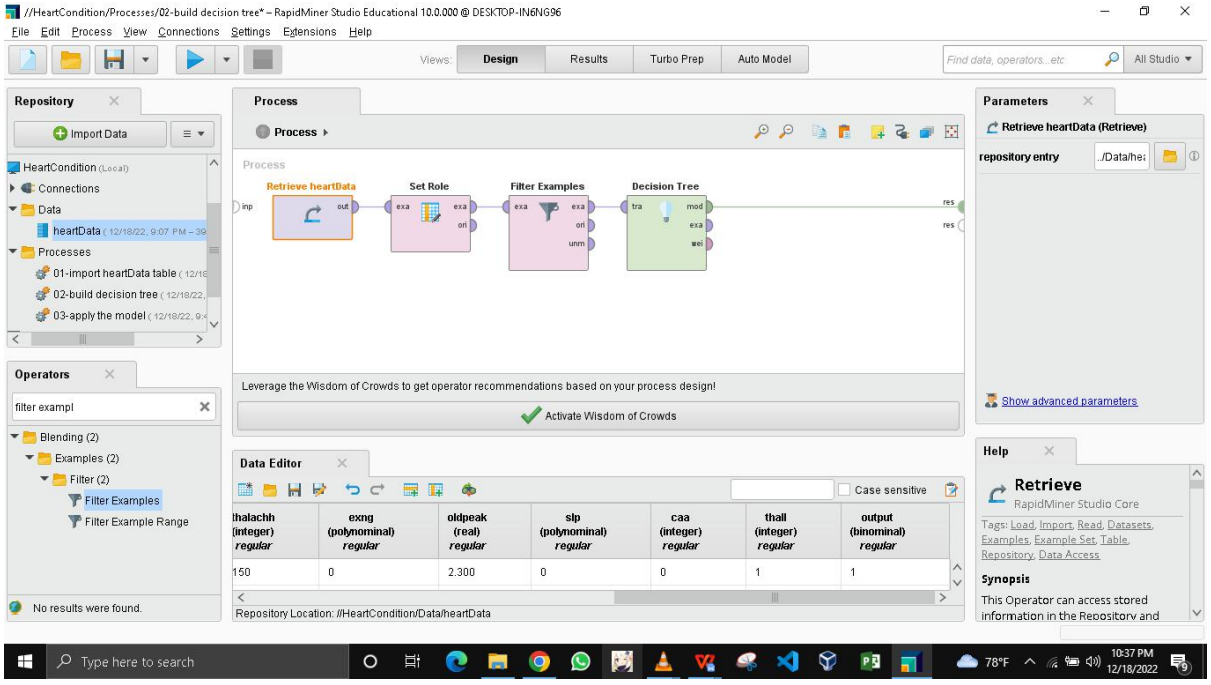
8. Created a Decision Tree of the heartData table using the “Decision Tree” operator after having changed the table attributes to better suited attributes



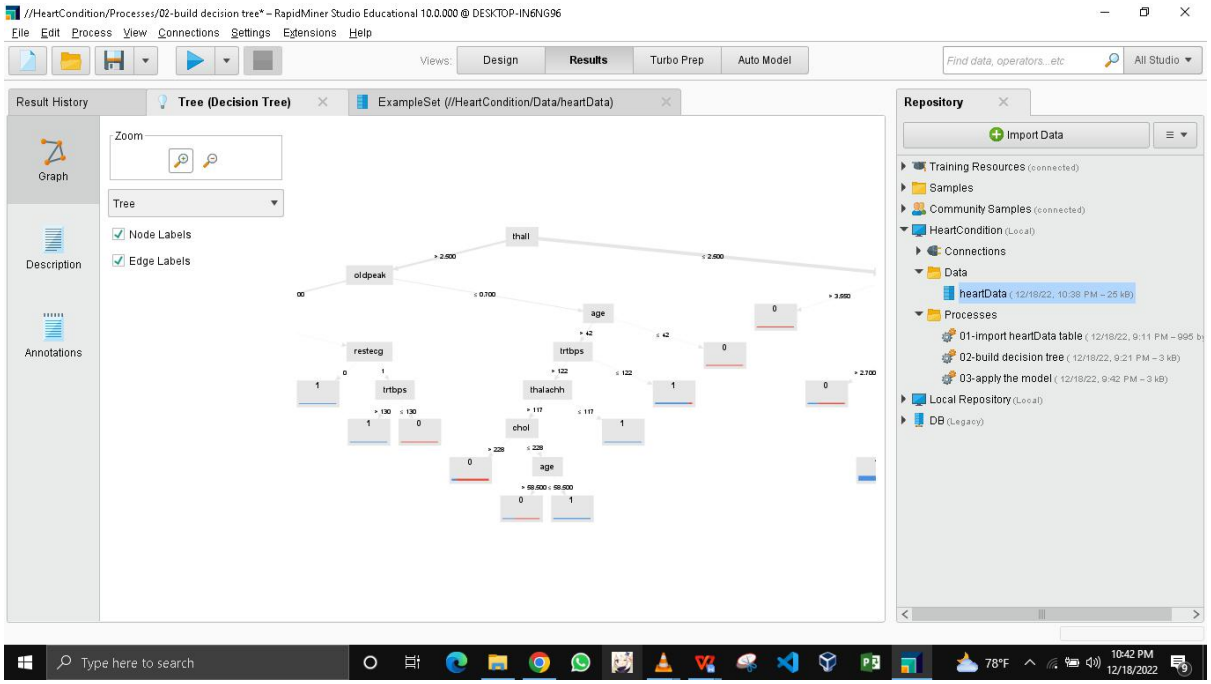


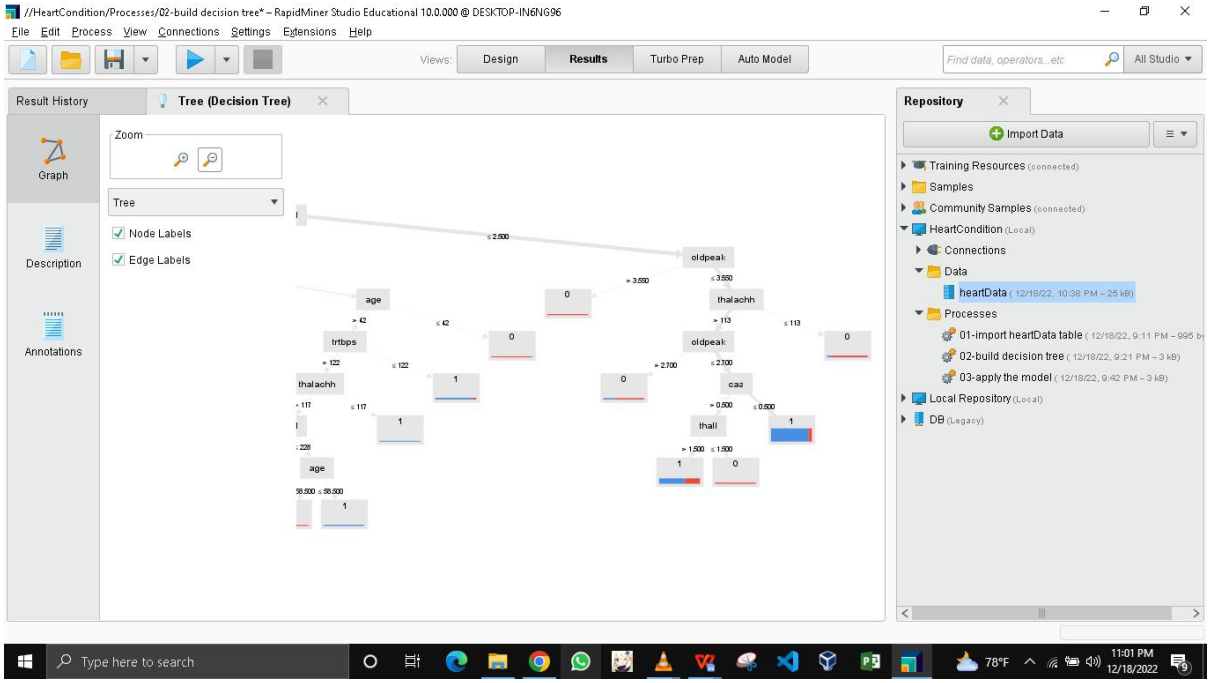
Abdhel Muthazz Mashood
Student ID : 21595621





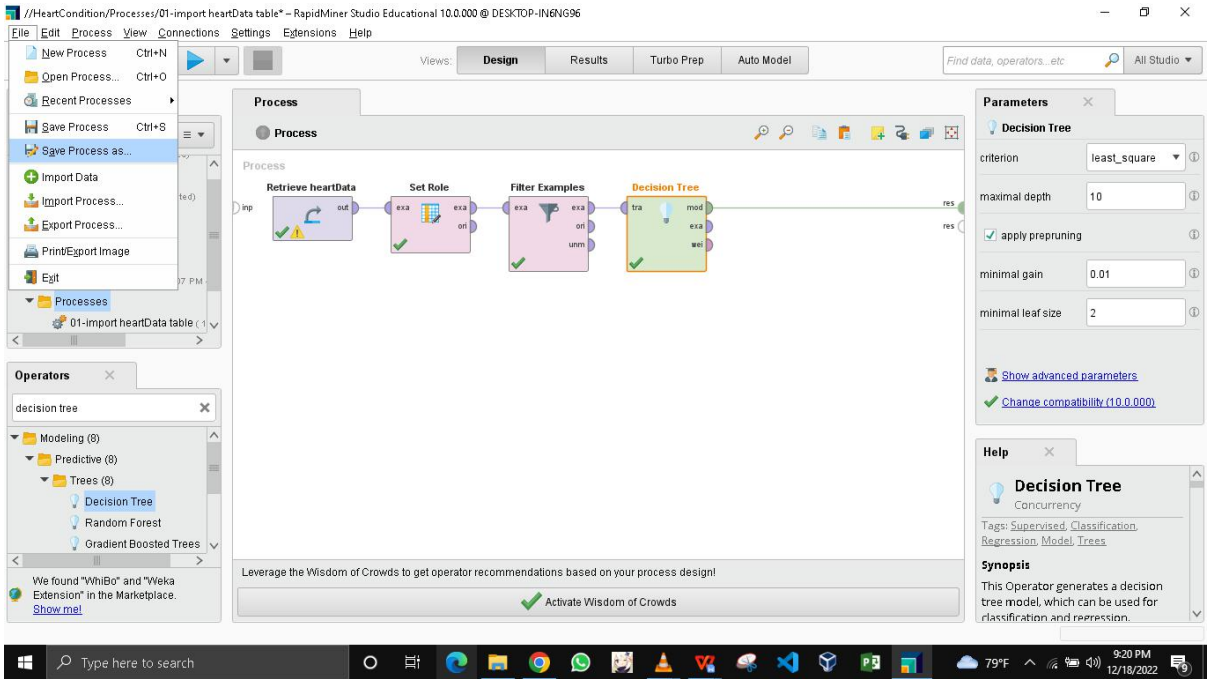
Abdhel Muthazz Mashood
Student ID : 21595621

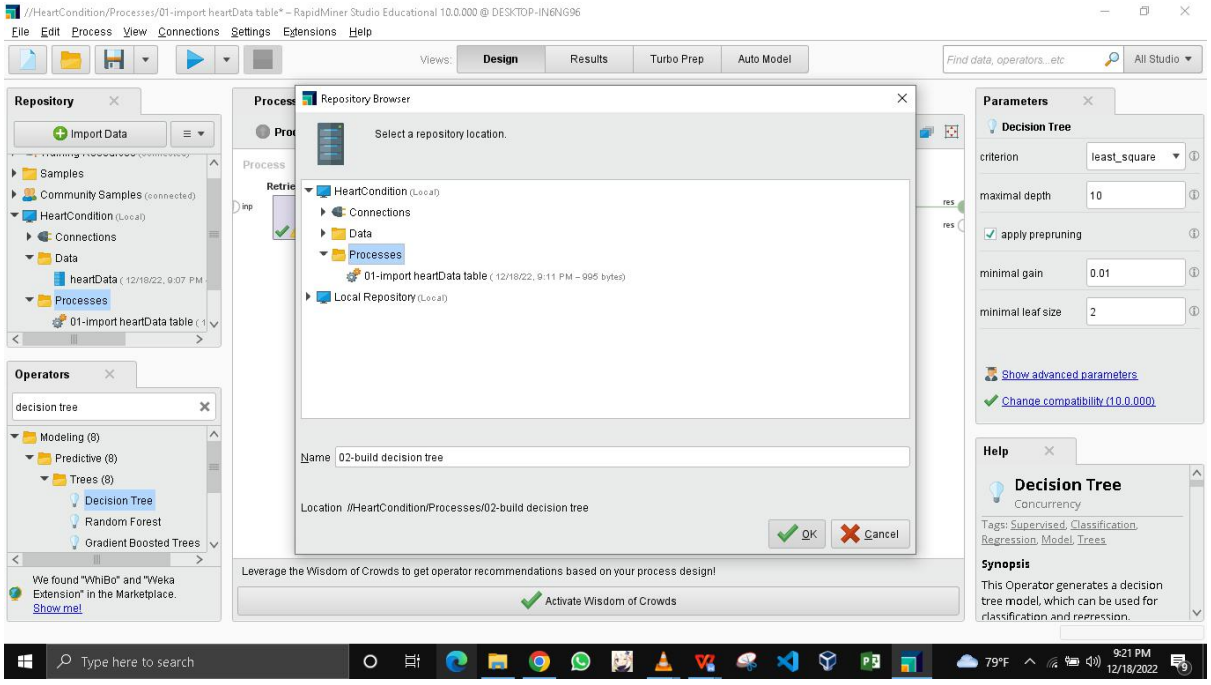


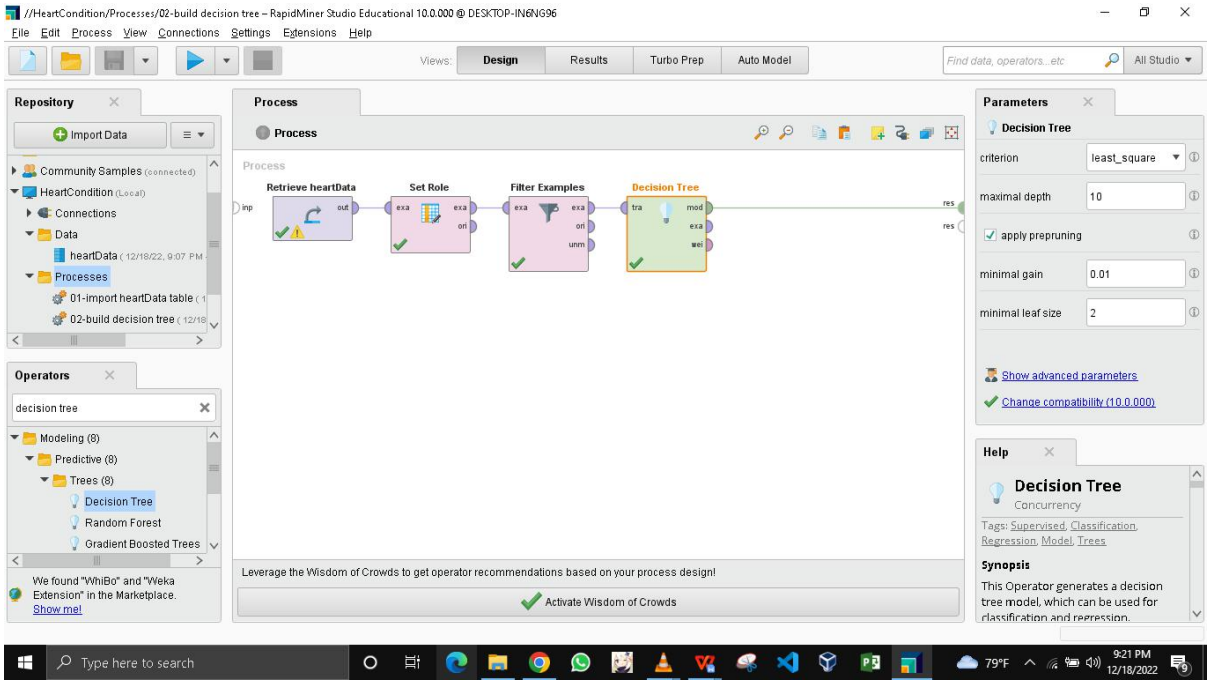


Abdhel Muthazz Mashood
Student ID : 21595621

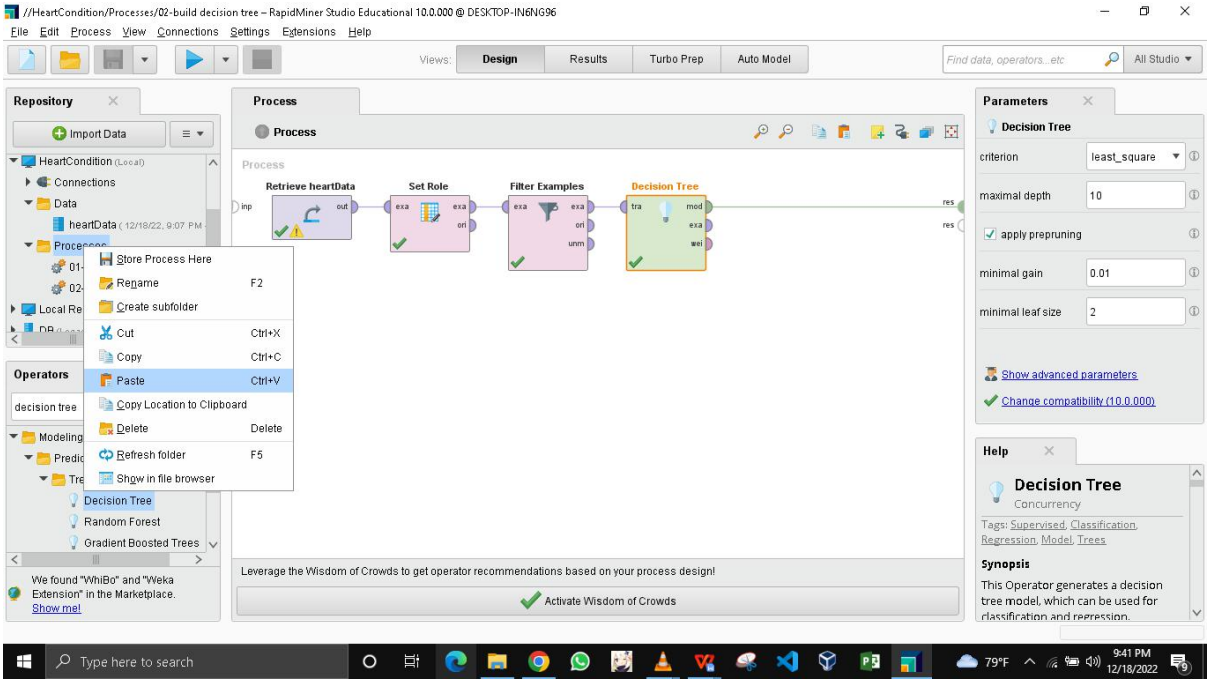
9. Saved the process as “02-build decision tree” in the “Processes” Folder



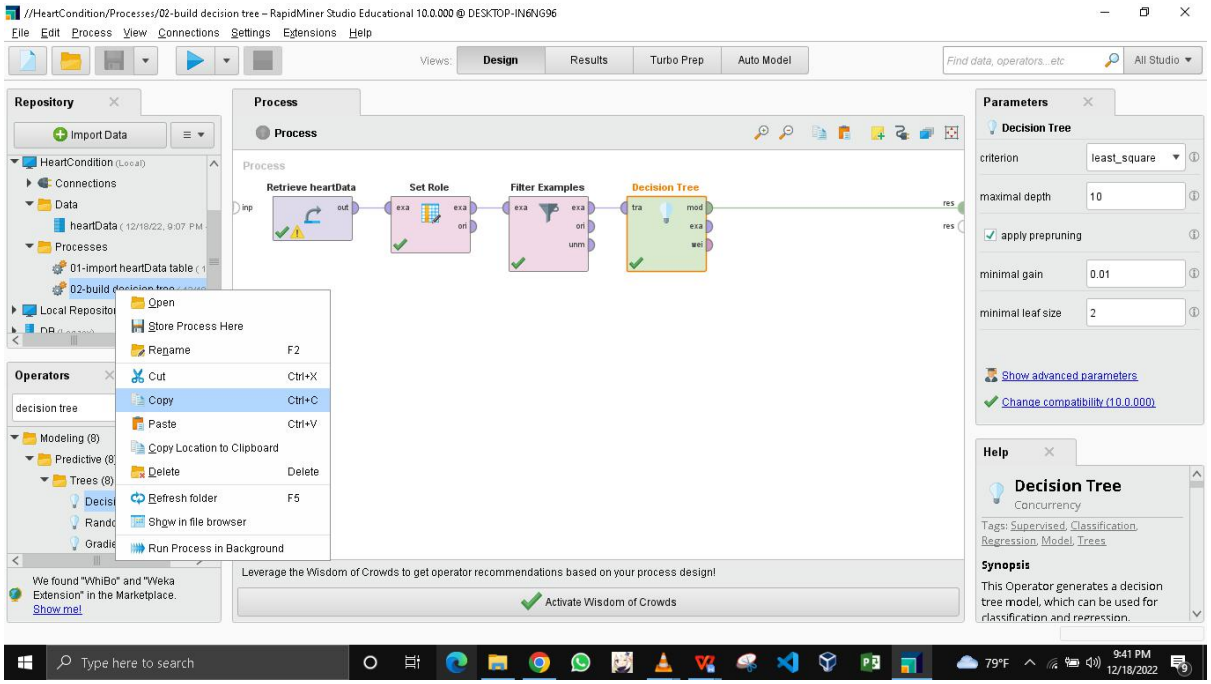


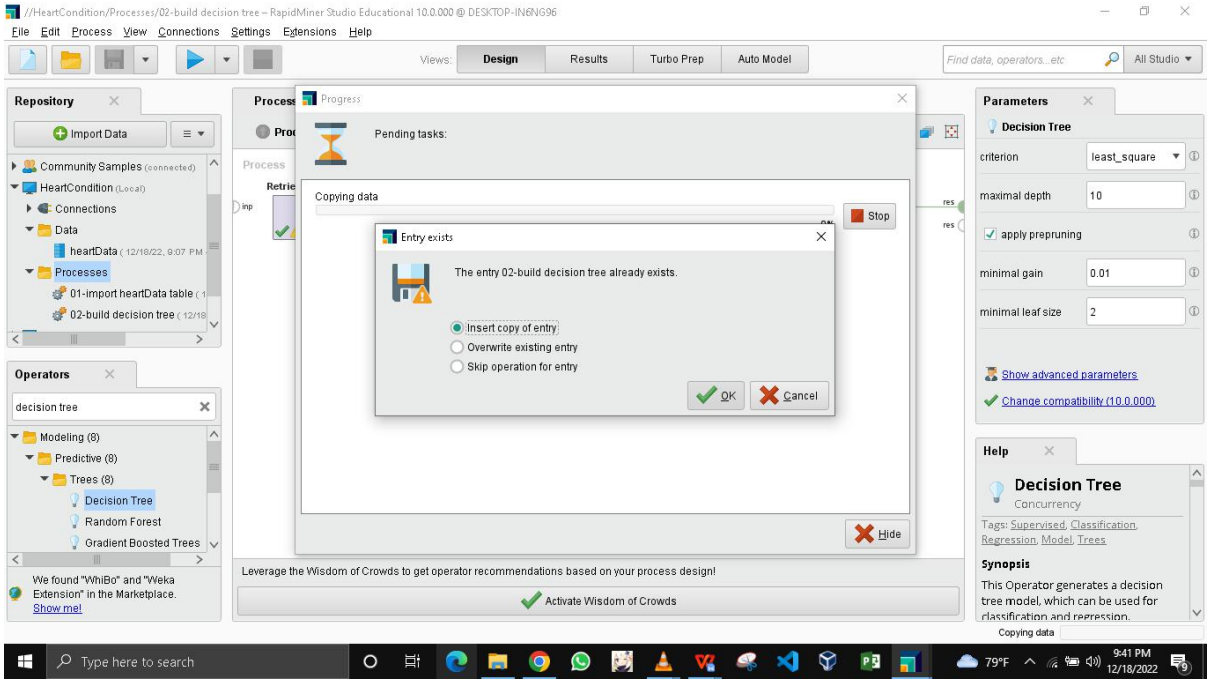


10. Created “03-apply the model” process using a copy of the “02- build decision tree” process

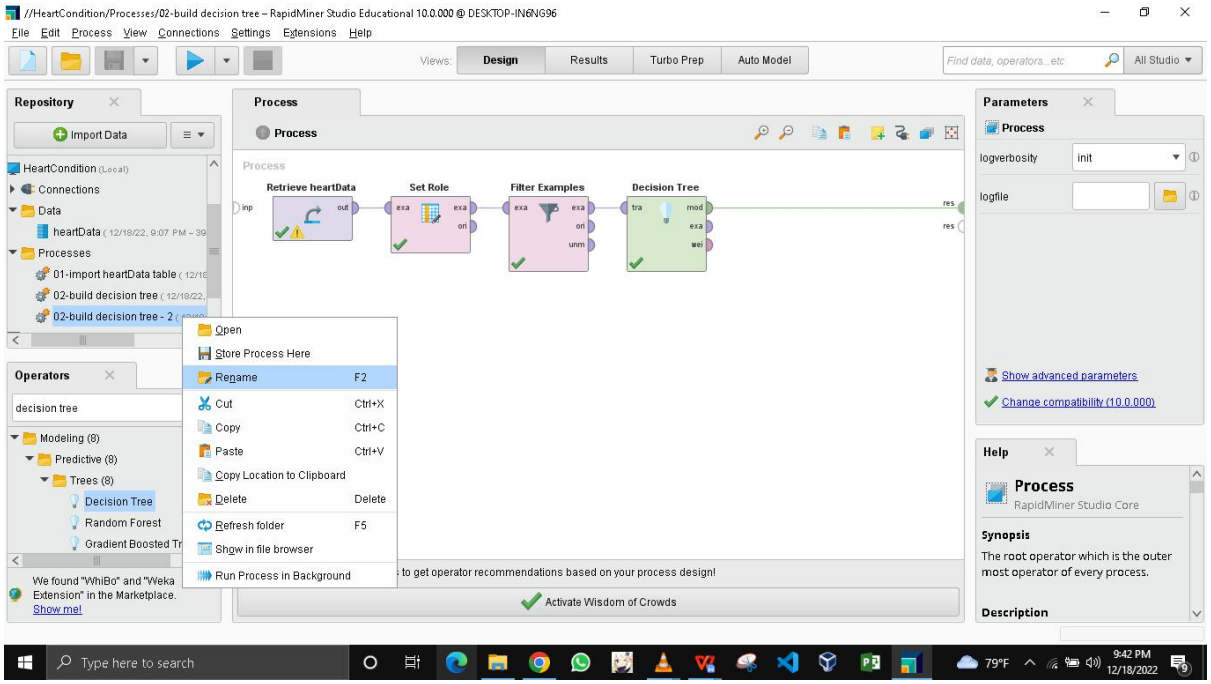


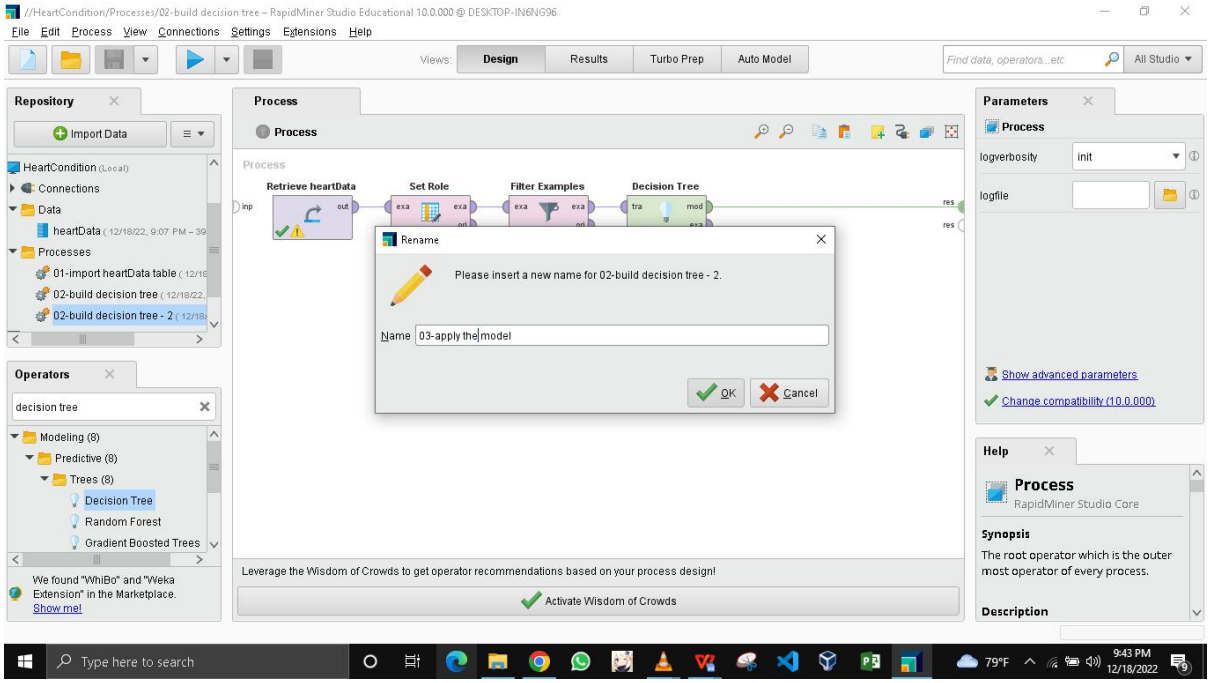
Abdhel Muthazz Mashood
Student ID : 21595621



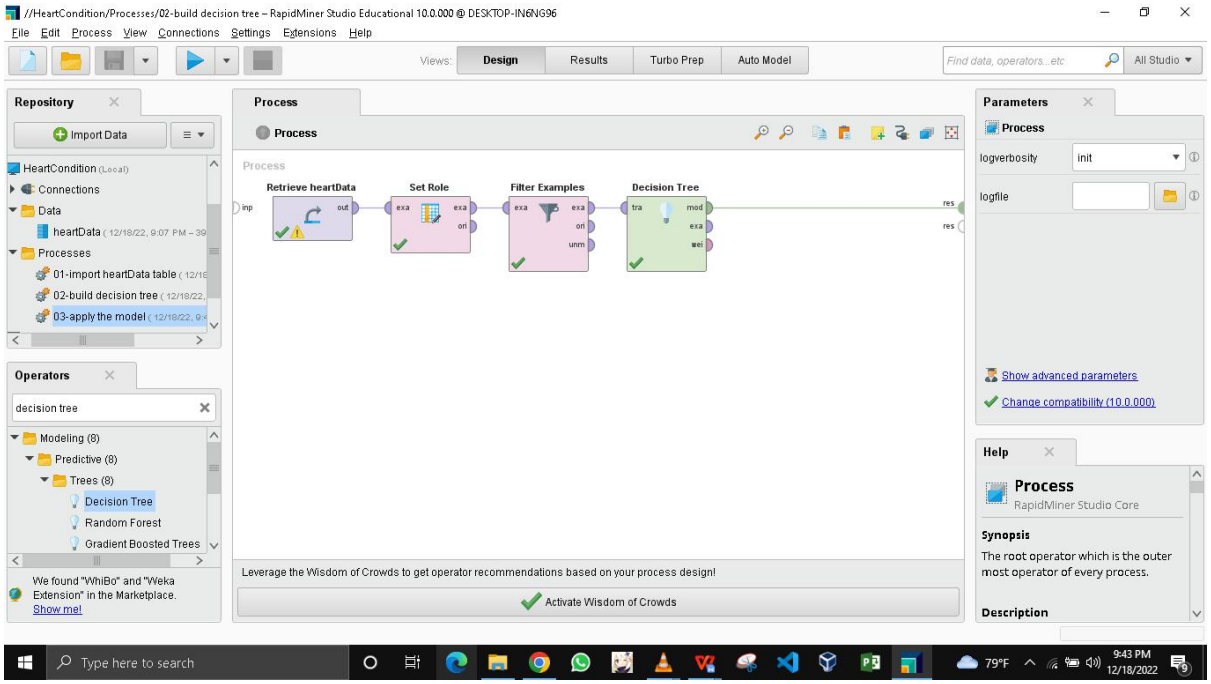


Abdhel Muthazz Mashood
Student ID : 21595621

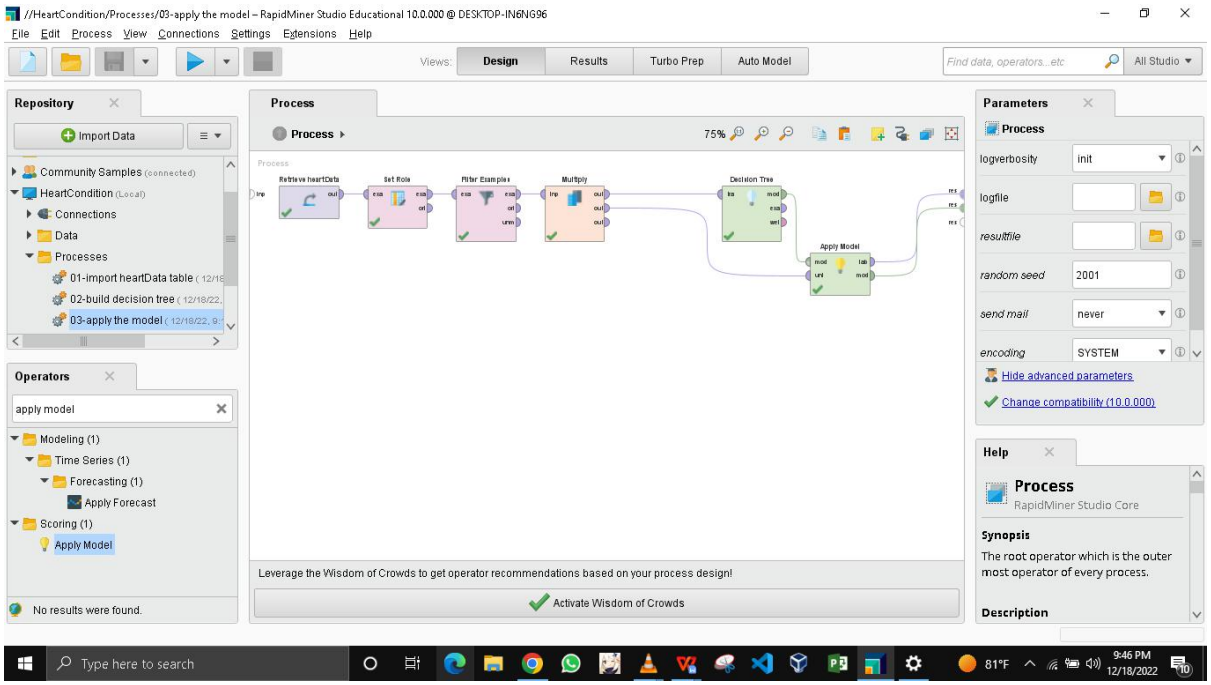




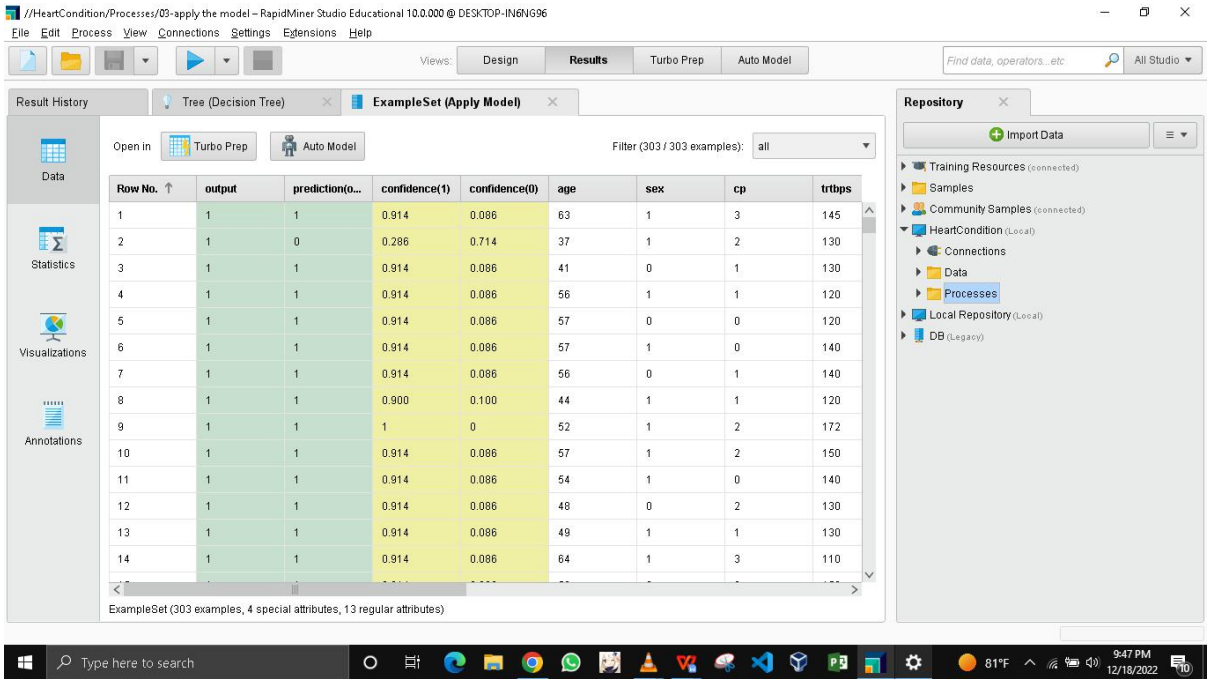
Abdhel Muthazz Mashood
Student ID : 21595621

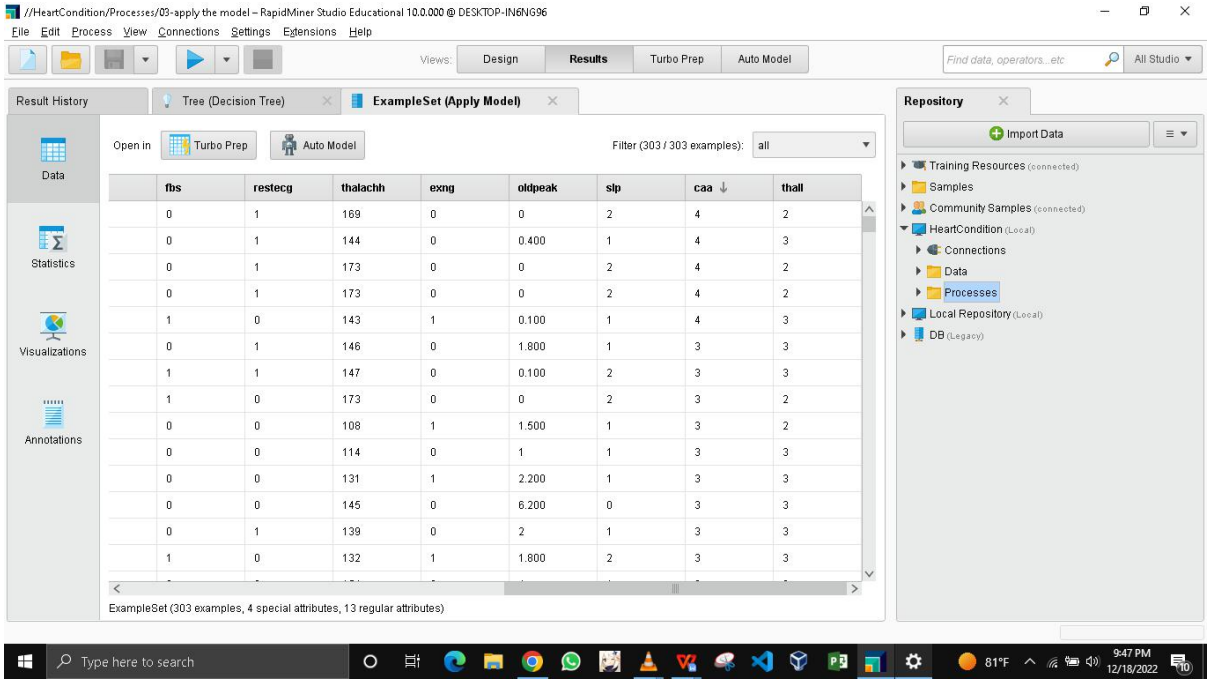


11. Used the “Multiply” and “Apply Model” operators to create a copy of the filtered data, and then get a prediction and the likeliness of the predicted results of the output respectively

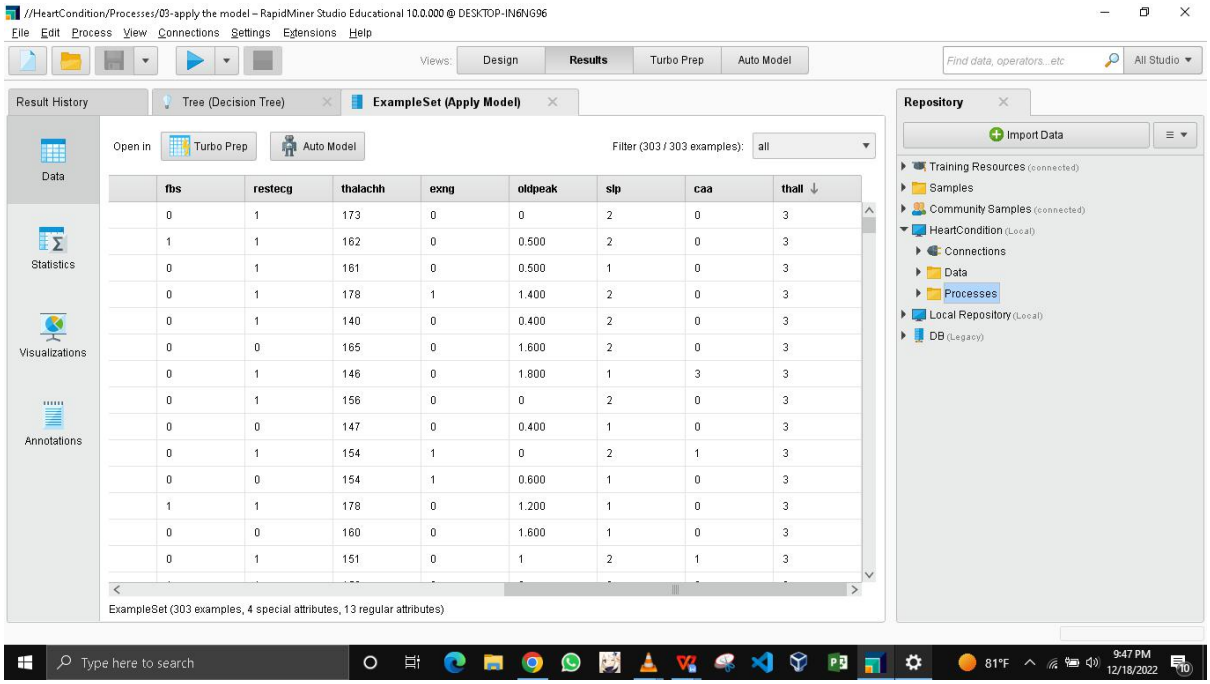


Abdhel Muthazz Mashood
Student ID : 21595621

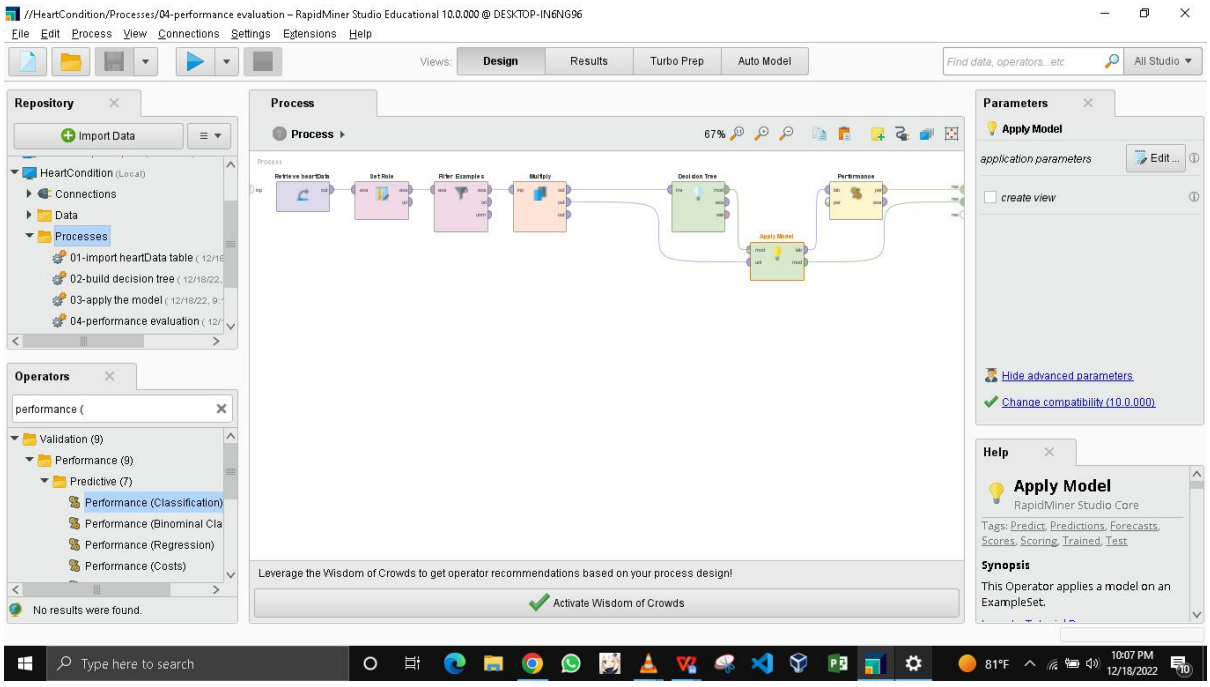




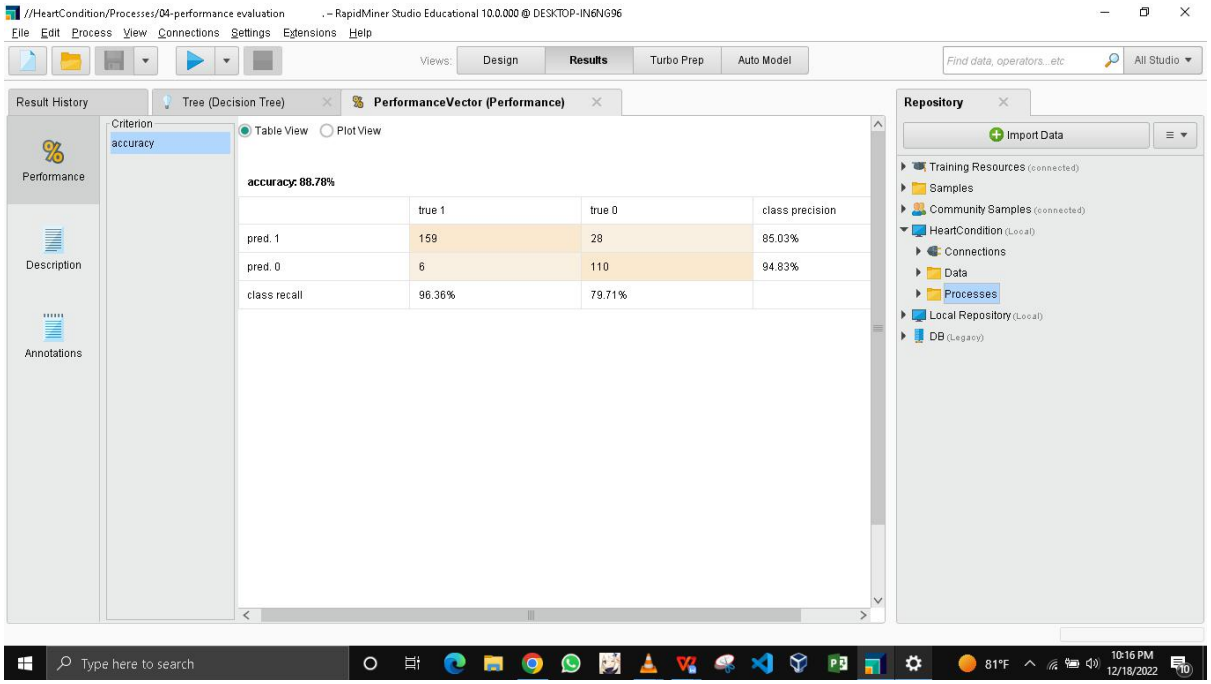
Abdhel Muthazz Mashood
Student ID : 21595621



11. Added the “Performance (Classification)” operator to the model to evaluate the model’s performance in terms of overall prediction accuracy and saved the process as “04-performance evaluation”

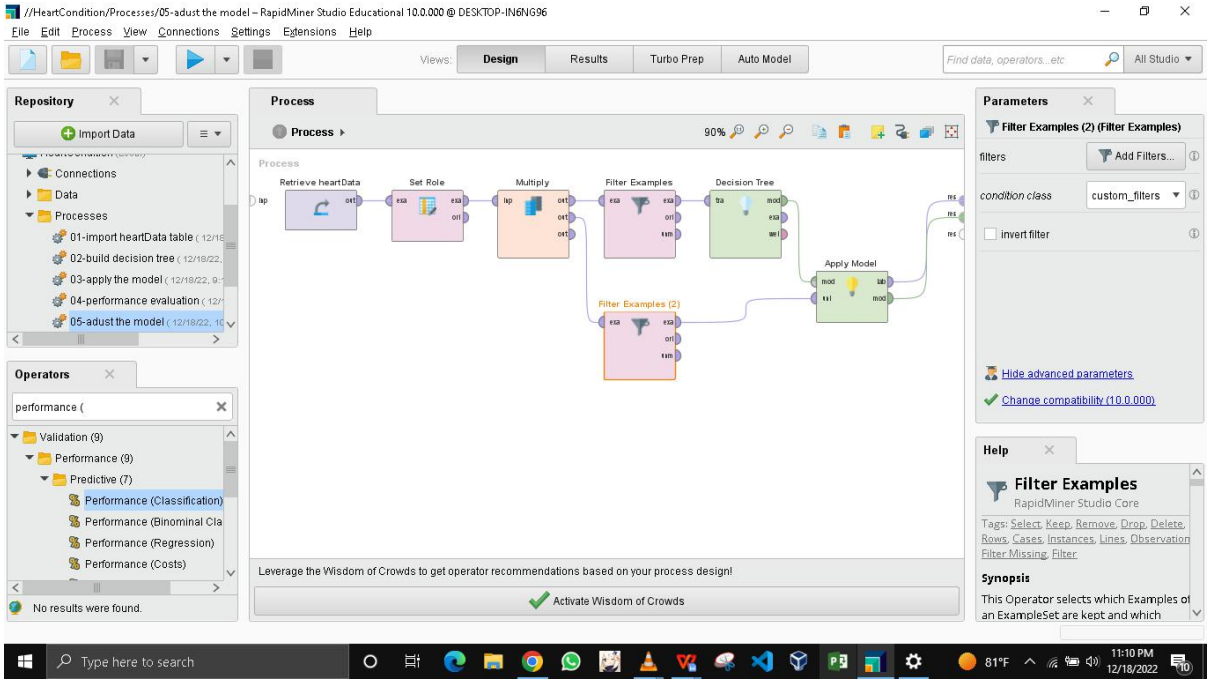


Abdhel Muthazz Mashood
Student ID : 21595621

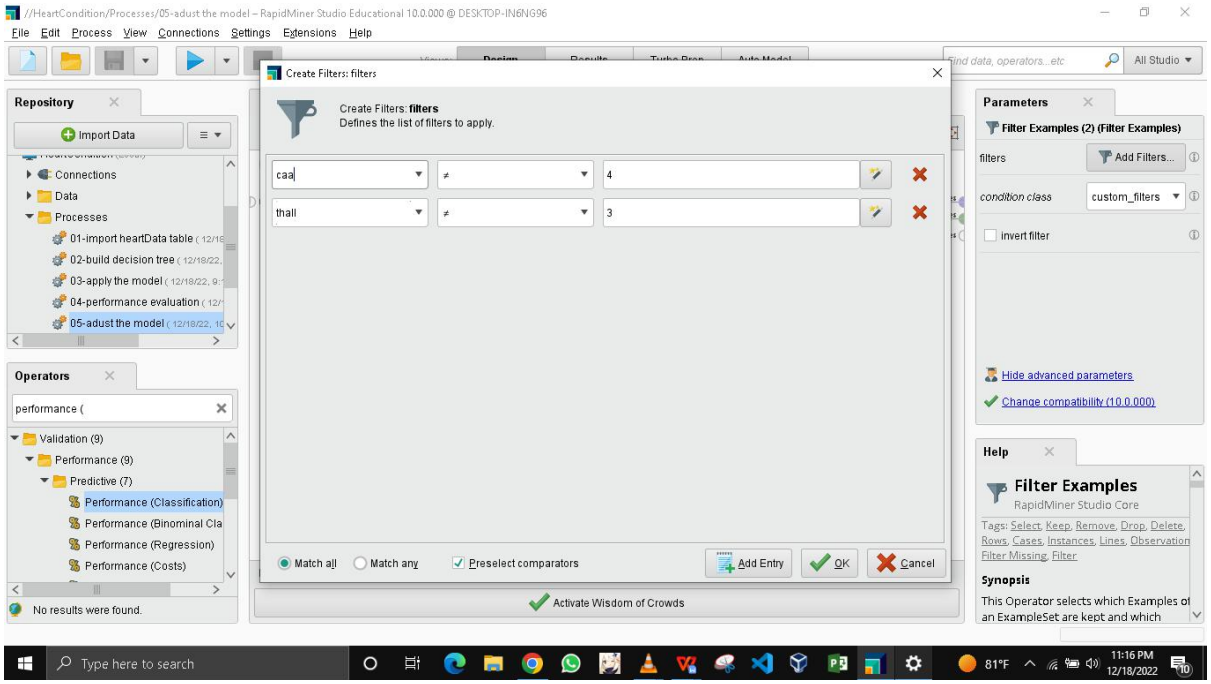


As indicated by the screenshot above the overall prediction accuracy of the current model is at 88.78% however this is not indicative of realistic results

12. Created an “05-adjust the model” process with the “Multiply” operator repurposed and re-positioned in order to be able to feed the same set of data into differently separate filters with the parameters of filtration for the bottom filter as shown in the following screen captures below



Abdhel Muthazz Mashood
Student ID : 21595621



With the reason being to eliminate and prevent undefined anomalies from interfering with the accuracy of the results as the defined range for the category “caa” as shown below is only in between (0-3) with the value 4 not included in the predefined range

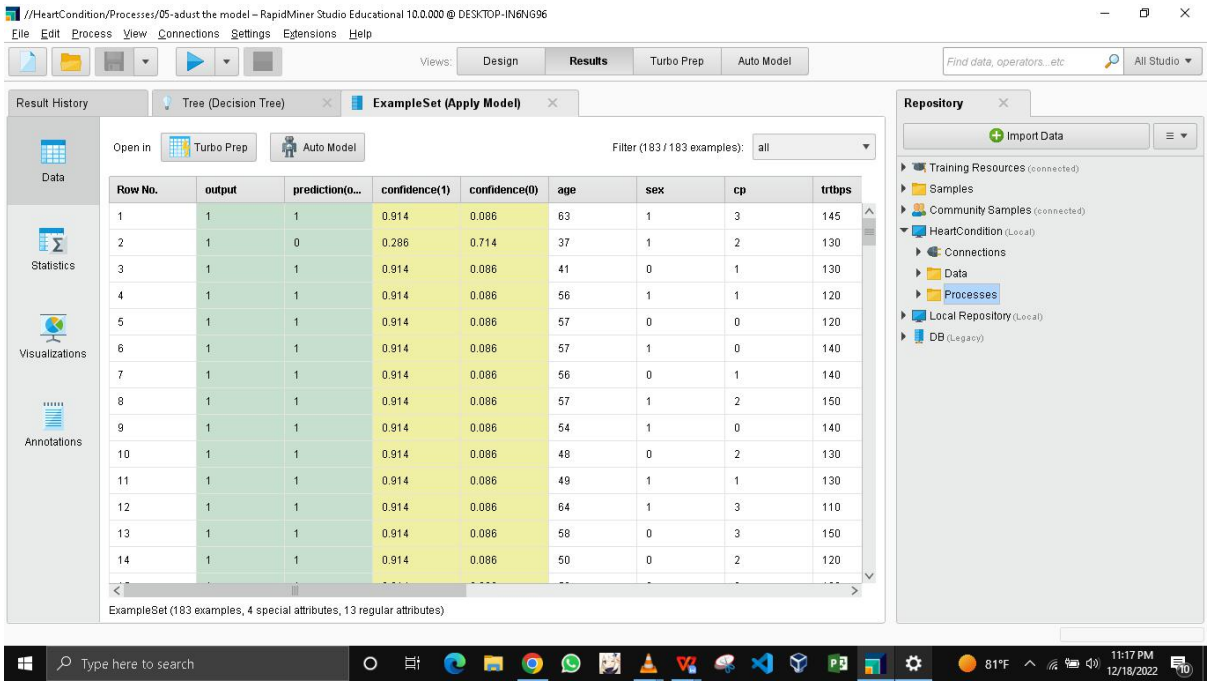
- caa: number of major vessels coloured by flourosopy (0-3)

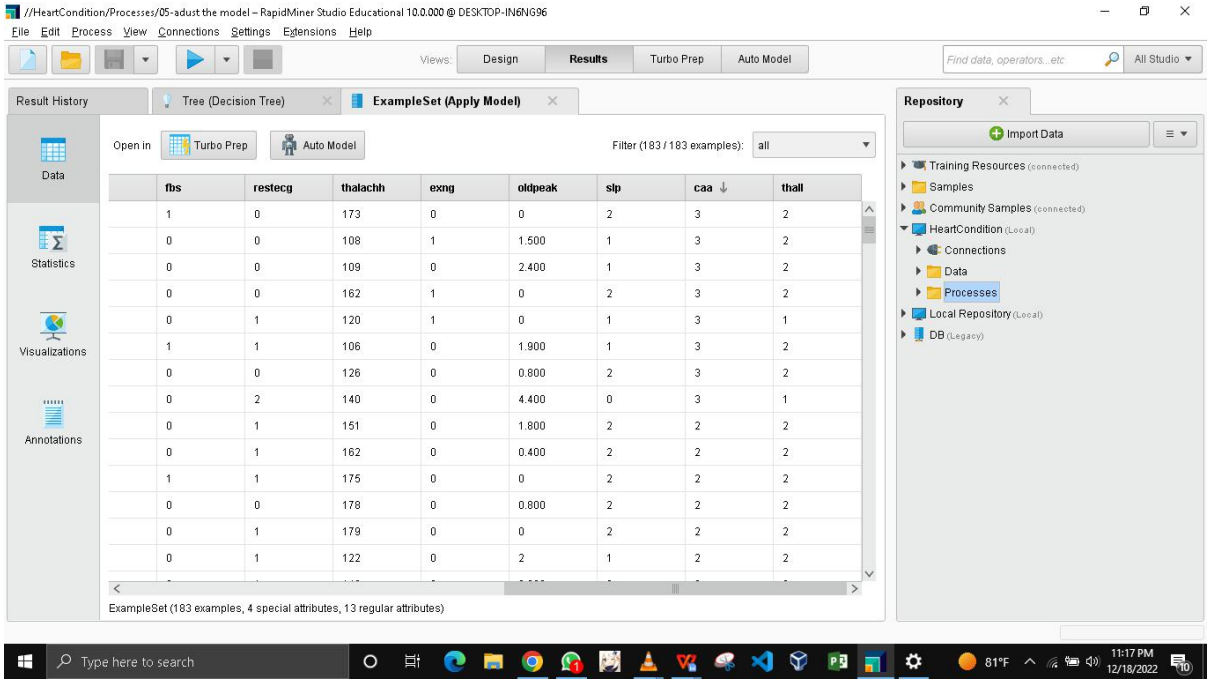
Whilst the category labeled “thall” consists only of definitions for the numbers 0-2 without a definition existing for the undefined integer 3 as shown below

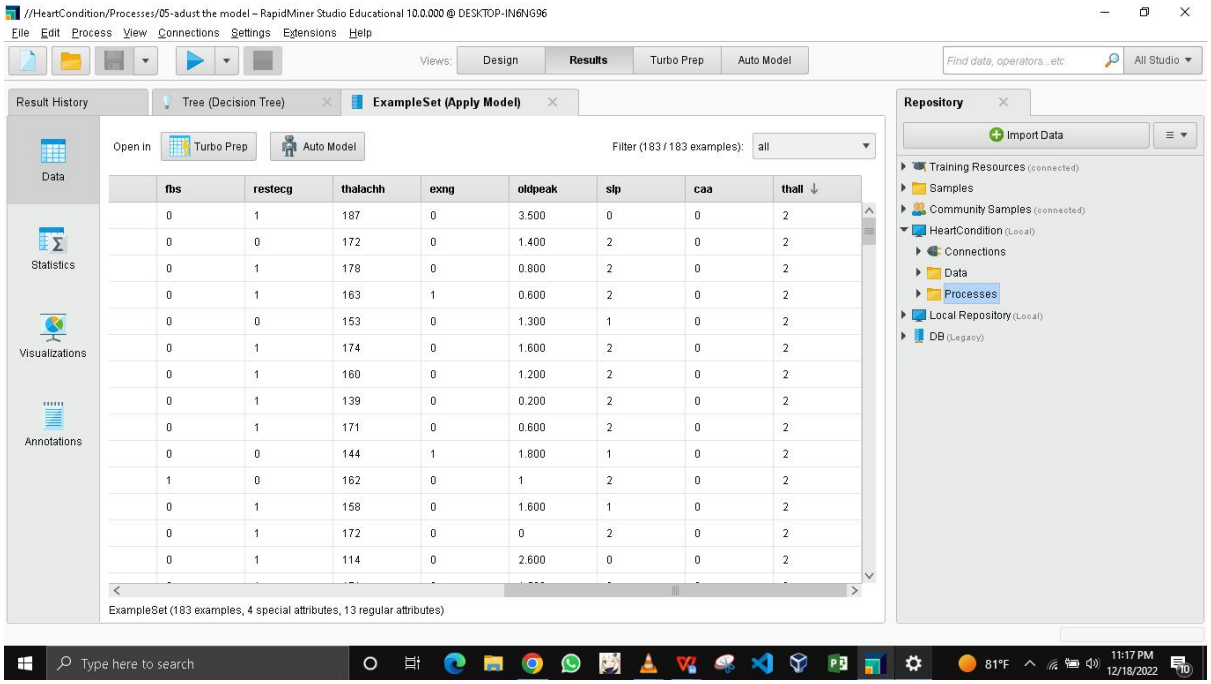
- thall: 0=normal; 1=fixed defect; 2=reversable defect

Abdhel Muthazz Mashood
Student ID : 21595621

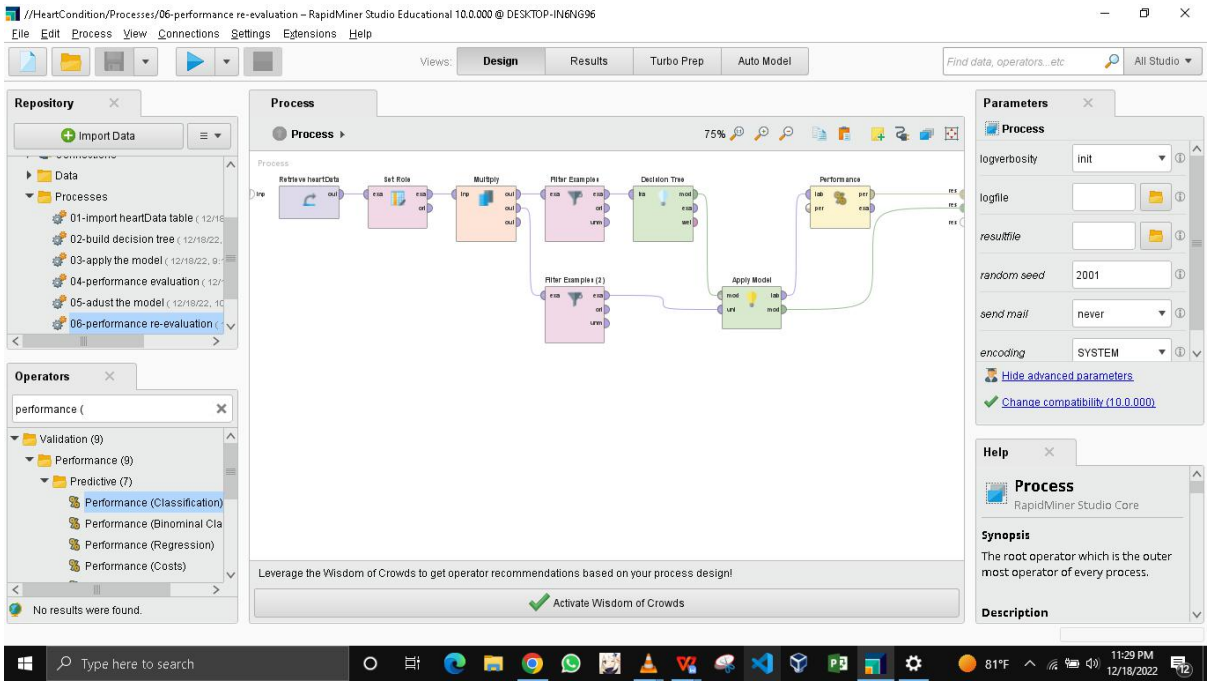
With these anomalies taken into consideration the results of the post filtered data are as follows with the number of examples having decreased from 303 to 183 in the ExampleSet







13. Tested the adjusted model using the “Performance (Classification)” operator and saved the process as “06-performance re-evaluation” with the results indicating an accuracy of 84.15% as shown below



Abdhel Muthazz Mashood
Student ID : 21595621