**Cairo University**

**Faculty of Computers and Artificial Intelligence**

**(2021-2022)**

***Forecasting & Predictive Analytics***

**Dr. Olivia Murad**

# Hotel Booking Prediction

## Table of Content



## Chapter 1: **Problem Definition**

In this project we have been working on freshen up the tourism in Egypt by working on Data Set - provided later on Chapter 2 – to guide either the Government or the Investors to which type of hotels – City Hotels or Resort Hotels – they need to pump their cash flow into, as here in Egypt we have many attractive sites to be visited, not only the wonderful beaches and sea landmarks, but we also have many tourist attractions sites in the city, every city in Egypt has at least one or two incredible sites which must be visited.

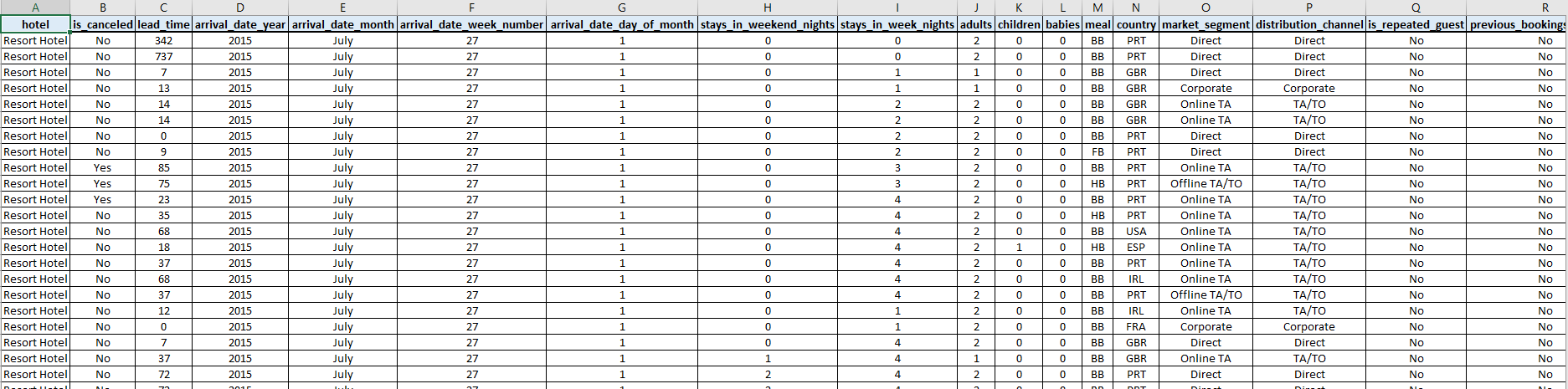
In our project we considered too many attributes that may affect our decision, the decision mainly will be use the time factor as a guide, if we filtered the data in details we will find out that in specific months most of hotels booking going to one type of hotels, this condition they called in business season, as we all know a city like Alexandria which fall under the resort hotels type most of people or let me be specific by saying most of tourists go there in specific months like June, July and August. We have another example to a city like Luxor, In Luxor most the tourists come in months like December, January, and February.

We will discuss in details in the upcoming chapter the data set, attributes, and the targeted variables, then we will move into the technical work which means the algorithm implementation, finally we will have a small comparison between algorithms and each other, in this comparison we will consider many factors to determine a rank for each algorithm. This comparison also we will use to get a conclusion for the whole project.

## Chapter 2: **Data Set**

This data set contains booking information for city & resort hotels, and includes information such as when the booking was made, length of residence, the number of adults, children, and/or babies, and the number of available parking spaces, etc.  
All personally identifying information has been removed from the data.

Here we will find a screenshot from the data



We have many attributes in this data set, we have 31 attributes. Some of these attributes is more efficient than others, and some of them is useless and we already ignored.

Our data set was 120,000 Rows, but we were obligated to remove almost 115,000 rows, because the max number of rows that can be processed by Rabid Miner is 5,000 Rows, so we worked on cleaning it by choosing the nearly best attributes which indicates to the best accuracy for algorithms.

Our Targeted Attribute is “Hotels”, which has two categories or two types “City Hotels” & Resort Hotels.

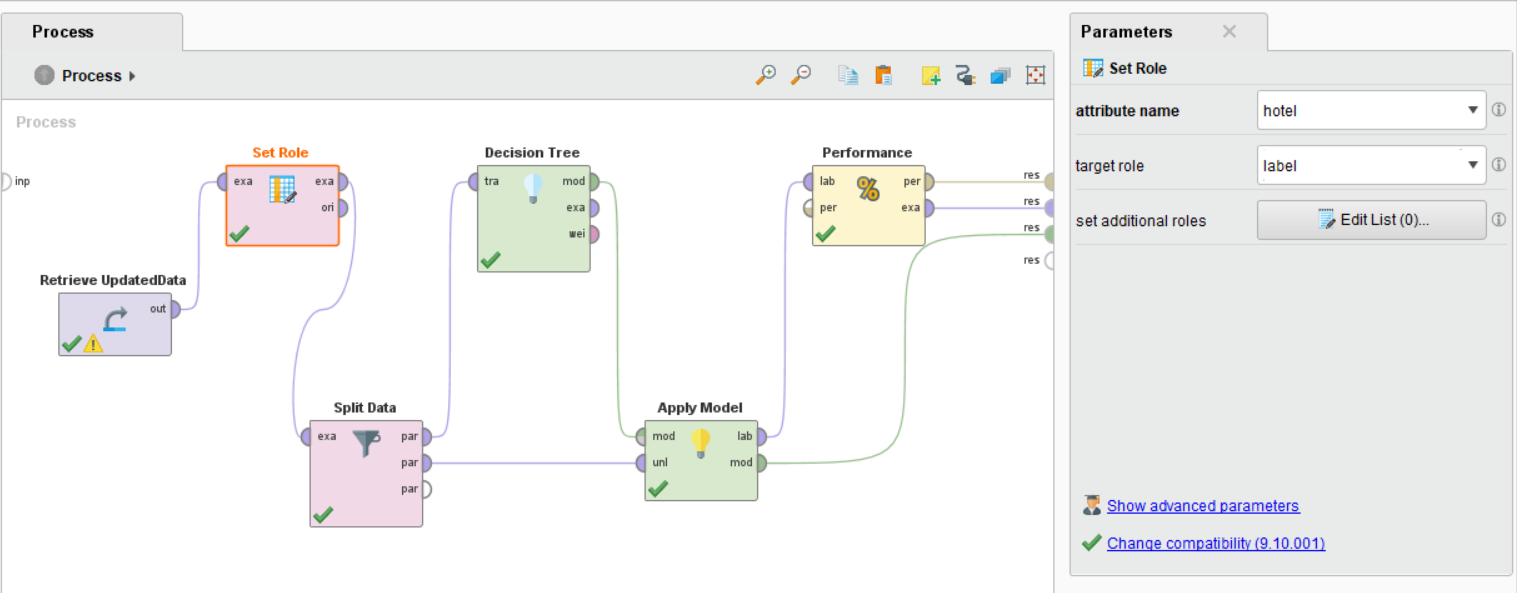
We will analysis multiple algorithms on this data set to determine our main objective of this project later on the upcoming chapter

## Chapter 3: **Algorithms Implementaion**

### **Decision Tree Algorithm**

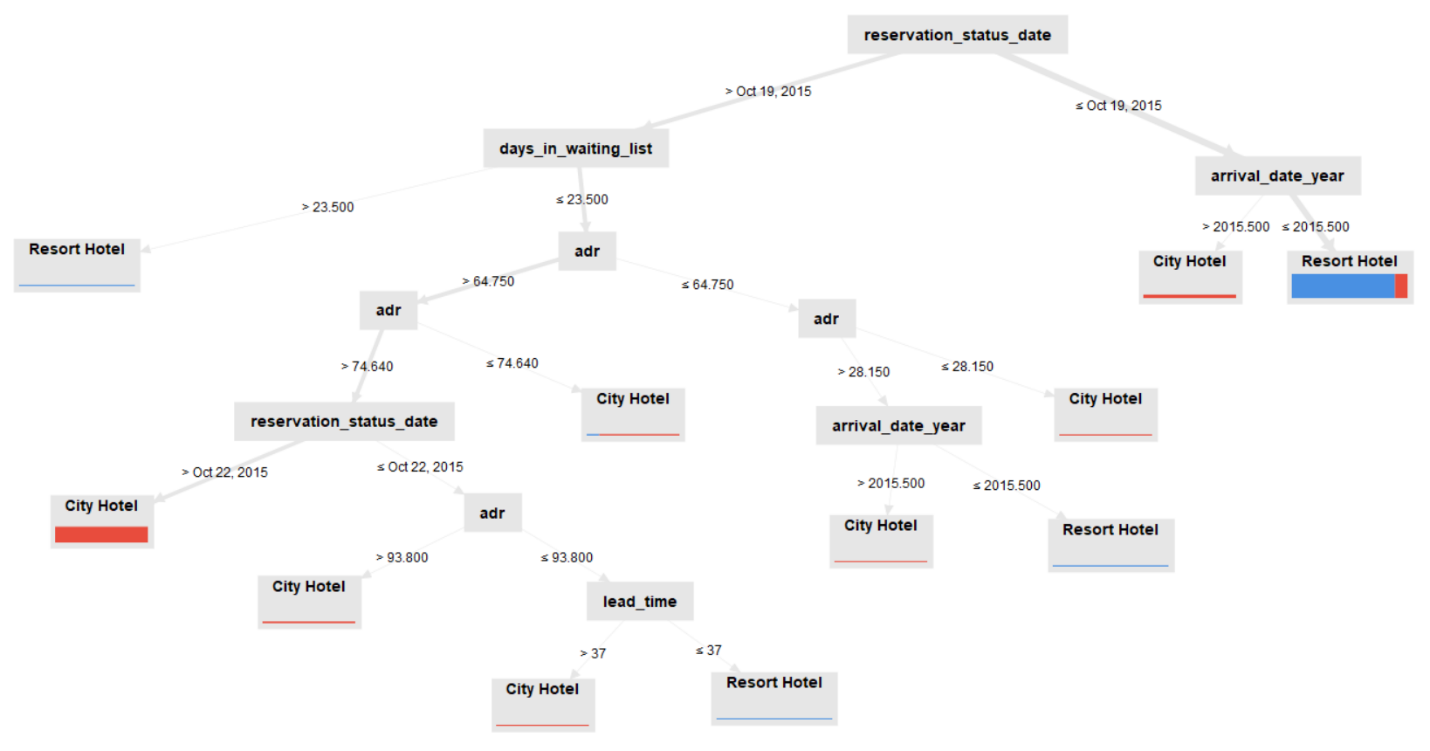
We will attach a screenshot for each step for the Decision Tree algorithm below.

Here we will find the process design:

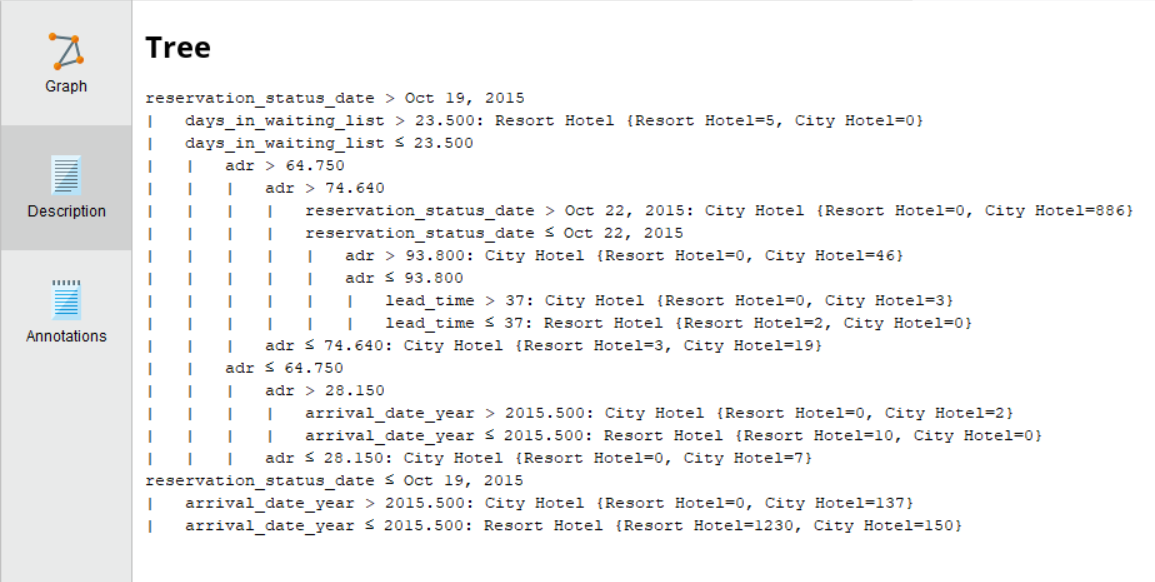


First of all, The UpdatedData refers to the last version of our data to make prediction on, then there’s the Set Role operator to determine which is the targeted variable - attribute: Hotel - then it goes to Split Data operator which splits data into two data sets, one called training data and it goes to the Decision Tree operator to train the data on, and the other one goes to operator called Apply Model to be predicted later after the Decision Tree operator got trained from the first splitted data, then we sent them after applying the model from the training data on the testing data we will go to the Performance operator to measure our model accuracy.

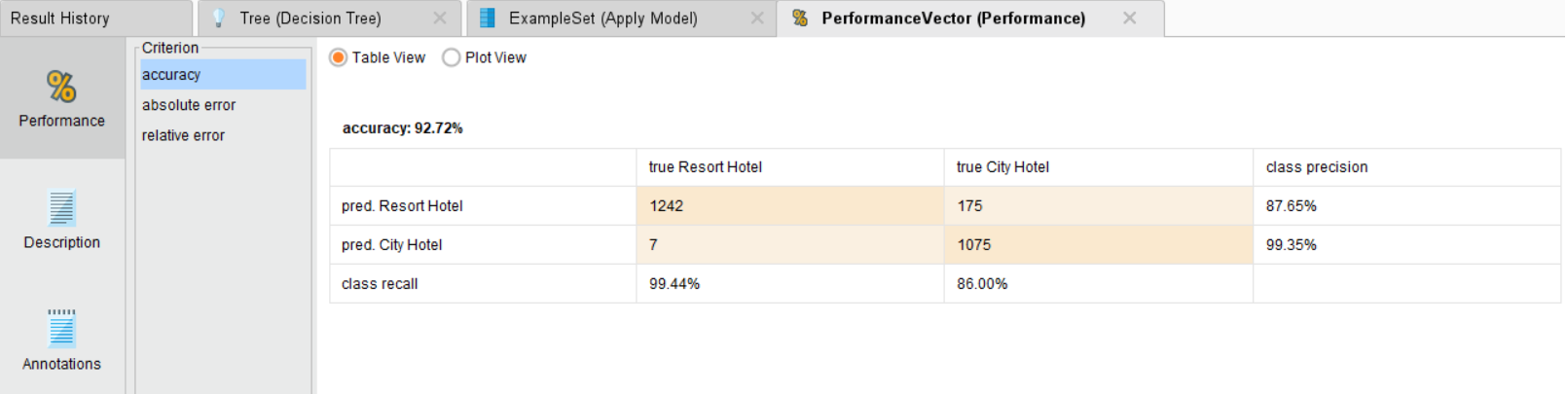
Then we will find below the Decision Tree Design that clarify which Hotels The investors and the Government Should Invest in, and which seasons would be better for each:



There also a description for the decision tree to be more clarified and for the simplicity here:

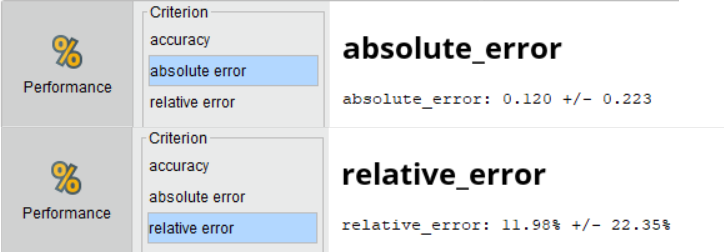


The previous three photos were discussing the process of designing the process of the Decision Tree and the Decision Tree itself. Here we will find the performance measures for the previous processes and a description after:

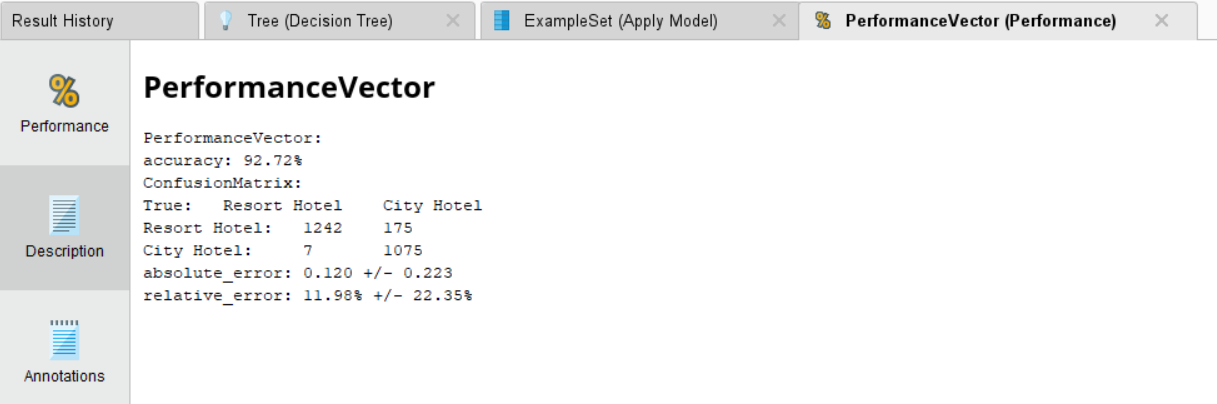


The last screenshot clarifies that this algorithm’s accuracy equal 92.72 %

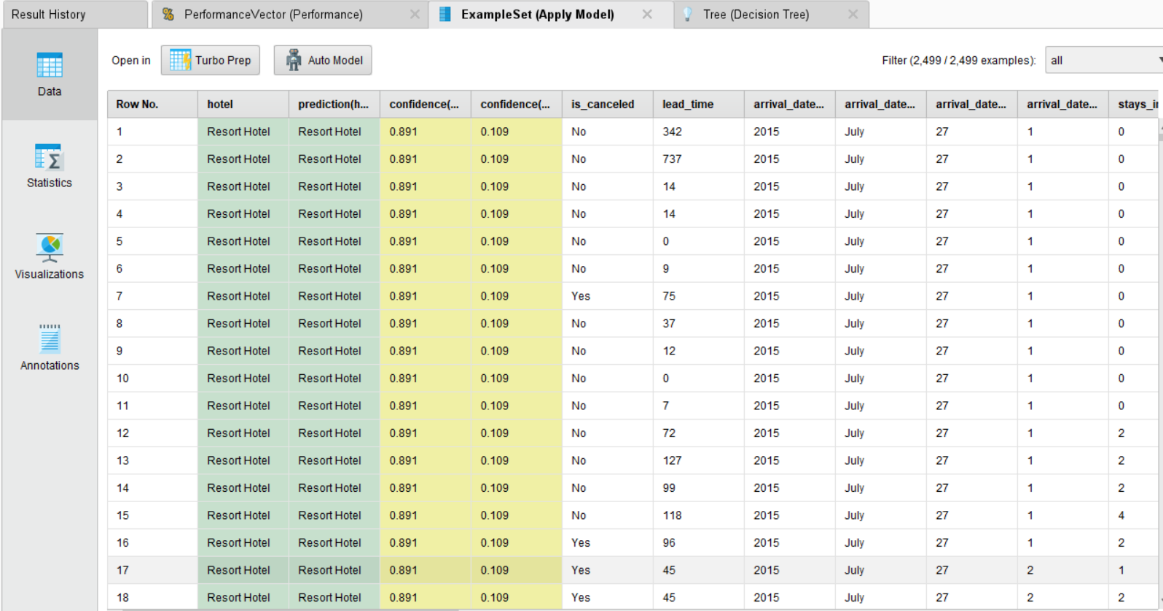
Here we will find the Absolute & the Relative Errors:



Here we will find the Performance Description:



After The performance and its description, we will return back to an operator called Apply Model to see the predicted data and statistics attached:



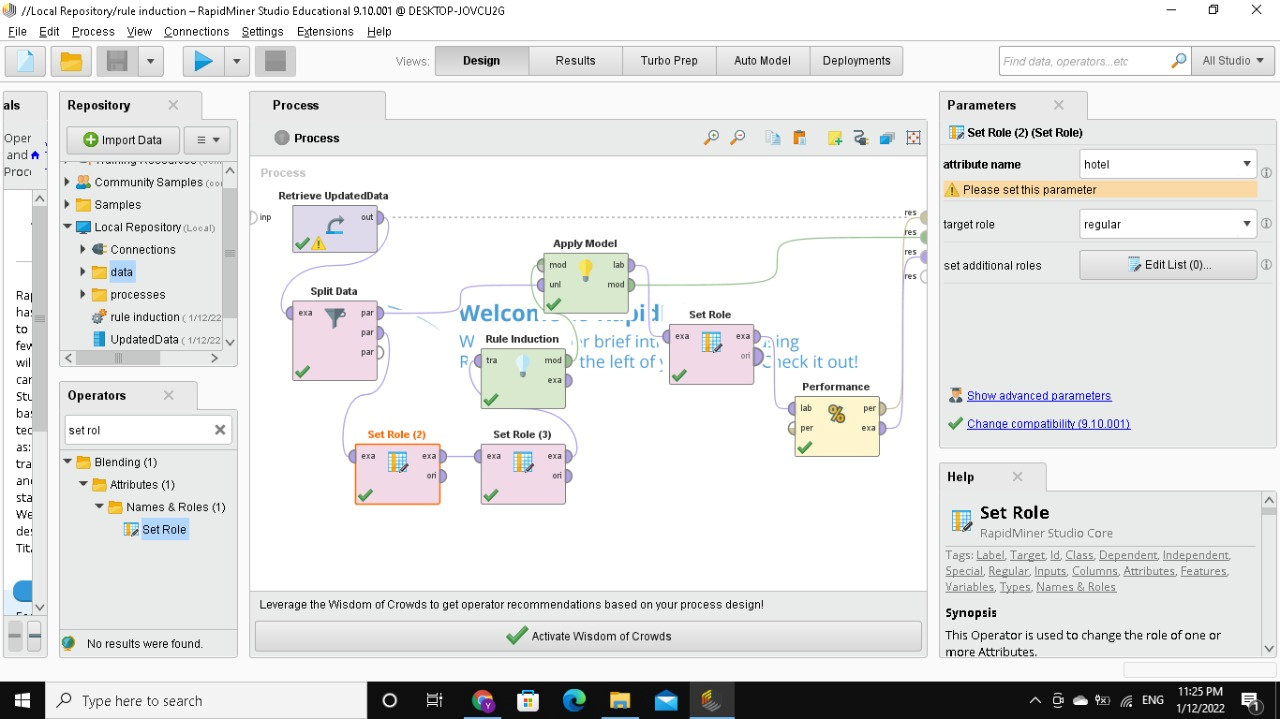
Statistics:



### **Rule Induction Algorithm**

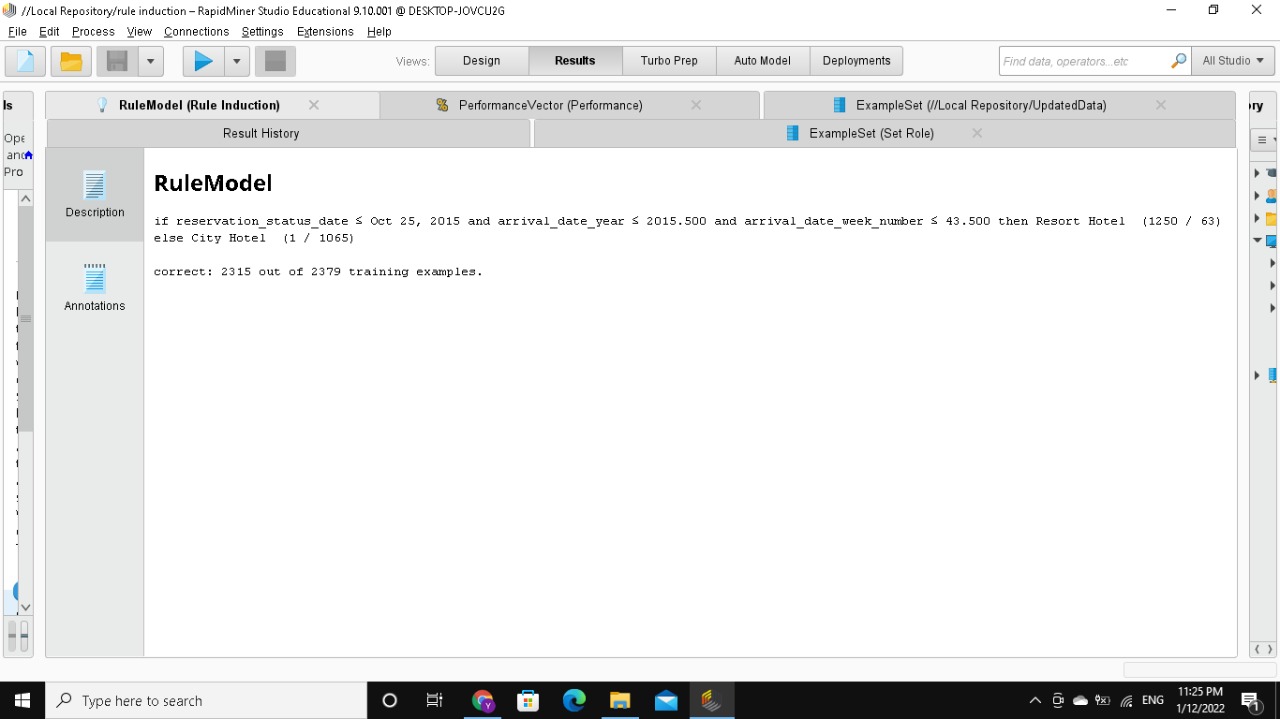
We will attach a screenshot for each step for the Rule Induction algorithm below.

Here we will find the process design:



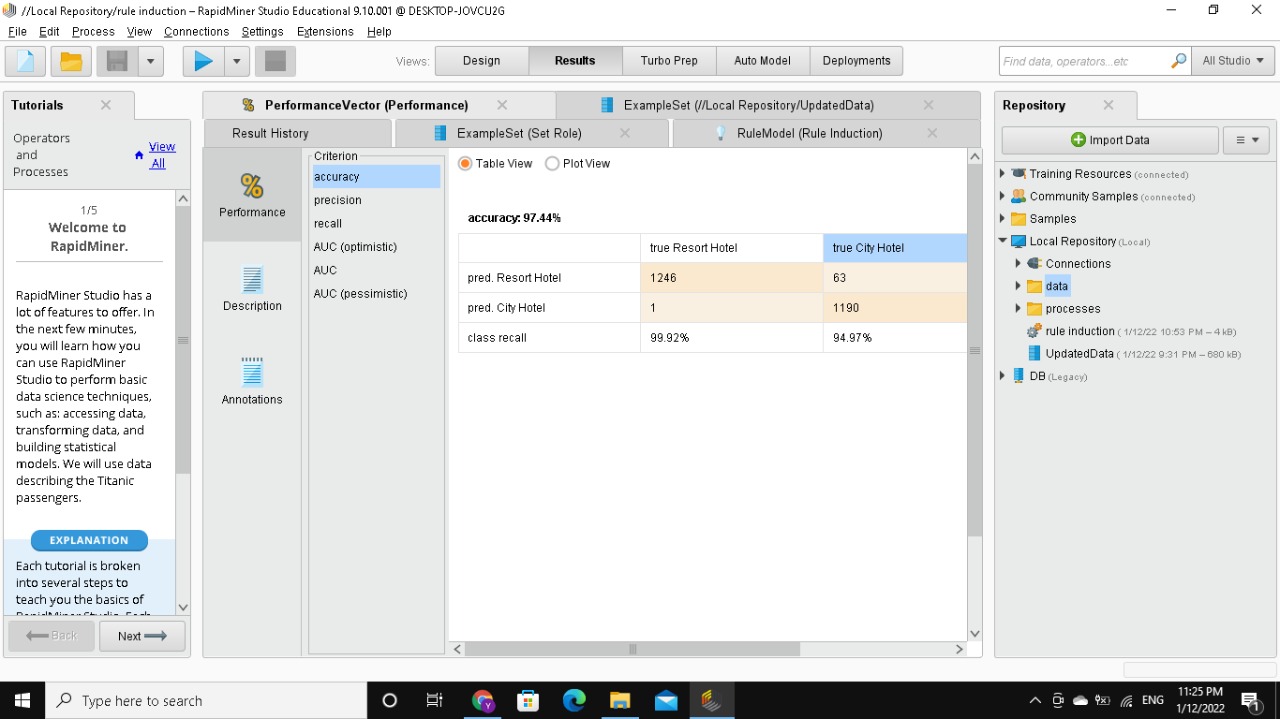
First of all, as the previous algorithm we will split the data into two data sets, training data to the Rule Induction operator before it goes to the Apply Model operator with the testing data from there, we will measure our performance by the Classification Performance operator.

Here we will find the Rule Induction Algorithm’s Description:



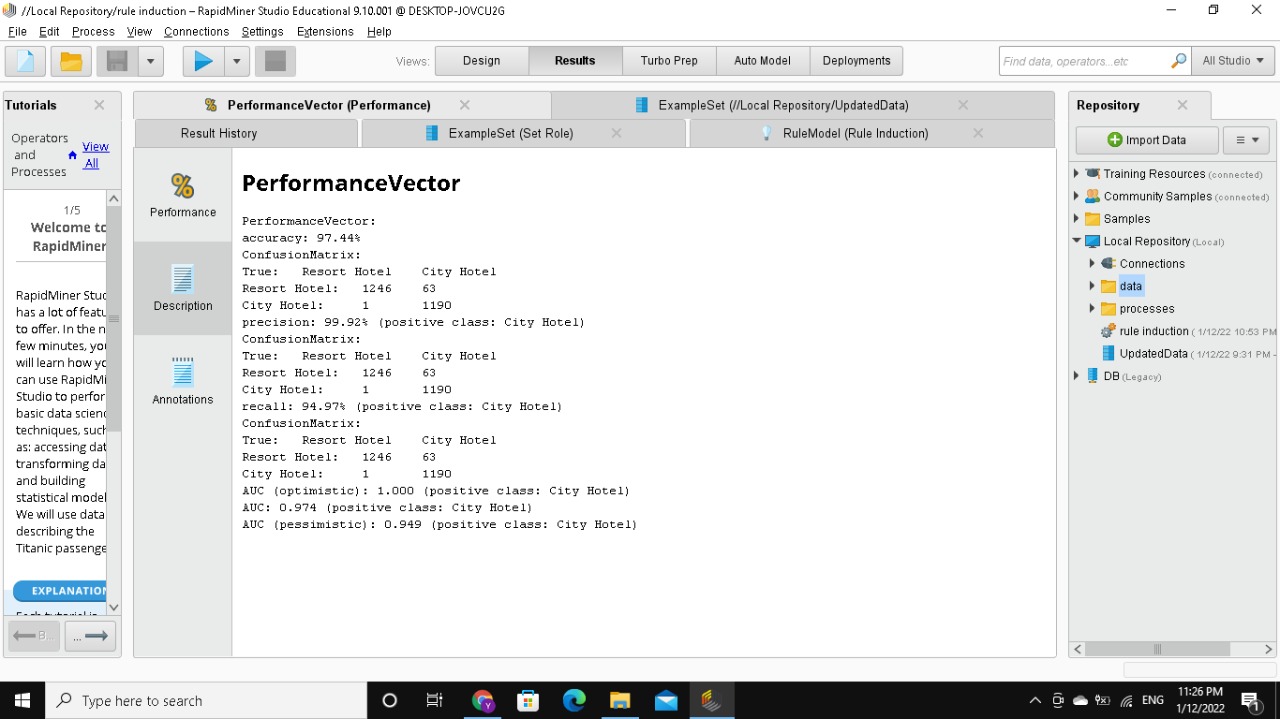
The previous discerption informs that 97.3% of the training example was correctly predicted.

The previous two photos were discussing the process of designing the process of the Rule Induction Algorithm and the discerption of it. Here we will find the performance measures for the previous processes and a description after:

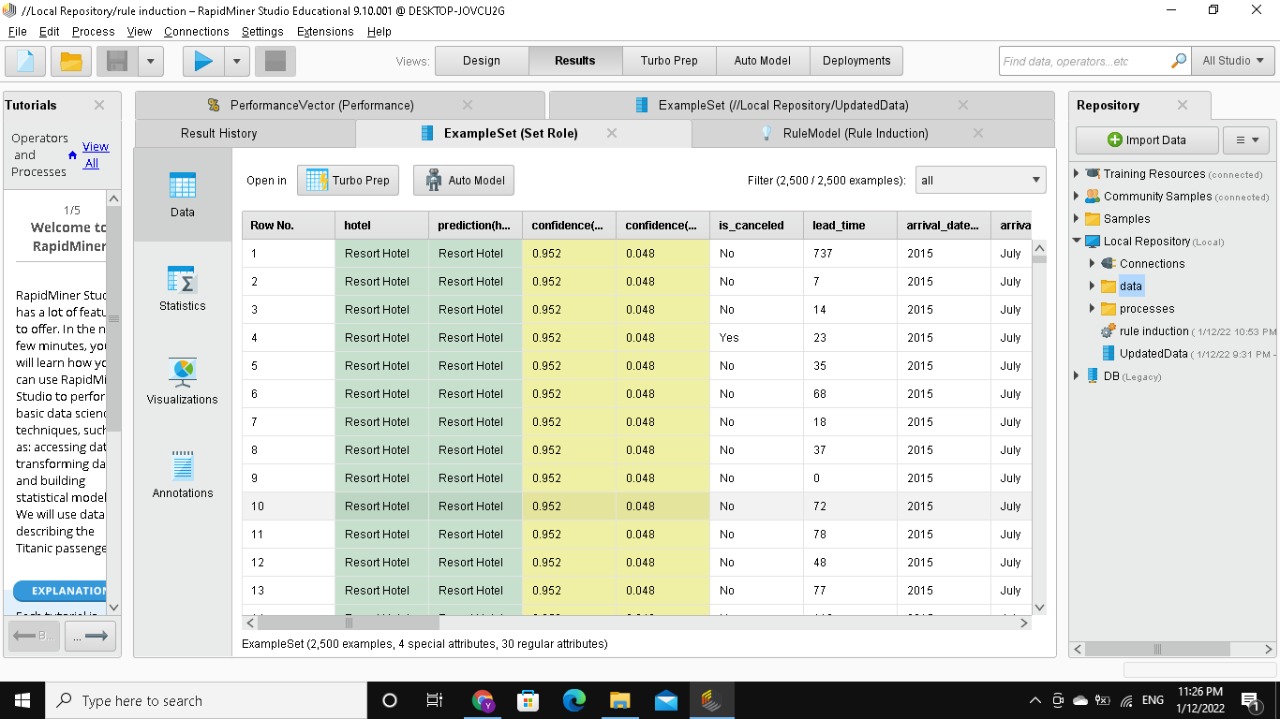


The last screenshot clarifies that this algorithm’s accuracy equal 97.44 %

Also attached here we will find the Performance Description:



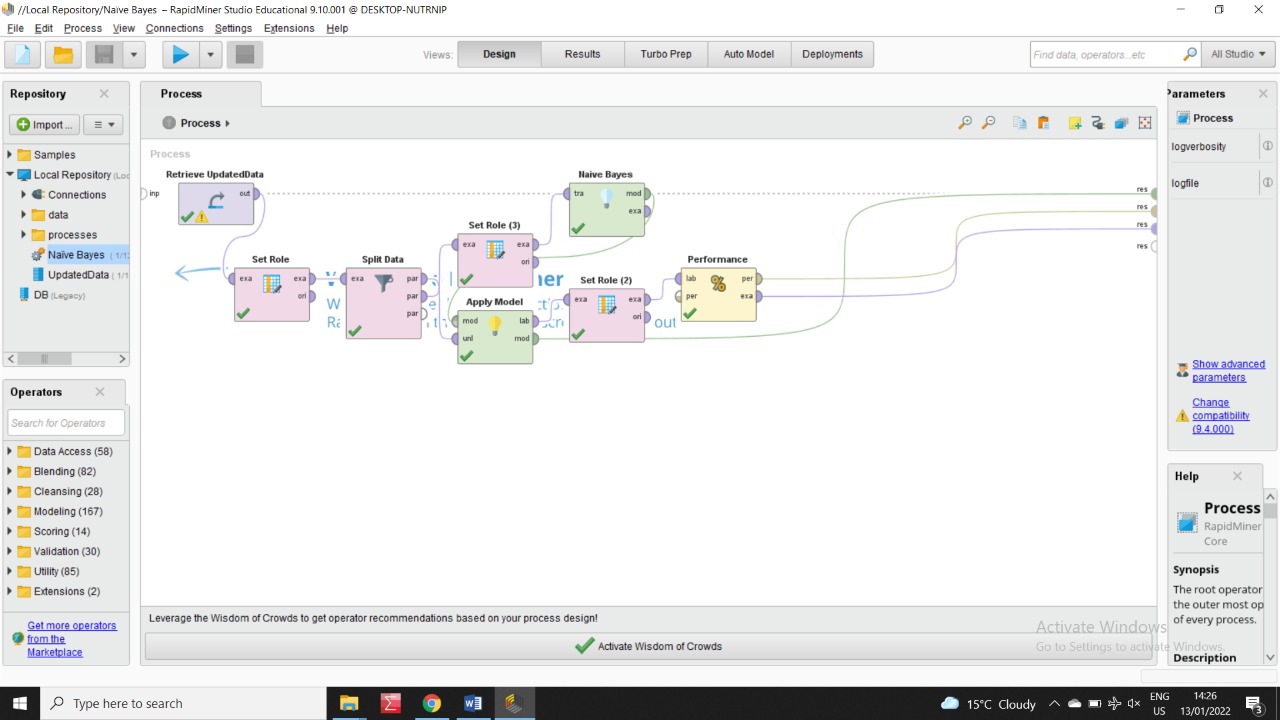
After The performance and its description, we will return back to an operator called Apply Model to see the predicted data:



### **Naïve Bayes Algorithm**:

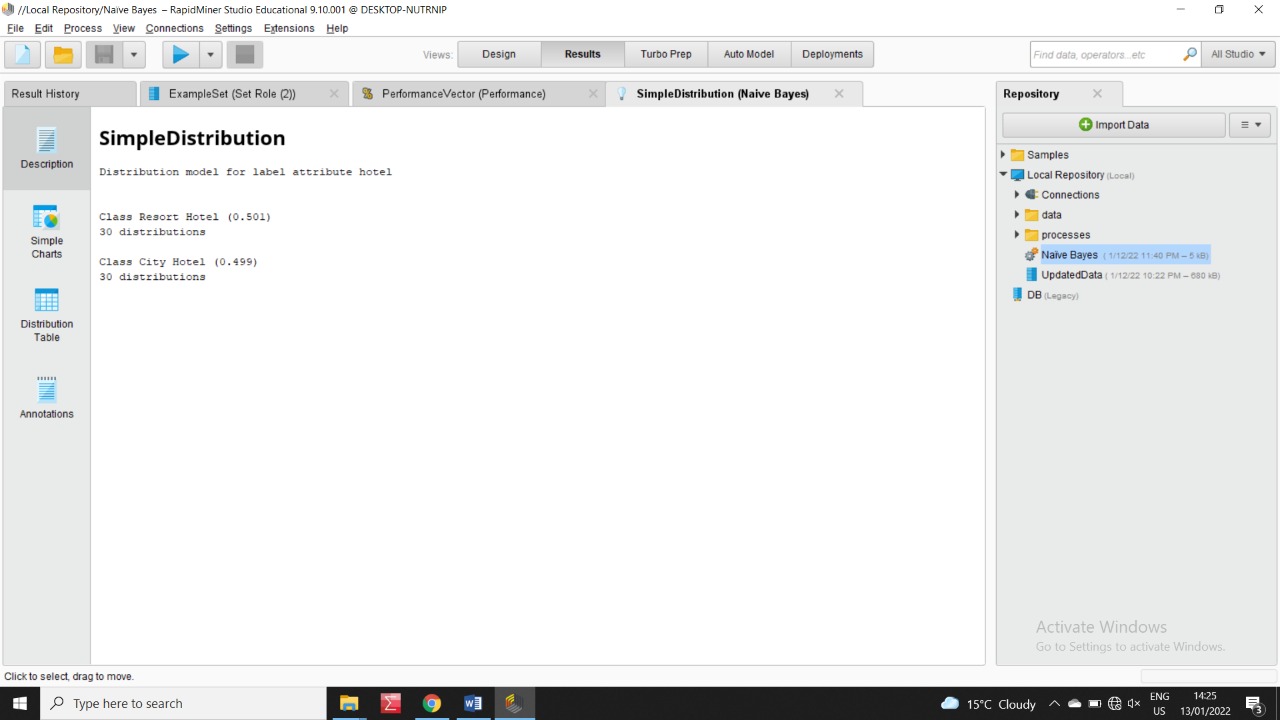
We will attach a screenshot for each step for the Naïve Bayes algorithm below.

Here we will find the process design:

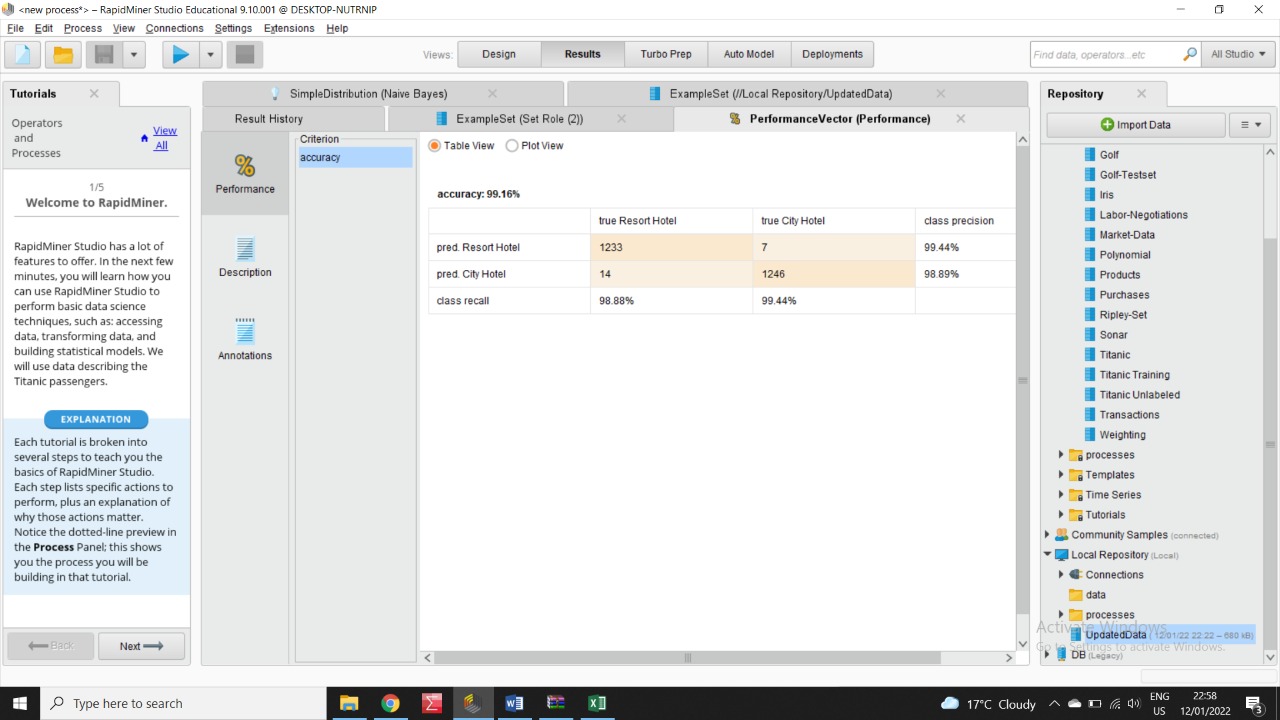


First of all, as the previous two algorithms we will split the data into two data sets, training data to the Naïve Bayes operator before it goes to the Apply Model operator with the testing data from there, we will measure our performance by the Classification Performance operator.

Here we will find the Rule Induction Algorithm’s Description:

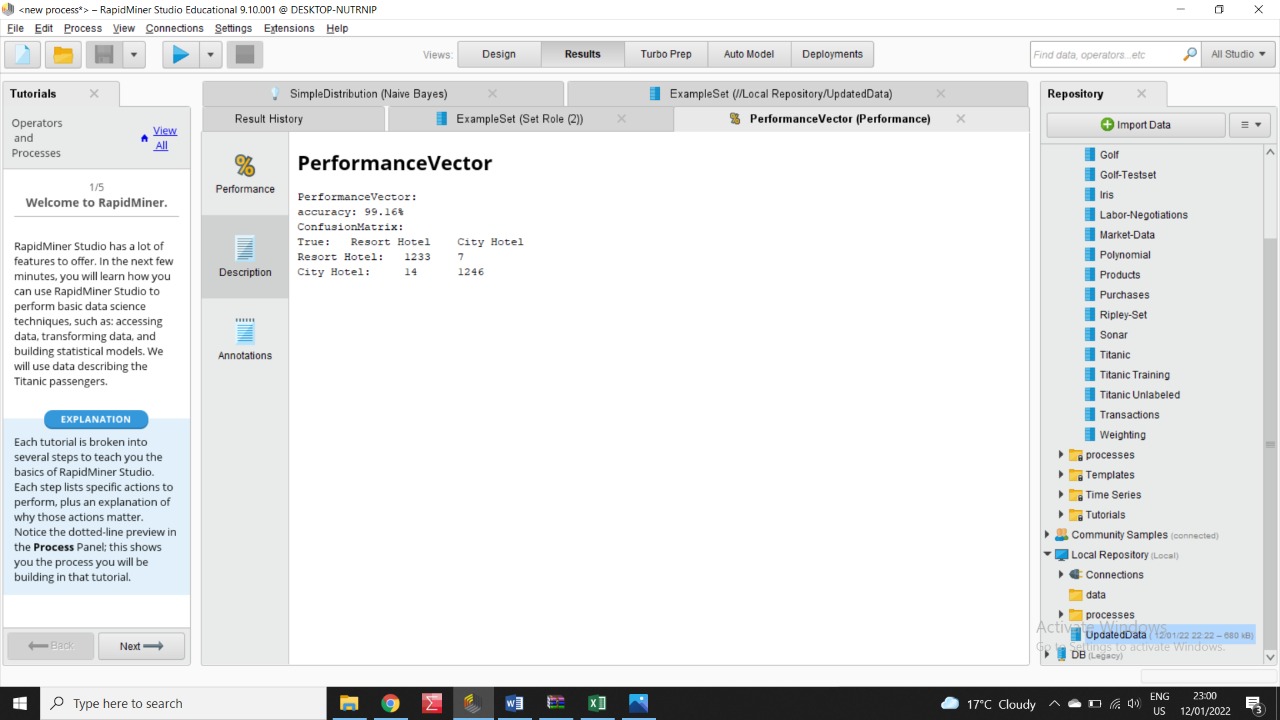


The previous two photos were discussing the process of designing the process of the Naïve Bayes Algorithm and the discerption of it. Here we will find the performance measures for the previous processes and a description after:

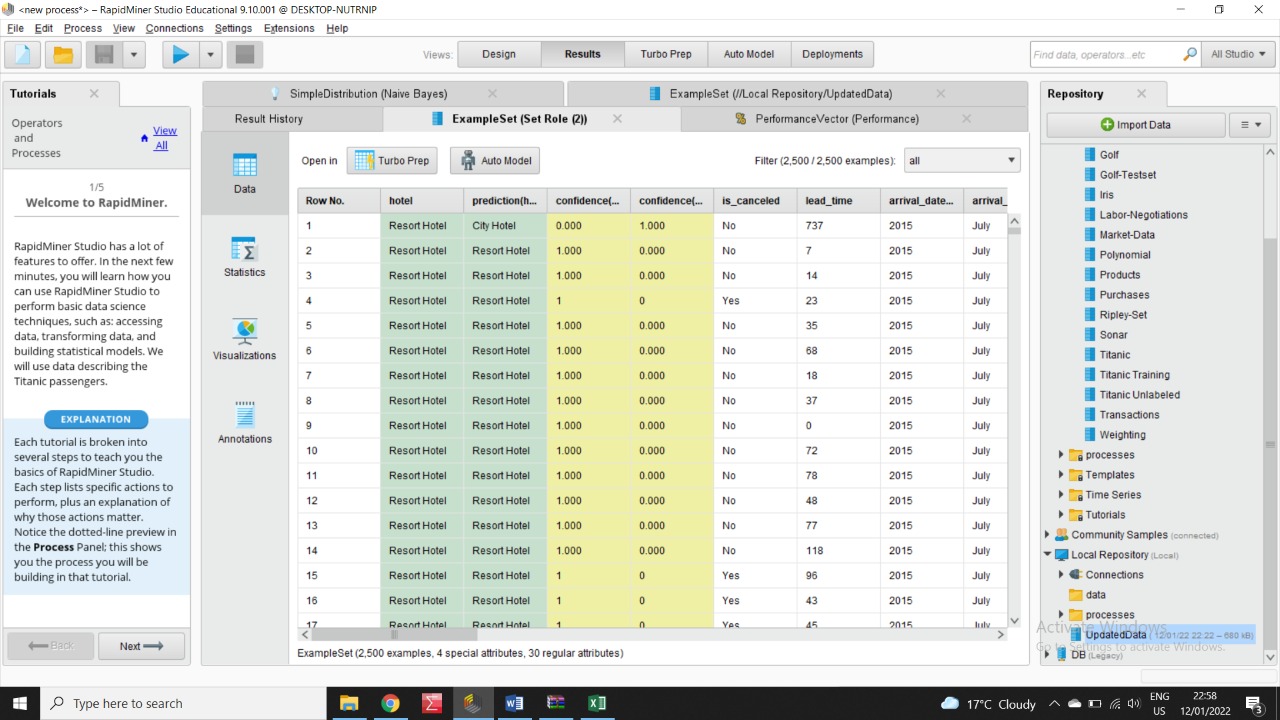


The last screenshot clarifies that this algorithm’s accuracy equal 99.16 %

Also attached here we will find the Performance Description:



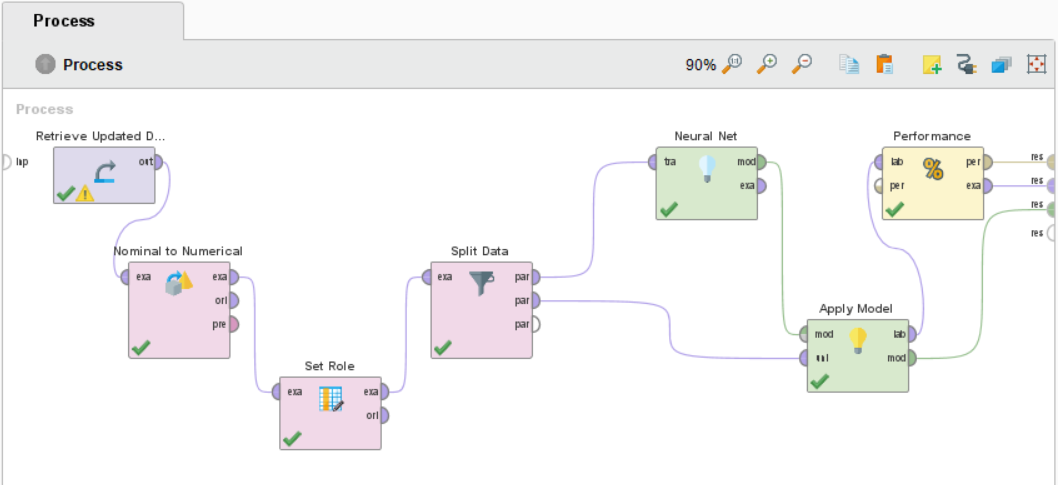
After The performance and its description, we will return back to an operator called Apply Model to see the predicted data:



### **Neural Network Algorithm**:

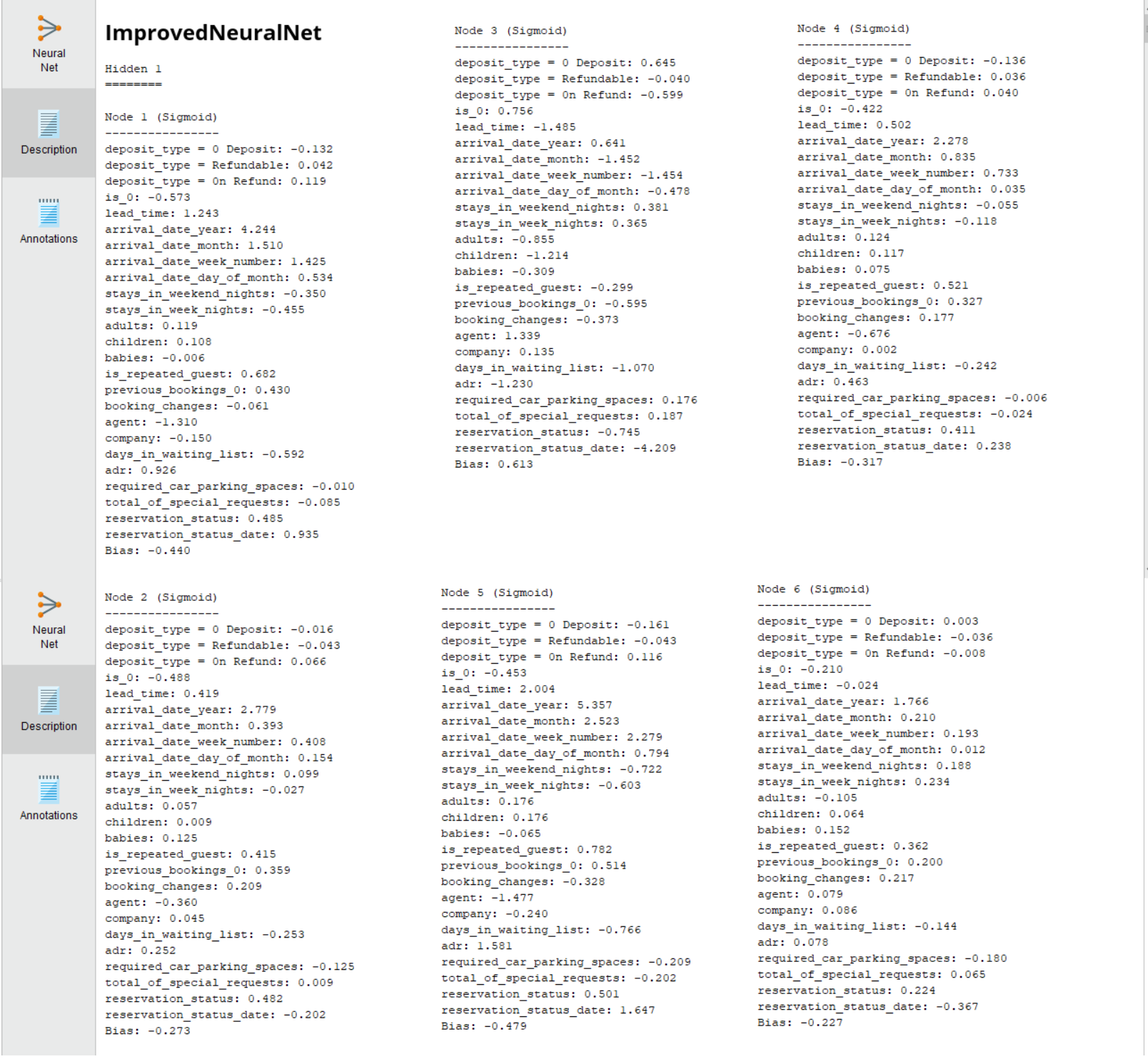
We will attach a screenshot for each step for the Neural Network algorithm below.

Here we will find the process design:

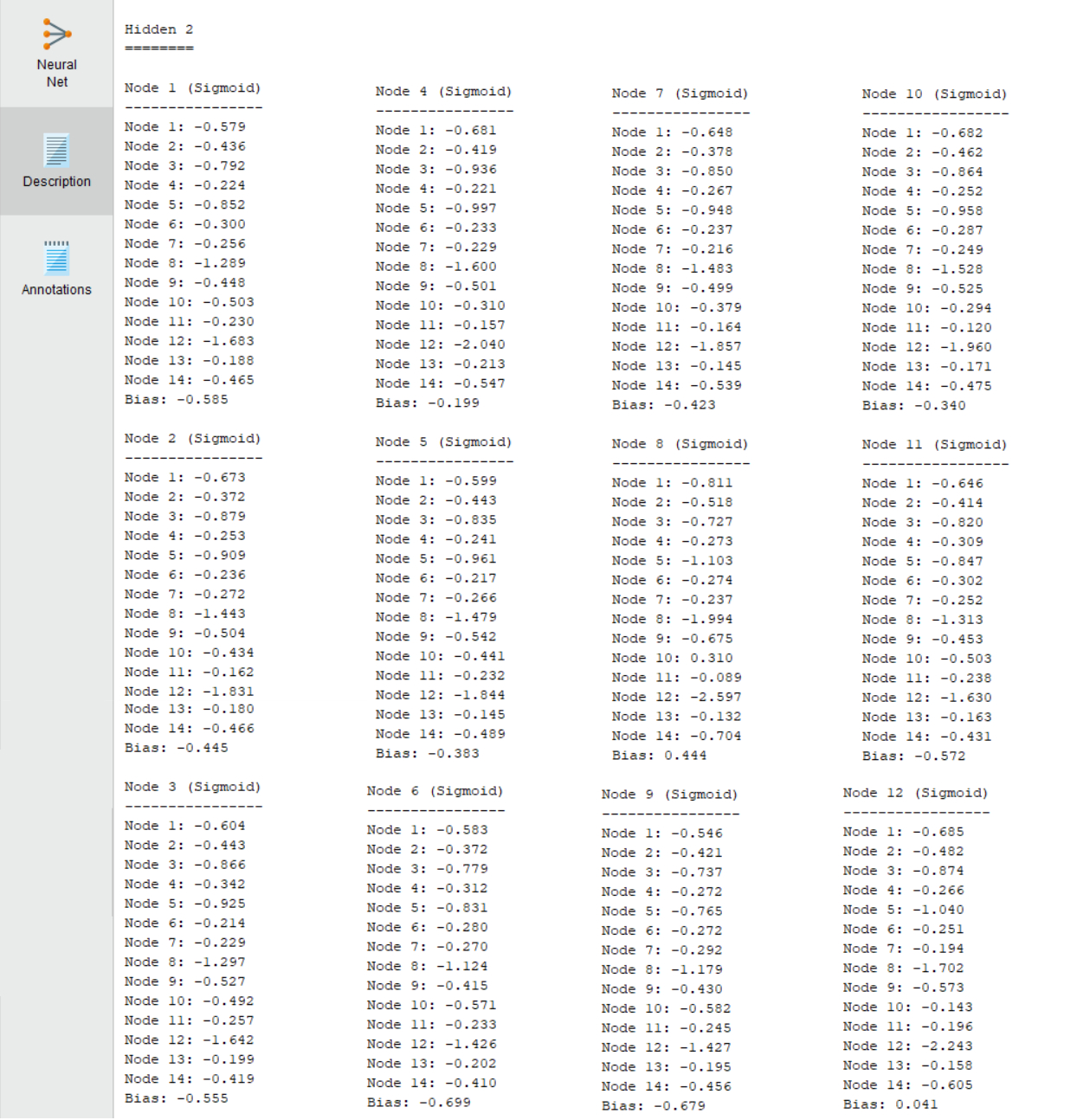


First of all, we will change the Nominal attributes to Numerical one, because The Neural Network algorithm uses only Numerical numbers, after converting we will split the data into two data sets, training data to the Neural Network operator before it goes to the Apply Model operator with the testing data from there, finally we will measure our performance by the Performance operator.

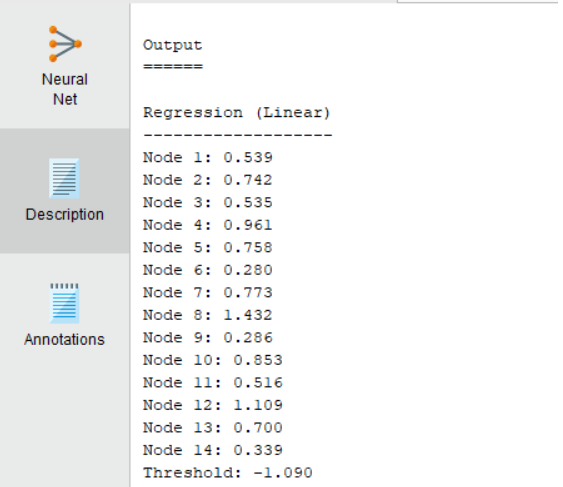
Here we will find part of the Neural Network’s Description:  
The first picture contains 6 Nodes description from the first Hidden Layer.



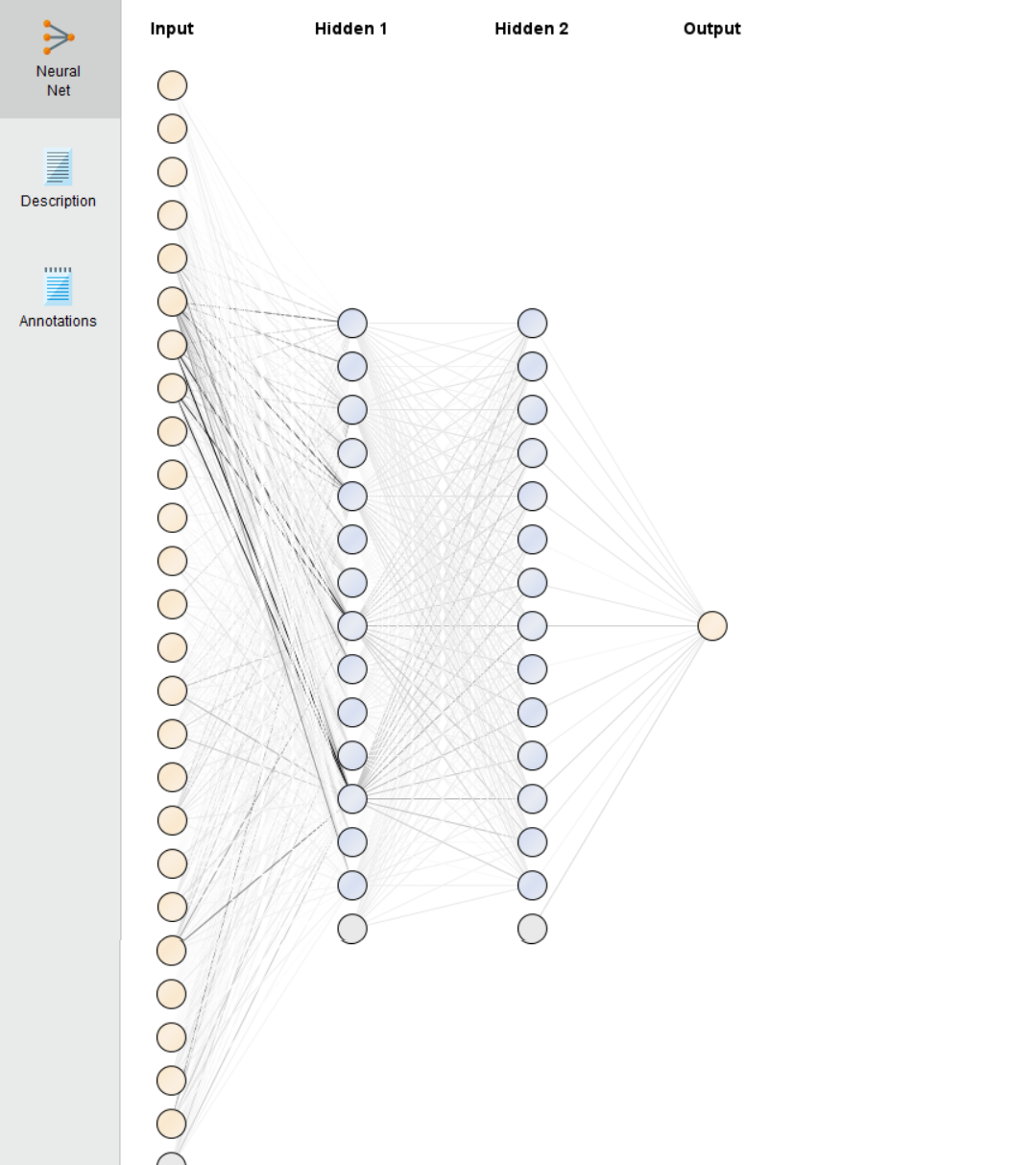
The Second picture contains 12 Nodes description from the Second Hidden Layer.



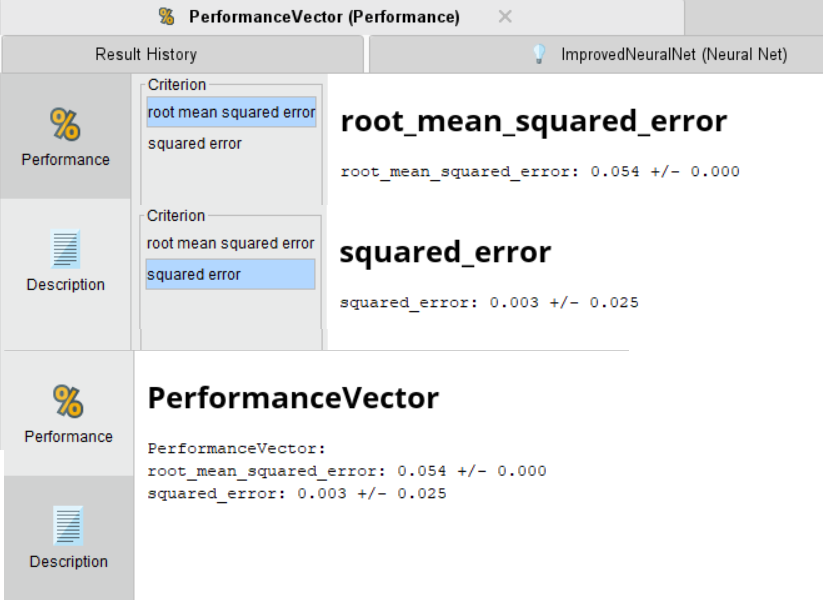
Then the third one contains the final output from the Neural Network algorithm:



The previous photos were discussing the process’s design for Neural Network algorithm, and the description of the process. Now we will find below the Neural Network Design itself:



Here we will find the performance measures for the previous processes and a description after:



## Chapter 4: **Results, Comparson & Conclusion**

