Amul Ghodasara

200493038

Assignment 2

*Part A*

**Read CSV: -** This operator is used to load data from speedsters. Here I load CSV file from my system which has car dataset.

**Select Attribute: -** This operator provides different types of filters to make attribute selection easy. In my work, my attribute filter type is subset, and attributes are Mileage and Price.

**Set Role: -** Main work of this operator is to change the role of attributes. The main role of attributes describes how other operators handle this attribute. Here, my attribute name is Price and target role is label.

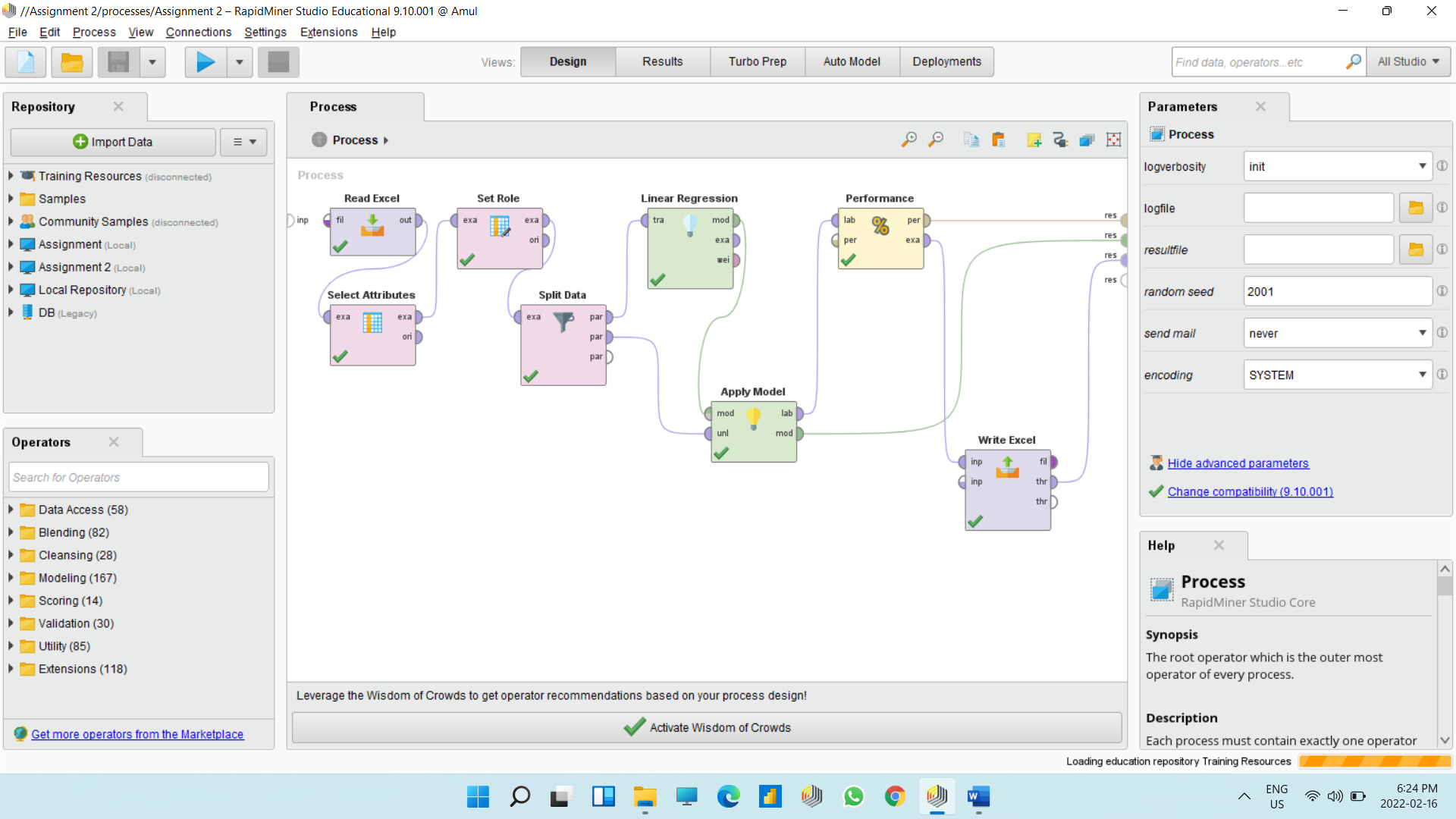
**Split Attribute: -** This operator creates the desired number of subsets of the given example. The example is partitioned into subsets according to the specified sizes. Here, my partitions are 0.7 and 0.3 and sampling type is shuffled sampling.

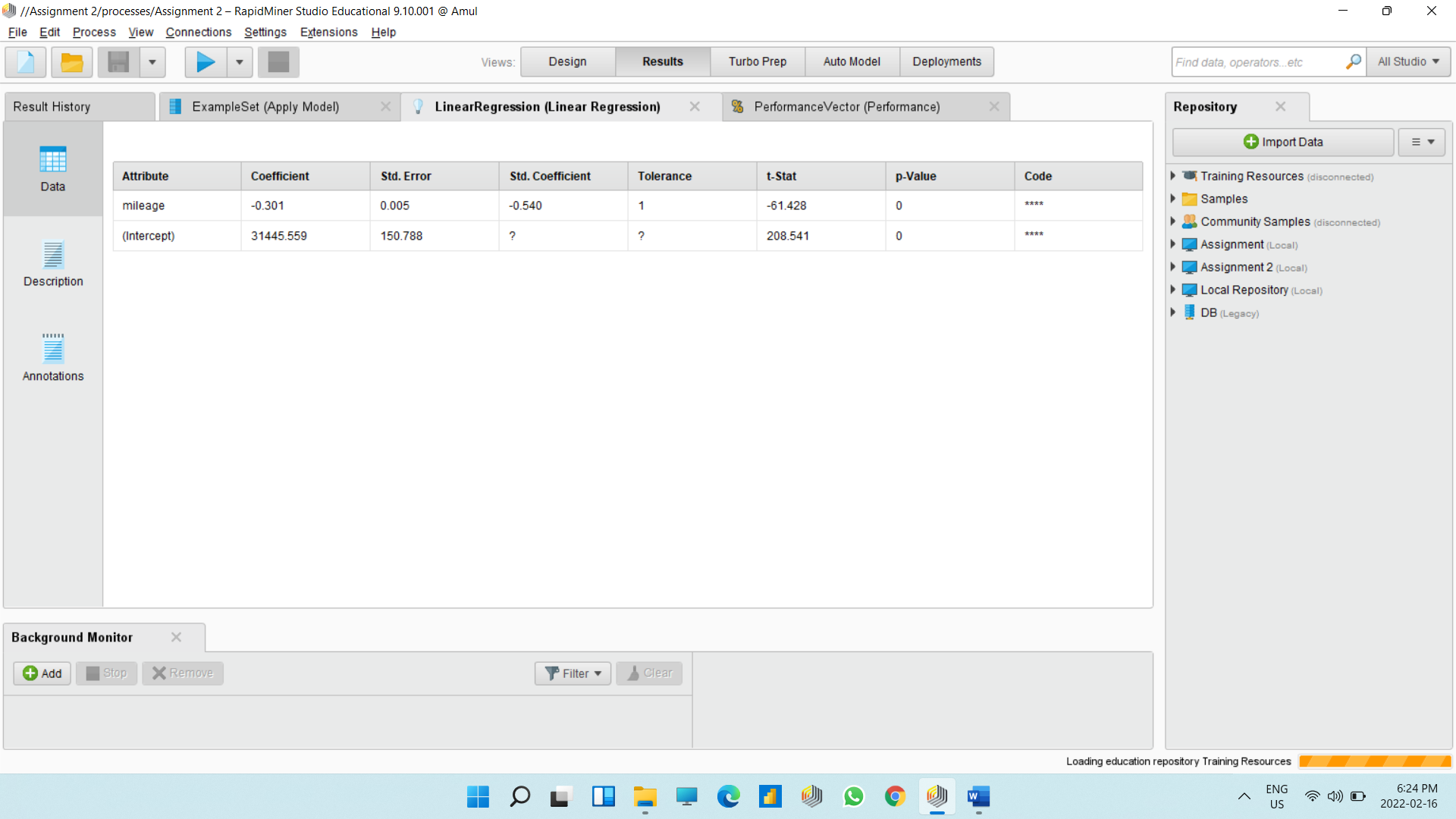
**Linear Regression: -** This operator calculates a linear regression model from the input example. This attribute is used for numerical prediction.

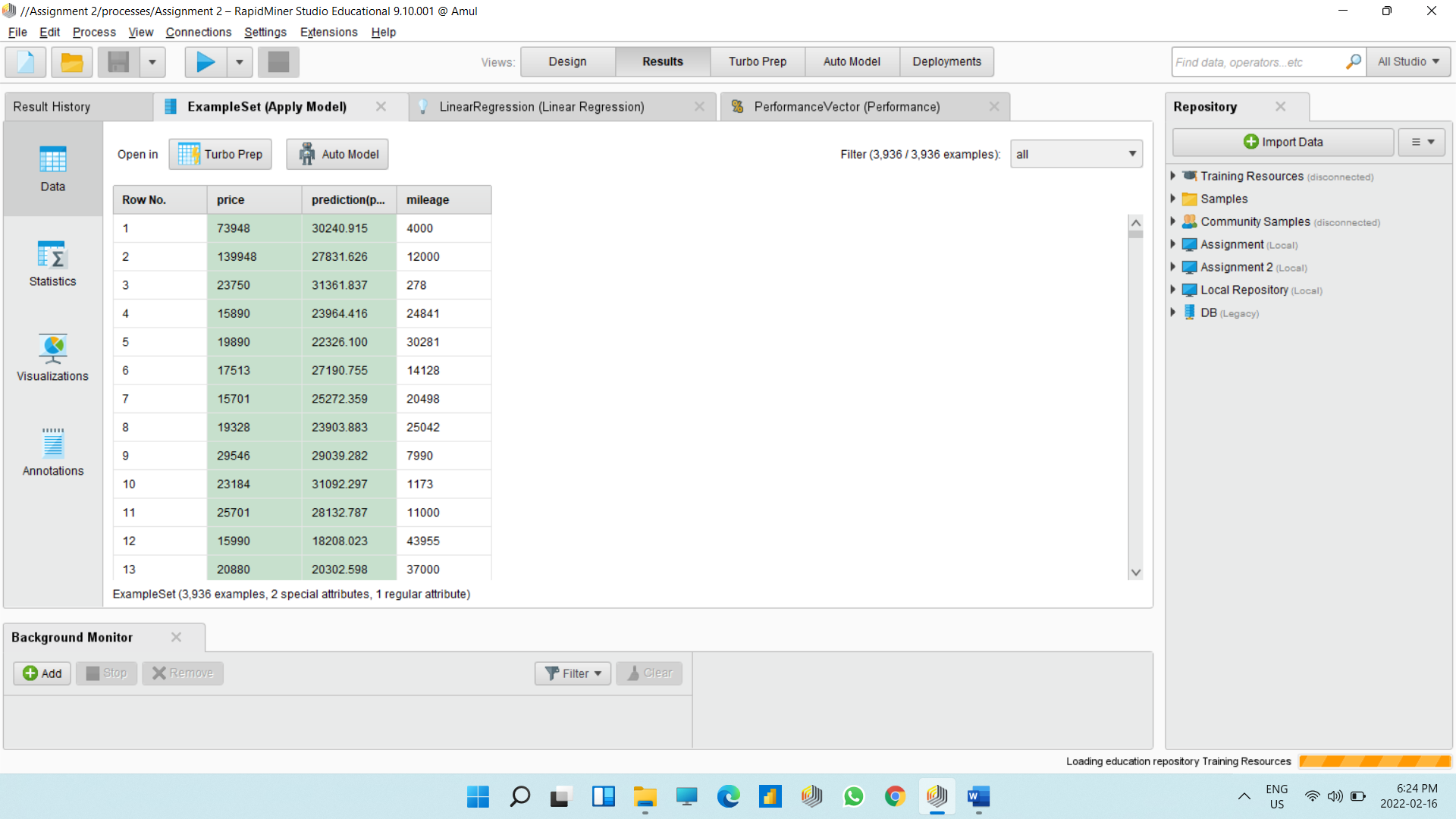
**Apply Model: -** This operator applies a model on an example set. Here, in apply model operator, it et connection from split data and linear regression operator. It passed connection to performance operator and process result.

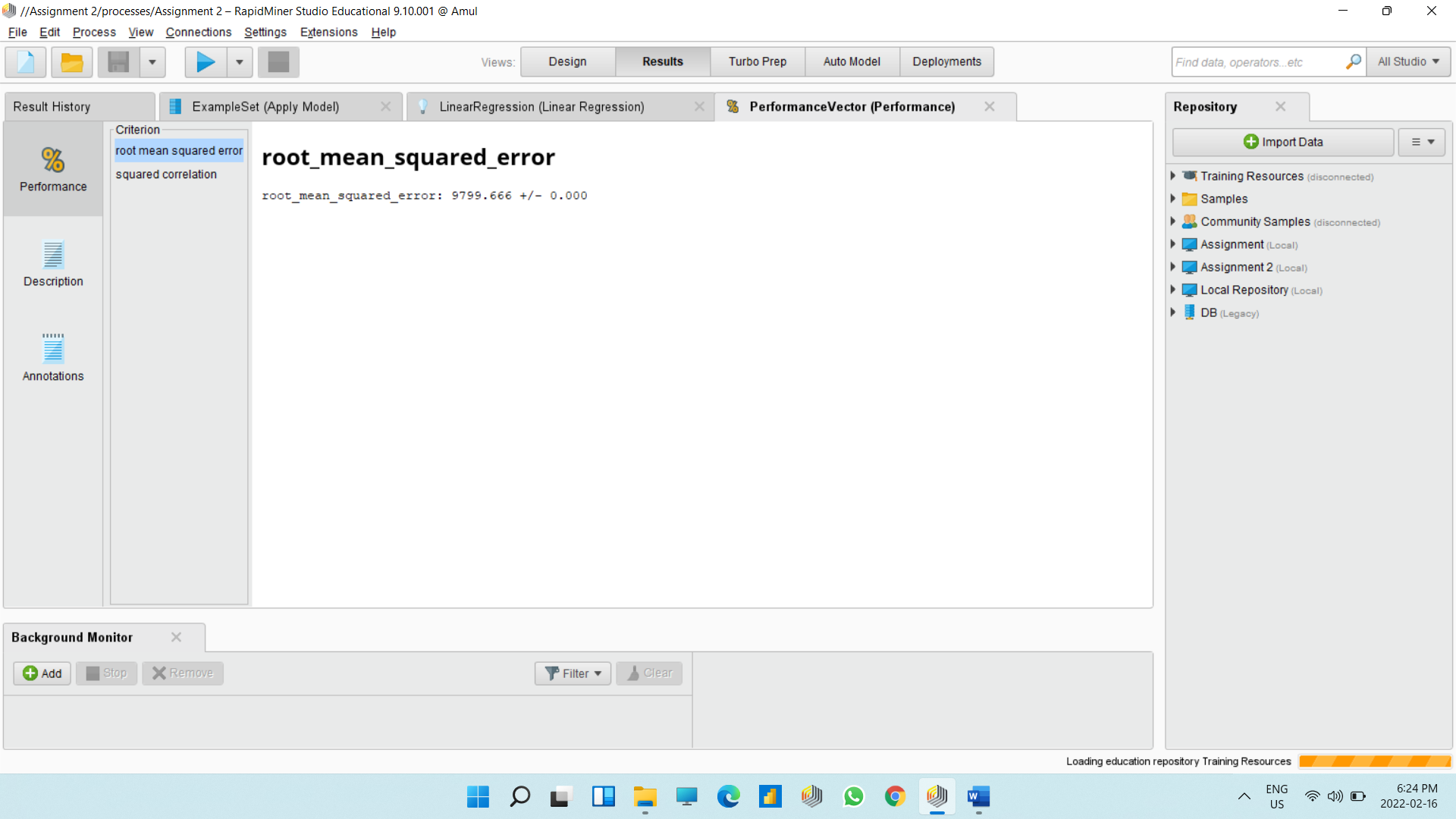
**Performance: -** It is used for statistical performance evaluation of regression tasks. Here, my main criterion is first and selected root mean squared error and squared correlation measures in it.

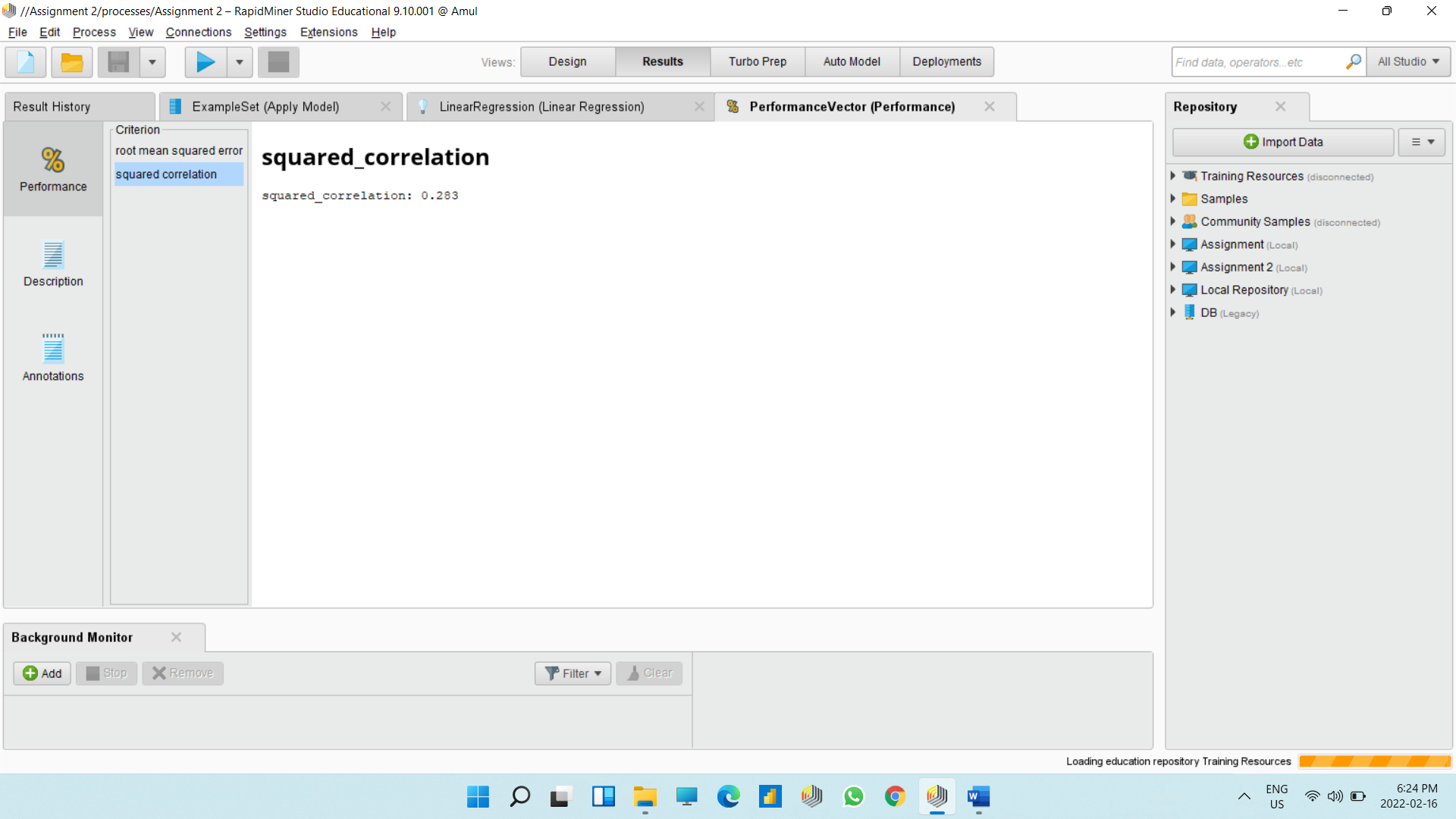
**Write Excel: -** This operator is used for saving that data set into excel spreadsheet. In short, this operator creates excel file that are readable in the user system.











**Explain the linear regression with standard coefficient for your calculation: -**

Here, if we are talking about linear regression with coefficient, I have selected mileage, price as my attribute and set as role to price and target role is label. After completion of all process as output I get standard coefficients for all other attribute in comparison of price attribute.

In output, I get standard coefficient of mileage is -0.540.

*Part B*

**Read CSV: -** This operator is used to load data from speedsters. Here I load CSV file from my system which has car dataset.

**Select Attribute: -** This operator provides different types of filters to make attribute selection easy. In my work, my attribute filter type is subset, and attributes are Mileage, engine size, mpg, price, tax, year. Here we are doing Multivalued regression, so I choose more attributes.

**Set Role: -** Main work of this operator is to change the role of attributes. The main role of attributes describes how other operators handle this attribute. Here, my attribute name is Price and target role is label.

**Split Attribute: -** This operator creates the desired number of subsets of the given example. The example is partitioned into subsets according to the specified sizes. Here, my partitions are 0.7 and 0.3 and sampling type is shuffled sampling.

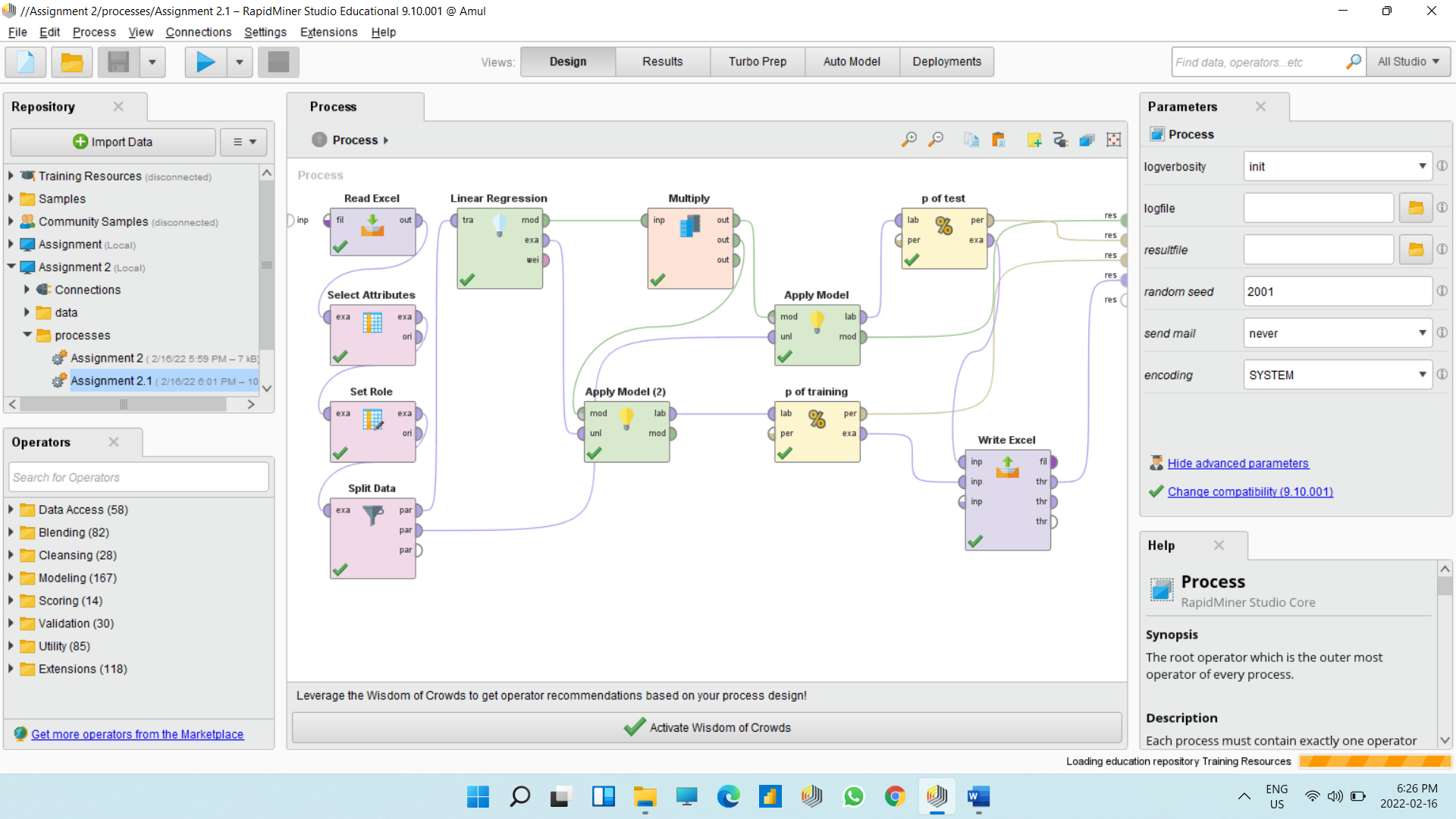
**Linear Regression: -** This operator calculates a linear regression model from the input example. This attribute is used for numerical prediction.

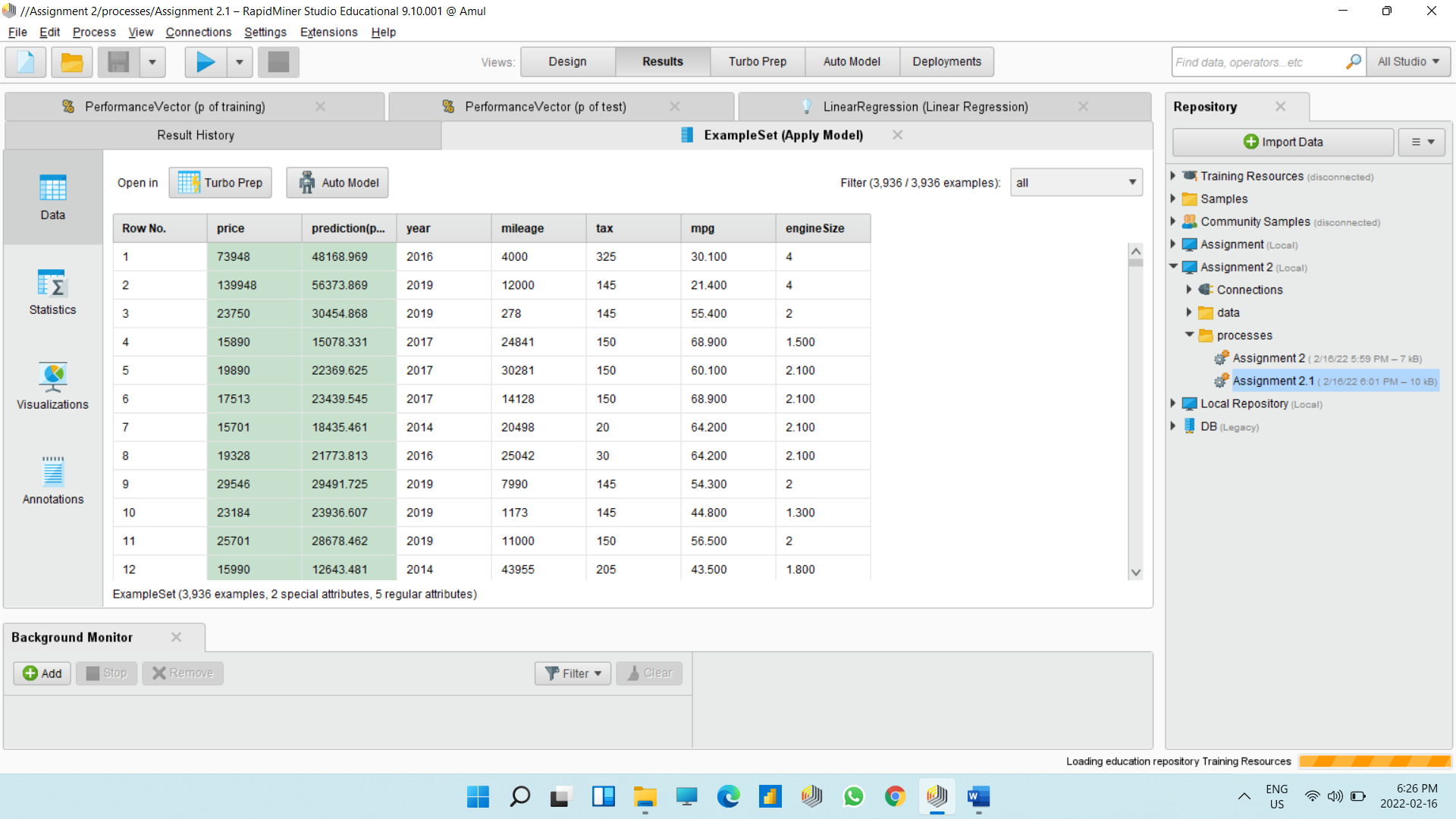
**Multiply: -** Thisoperator creates copies of objects. In short, it takes the RapidMiner object from the input ports and delivers copies of it to output ports. Each connected port creates an independent copy.

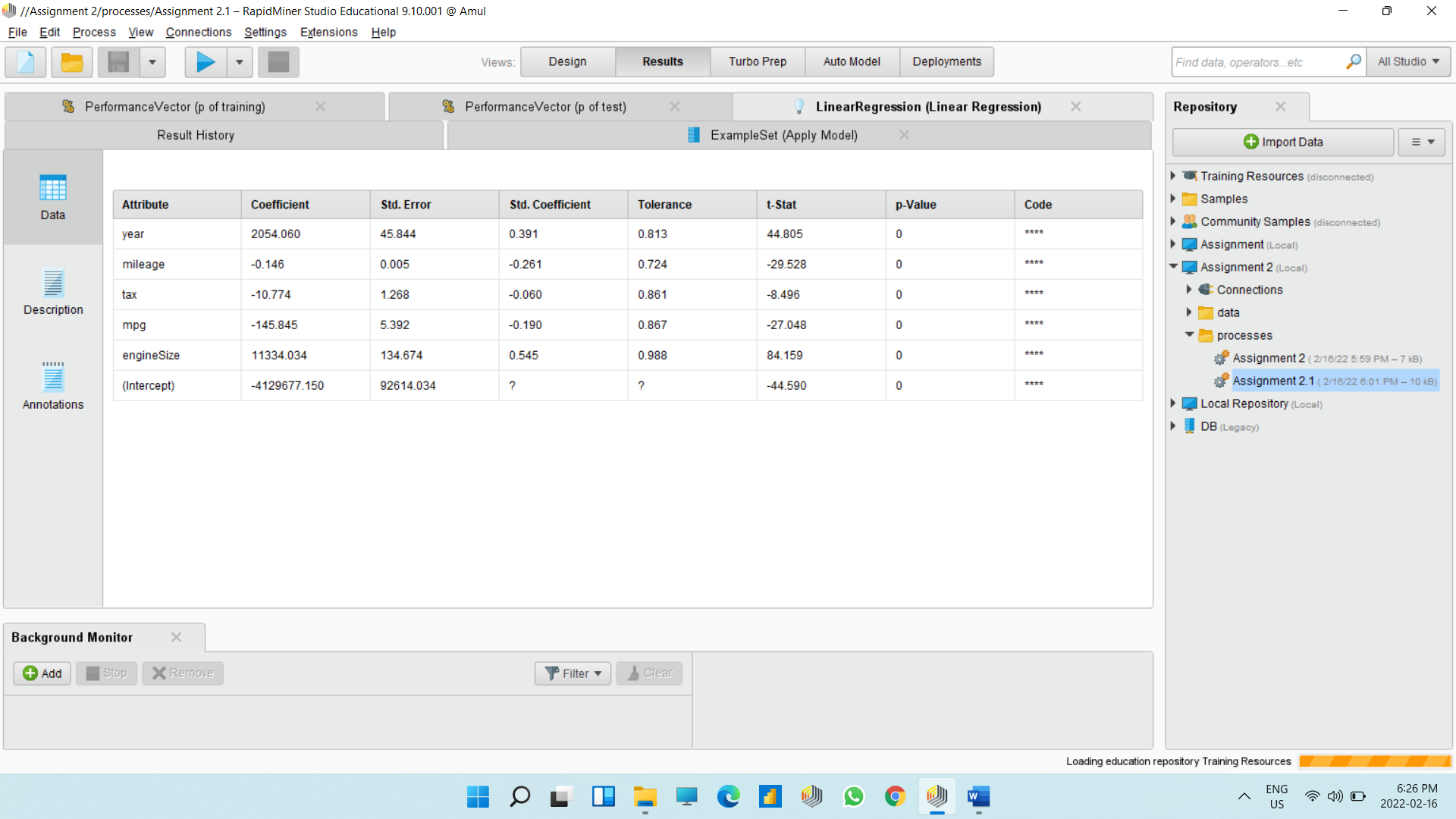
**Apply Model: -** This operator applies a model on an example set. For this process, I take 2 apply model because here we are performing Multivalued regression and we have multiply operator as well. Apply model 1 get connection from multiply and split data operators. Apply model 2 get connection from linear regression and multiply operators.

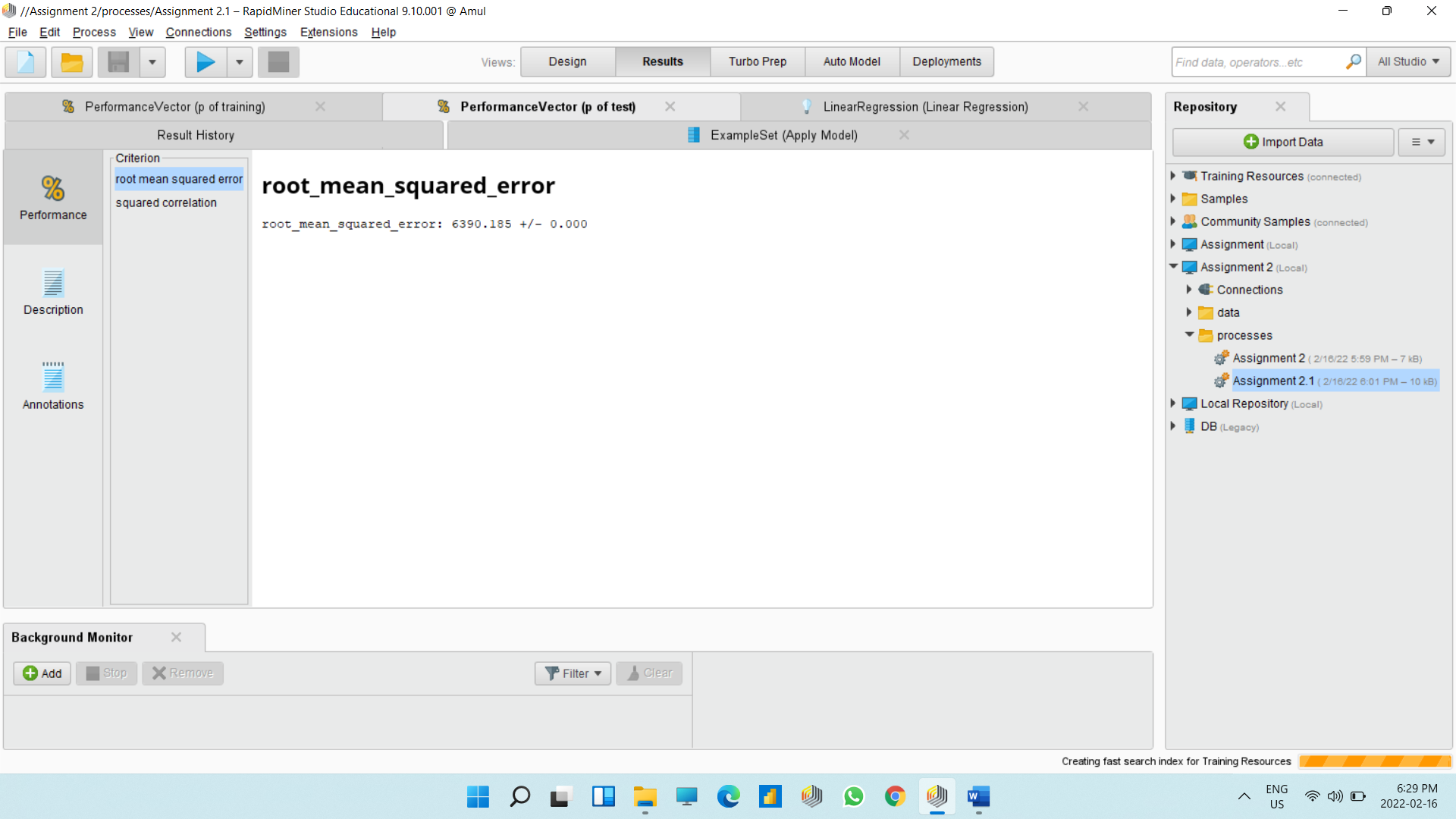
**Performance: -** It is used for statistical performance evaluation of regression tasks. For this process, I take 2 performance operators because here we are performing Multivalued regression and we have multiply operator as well Here, my main criterion is first and selected root mean squared error and squared correlation measures in it.

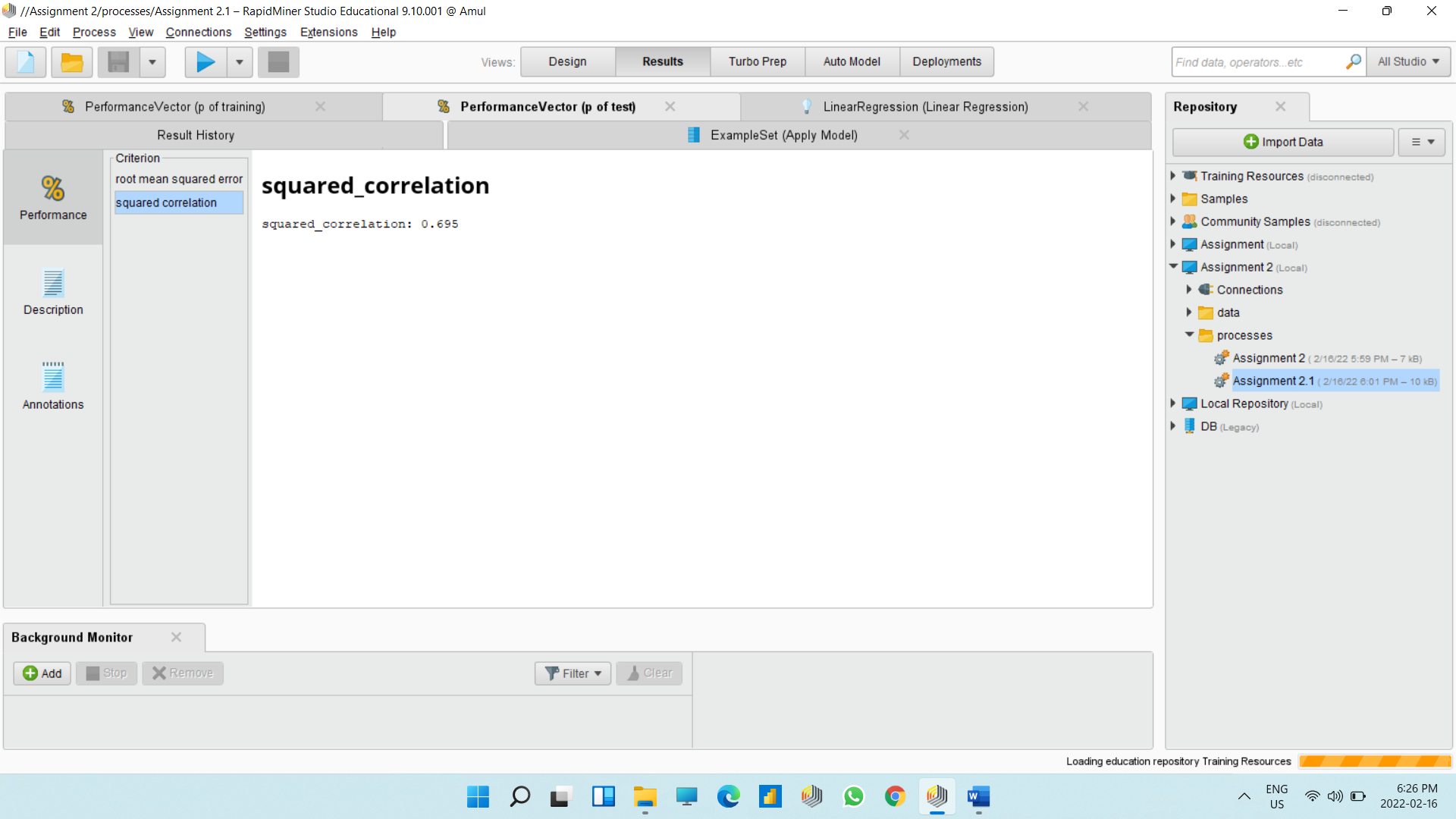
**Write Excel: -** This operator is used for saving that data set into excel spreadsheet. In short, this operator creates excel file that are readable in the user system. Here, I have got 2 connections. One from p of test and another is from p of training. And pass one output from it.

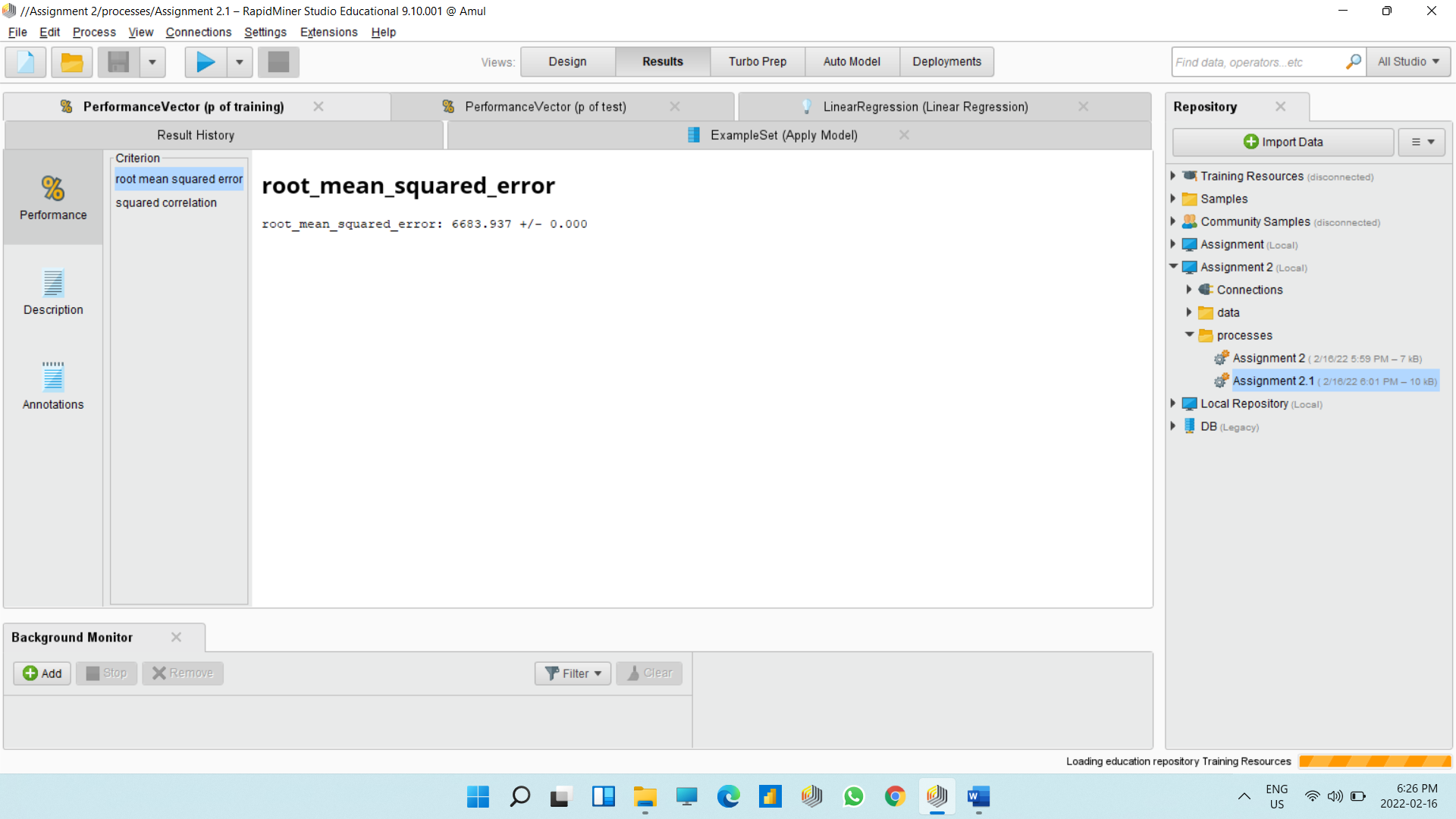


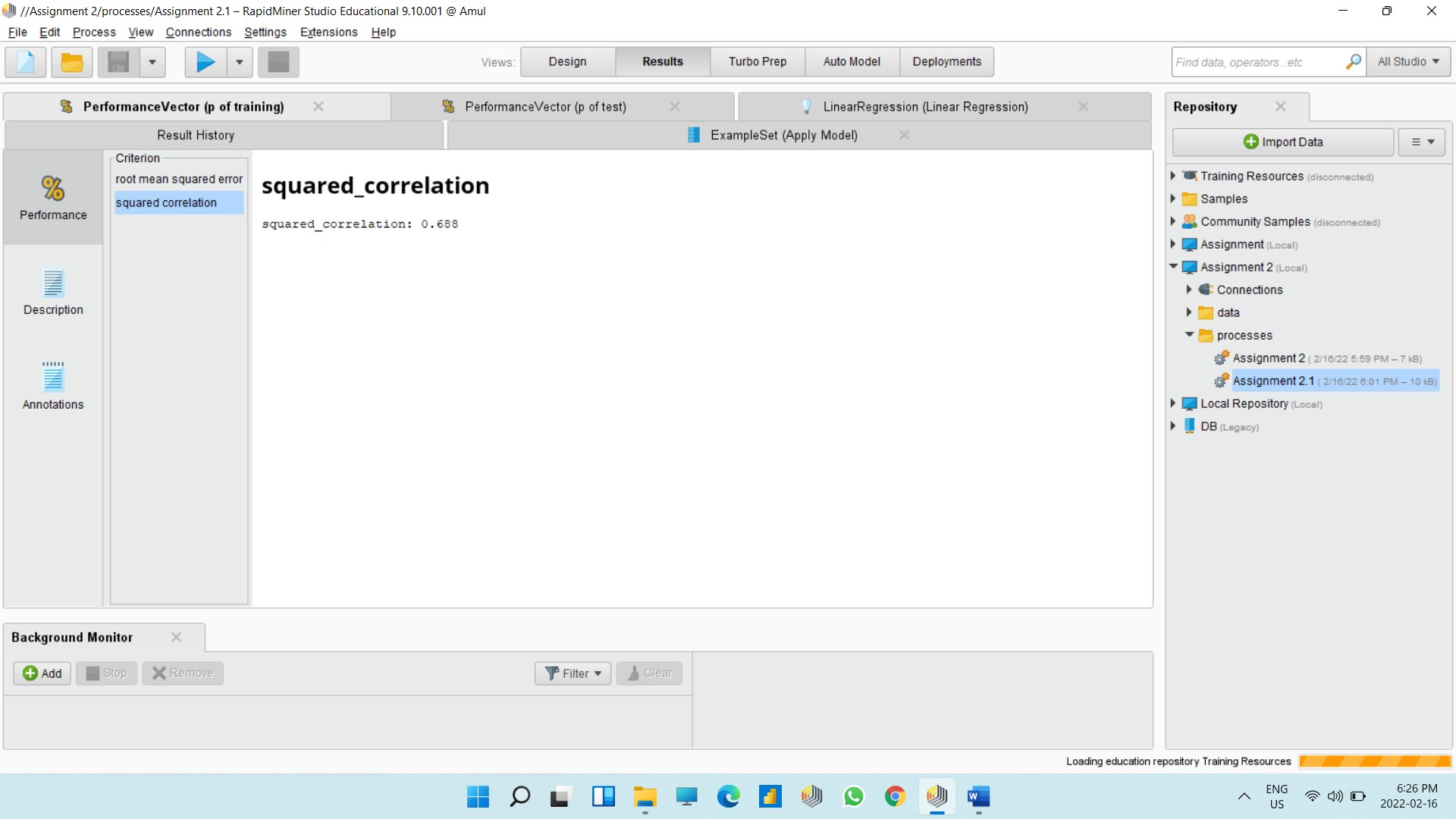












**Explain the linear regression with standard coefficient for your calculation: -**

Here, if we are talking about linear regression with coefficient, I have selected engine size, mileage, price, year, Tax, mpg as my attribute and set as role to price and target role is label. After completion of all process as output I get standard coefficients for all other attribute in comparison of price attribute.

In output, I get standard coefficient of year is 0.391, mileage -0.261, tax -0.060, mpg -0.190 and for mpg is 0.545.