**EXPERIMENT 5**

**Aim:** Experiment to explore Rapid Miner and implement classification models like Decision Tree and Naive Bayes etc.

**To do:**

1. Preprocess data. Split data into train and test set
2. Build Classification model using Rapid miner on training data (use decision tree and naive Bayes method)
3. Calculate metrics based on test data
4. Compare the models based on metrics and find which model is best suited for the dataset.

**Dataset:**

<https://www.kaggle.com/datasets/zaurbegiev/my-dataset?select=credit_test.csv>

This dataset of different columns: Loan ID, Customer ID, Current Loan Amount, Term, Credit, Score, Annual Income,Years in current job, Home Ownership, Purpose and Monthly Debt. This dataset contains information about loans and customers who have taken out those loans.

**Theory:**

**RapidMiner:**

RapidMiner is a data science platform that provides a comprehensive set of tools for data preparation, machine learning, and predictive analytics. It is designed to help users build predictive models quickly and easily without needing advanced programming skills.

The RapidMiner platform includes a drag-and-drop interface that allows users to easily build data pipelines for data preprocessing, transformation, and modeling. The platform supports a wide range of data sources, including CSV, Excel, and databases like MySQL and PostgreSQL.

RapidMiner offers a wide variety of machine learning algorithms, including regression, classification, clustering, and association analysis. It also includes advanced analytics features like text analytics, image analytics, and time series analysis.

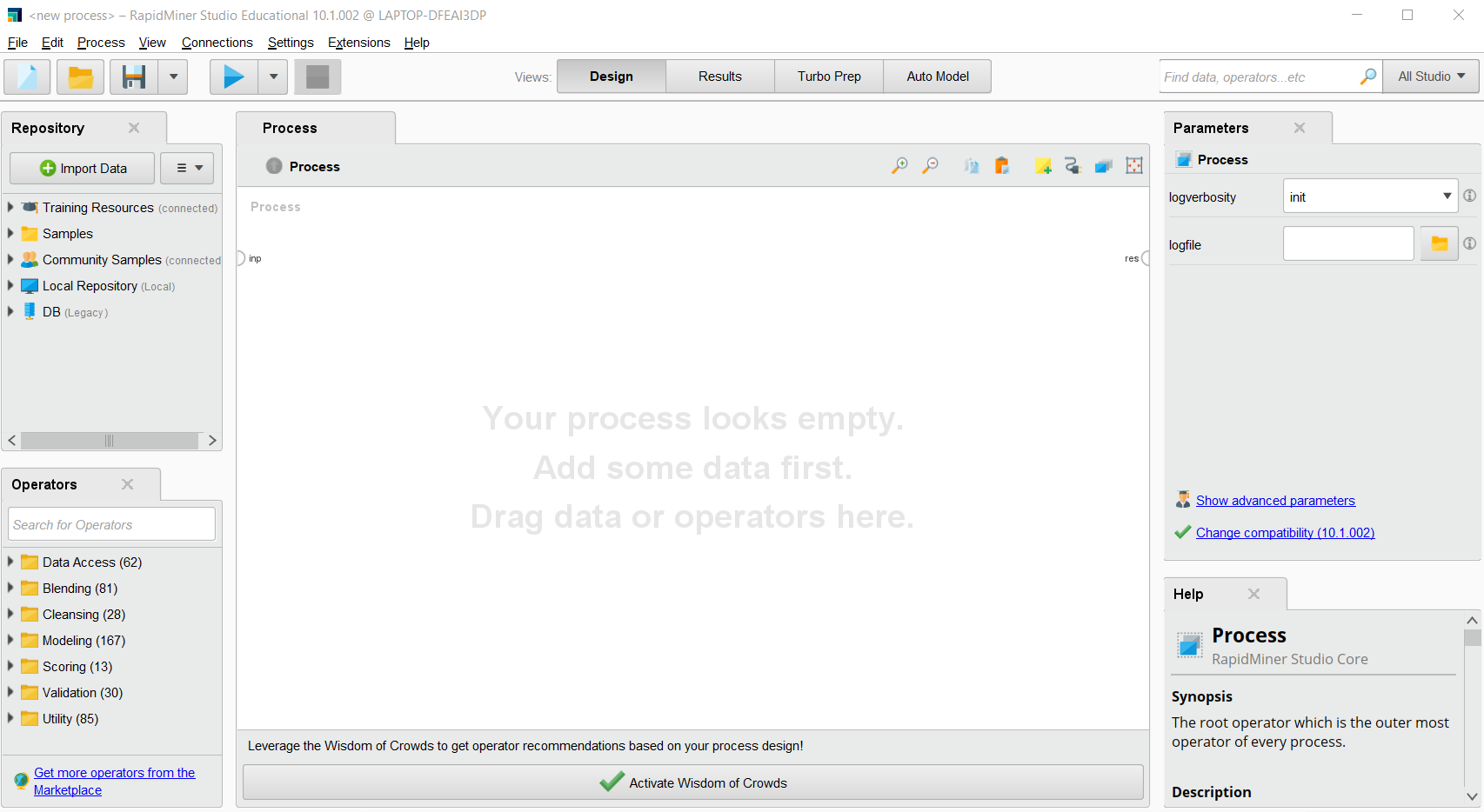
RapidMiner is available as both a desktop application and a cloud-based platform, and it is used by businesses, researchers, and educators worldwide to analyze and make predictions on various data types.

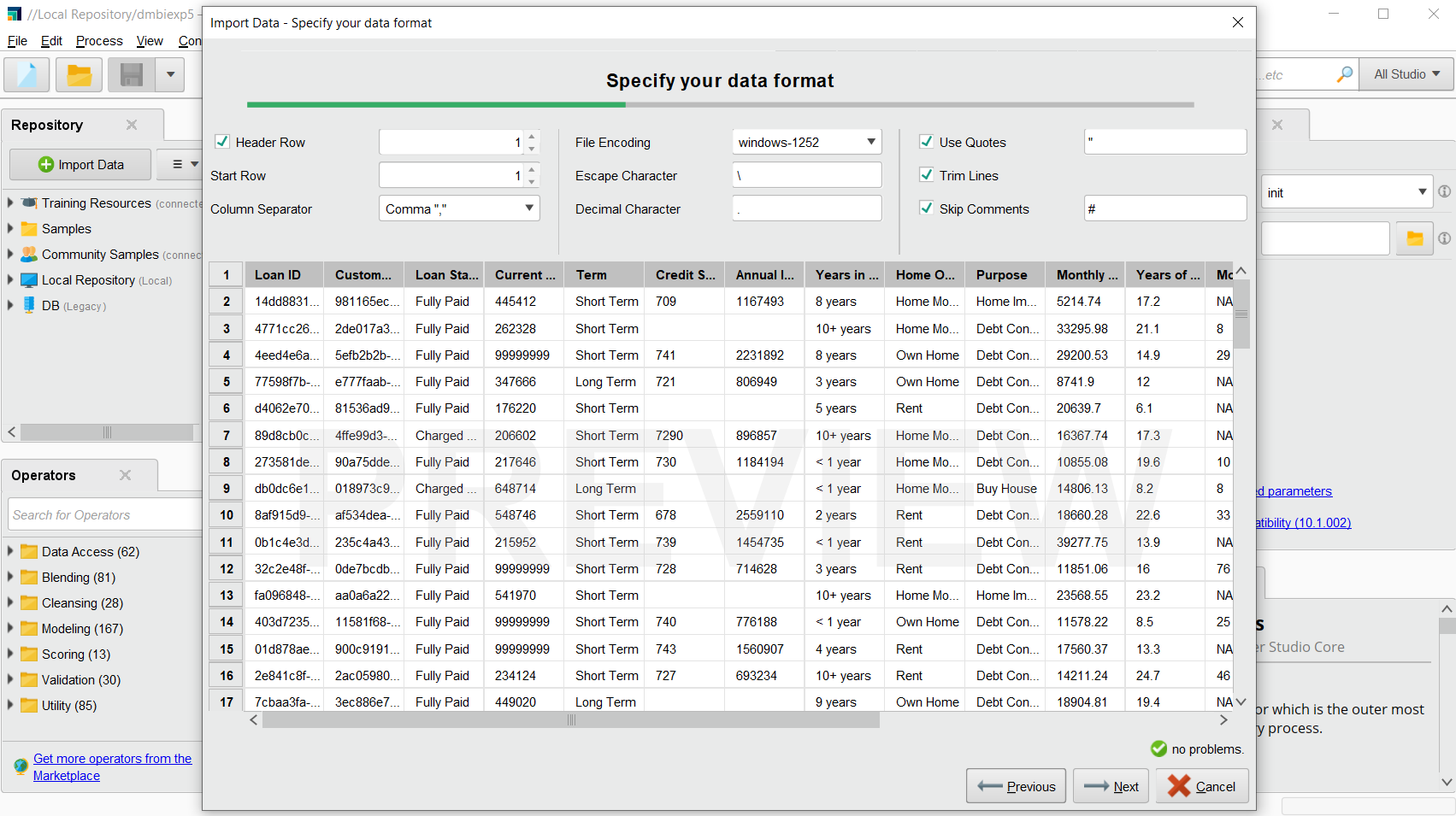
Steps to work with RapidMiner:

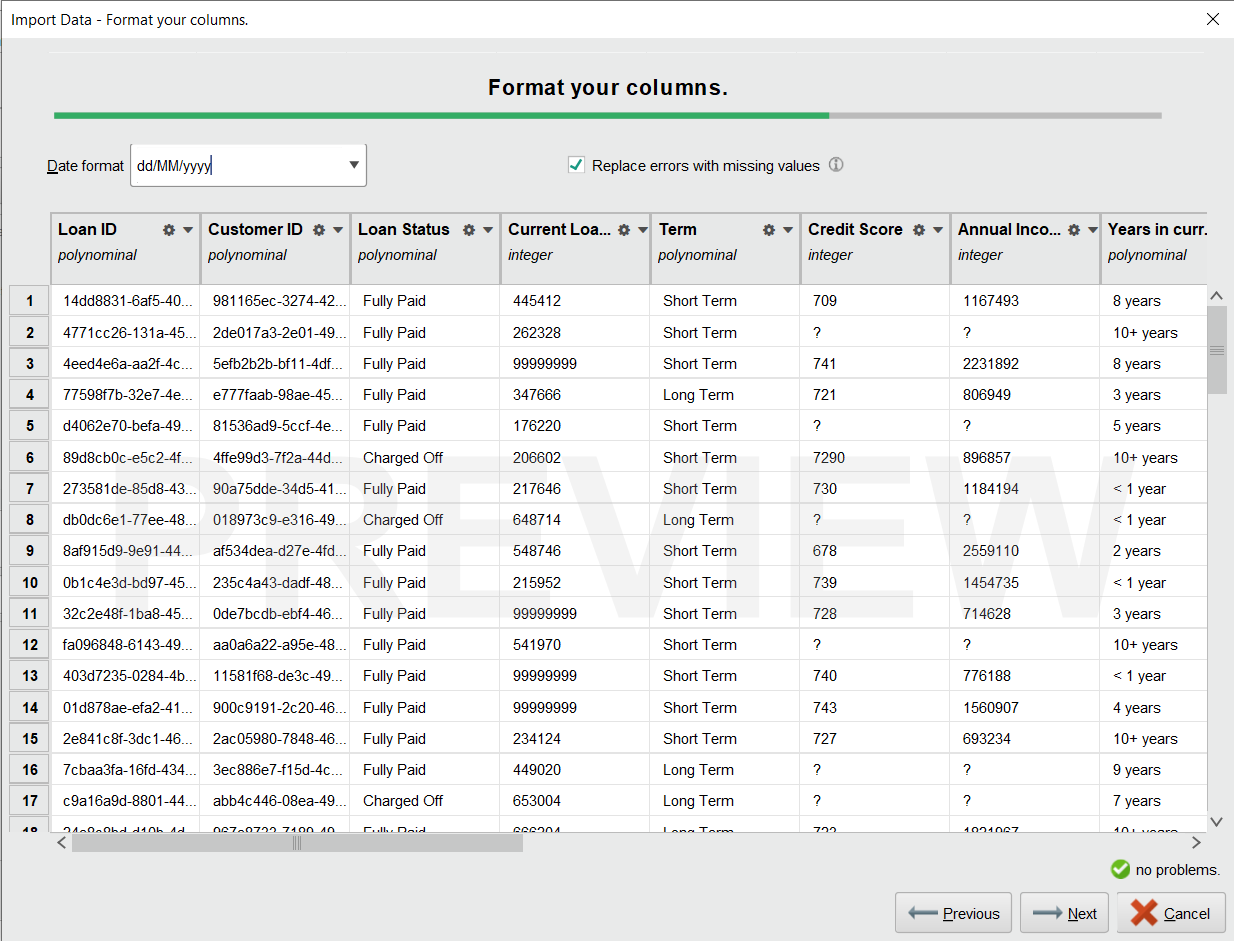
1. Load Data: Open RapidMiner Studio and load your data into the workspace. You can import data from a variety of sources including CSV, Excel, and databases.
2. Data Preprocessing: Preprocess your data by cleaning, transforming, and aggregating it as necessary. You can use a variety of RapidMiner operators to perform these tasks, including filtering, aggregation, and normalization.
3. Model Building: Build a predictive model using RapidMiner's modeling operators. These operators include algorithms for classification, regression, clustering, and association rule mining.
4. Model Evaluation: Evaluate the performance of your model using RapidMiner's evaluation operators. These operators include metrics for accuracy, precision, recall, and F1 score.
5. Visualization: Use RapidMiner's visualization operators to create charts and graphs to help you better understand your data and the results of your analysis.
6. Deployment: Deploy your model by integrating it with other systems or exporting it as a web service using RapidMiner Server.
7. Monitoring and Maintenance: Monitor the performance of your model over time and make adjustments as necessary to ensure that it remains accurate and effective.

**IMPLEMENTATION:**

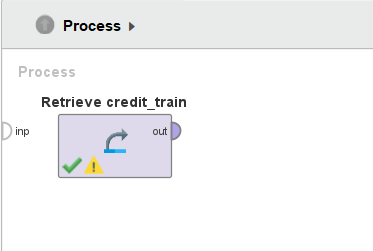
Step 1 : Import the dataset from your local directory



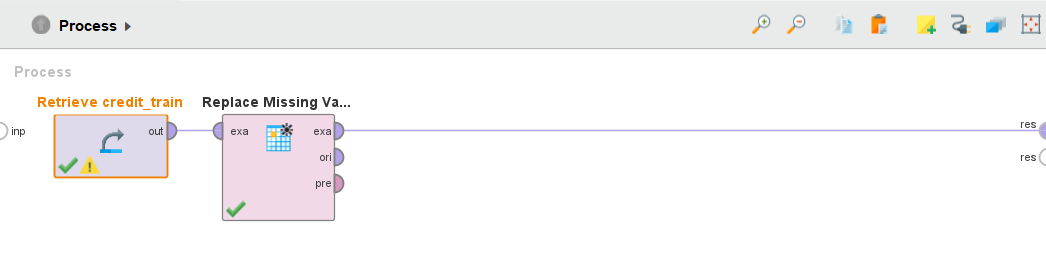




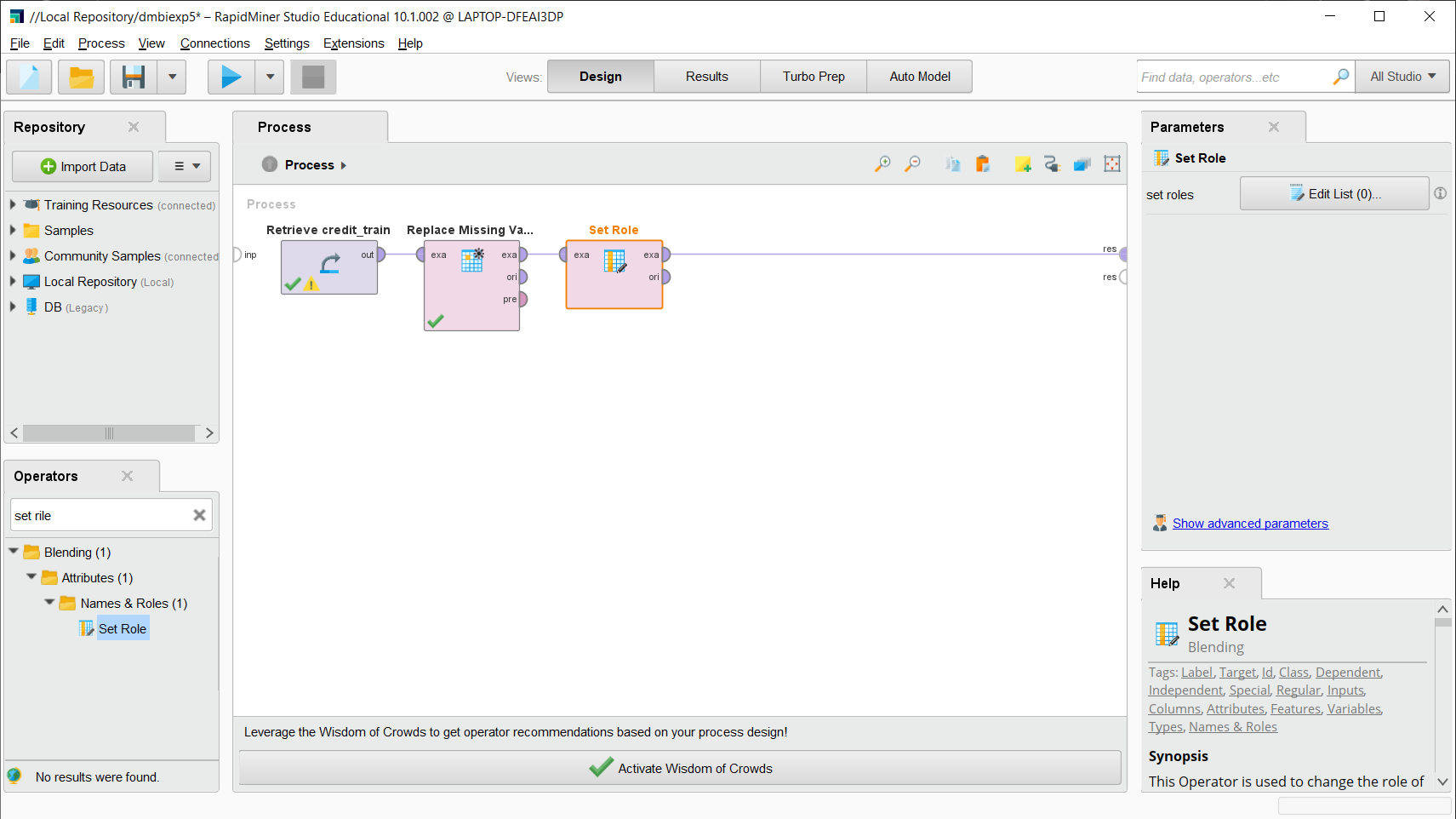
Step 2: Drag and drop the desired data set

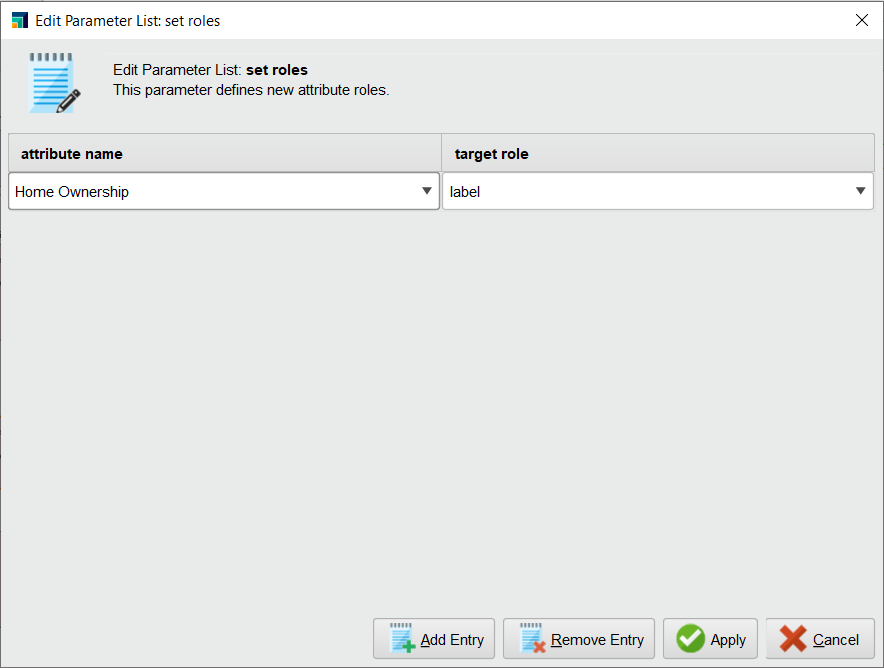


Step 3: Search ‘Replace Missing Values’ to deal with the question marks in the data(missing values).

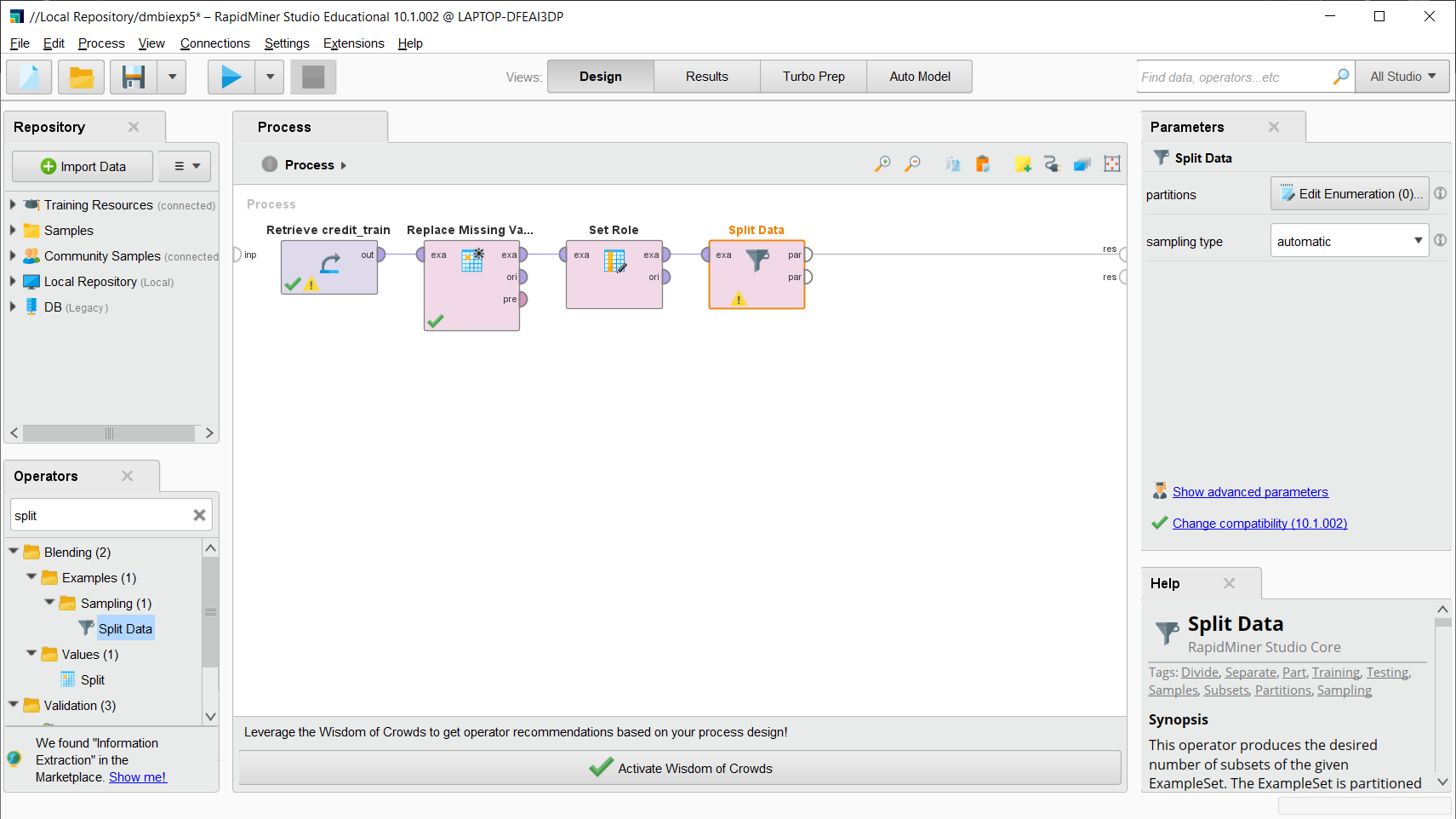


Step 4: Now search Set role and then drag and drop it

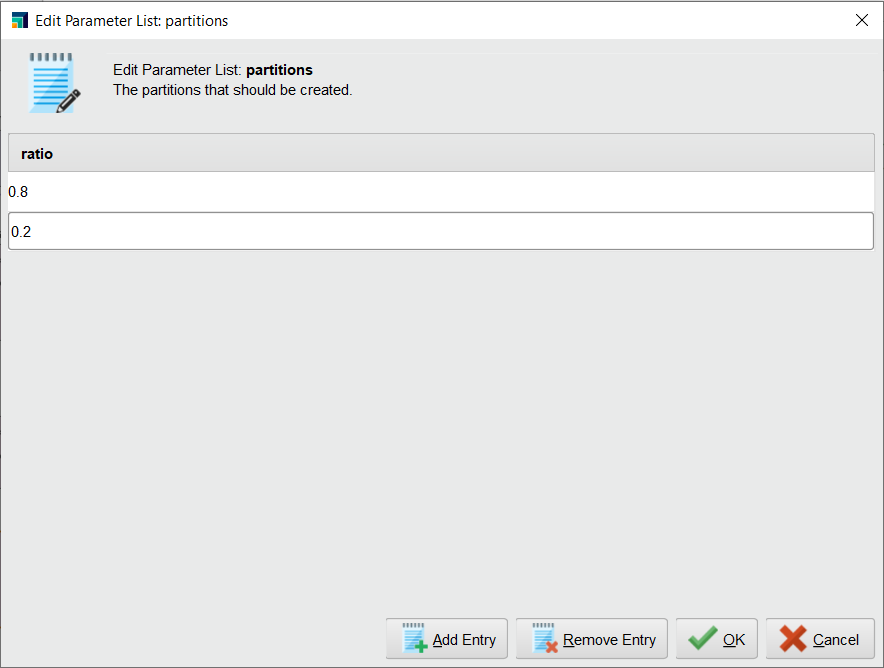




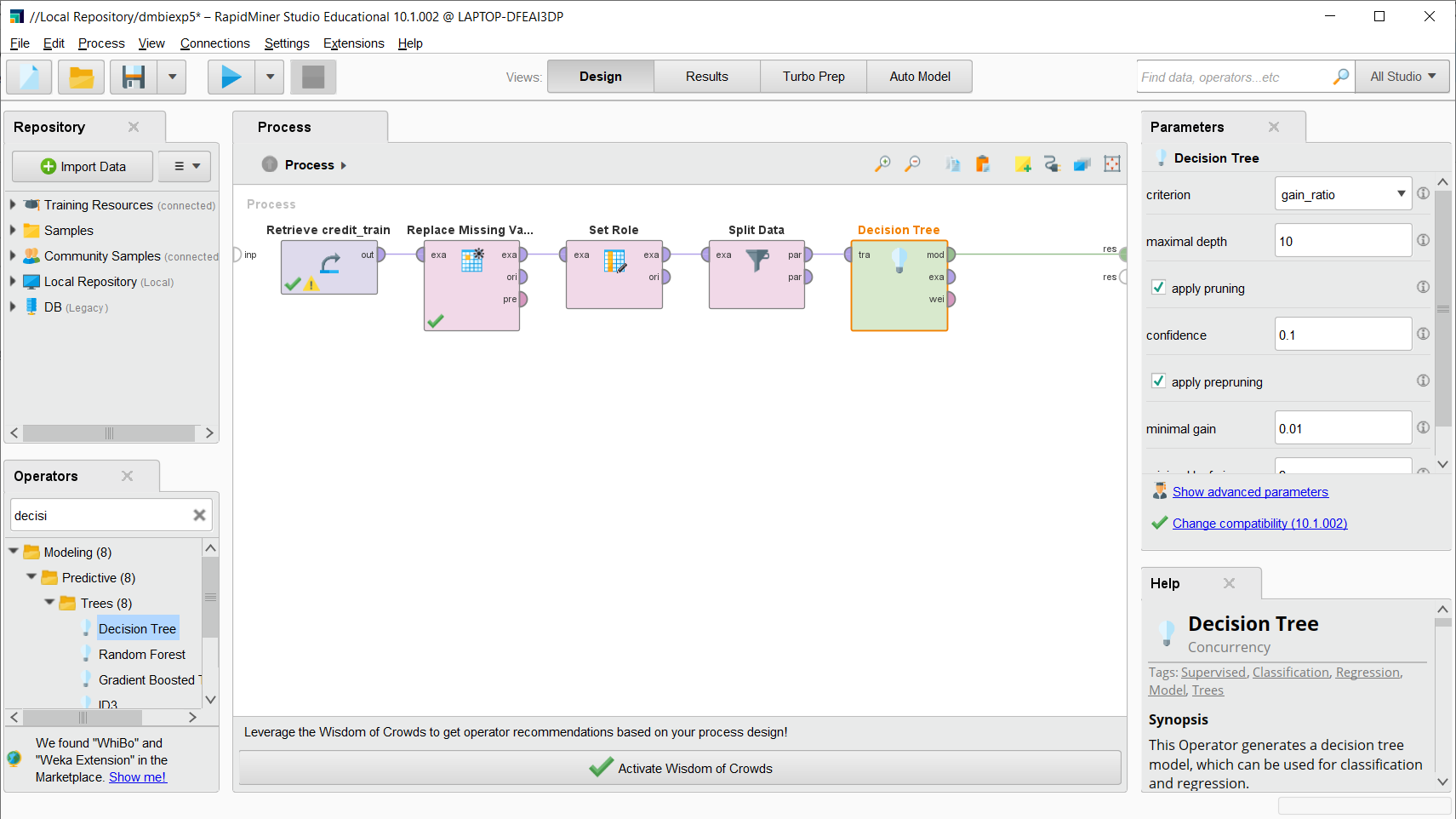
Step 5: As now we want to split the data we will use the split operator



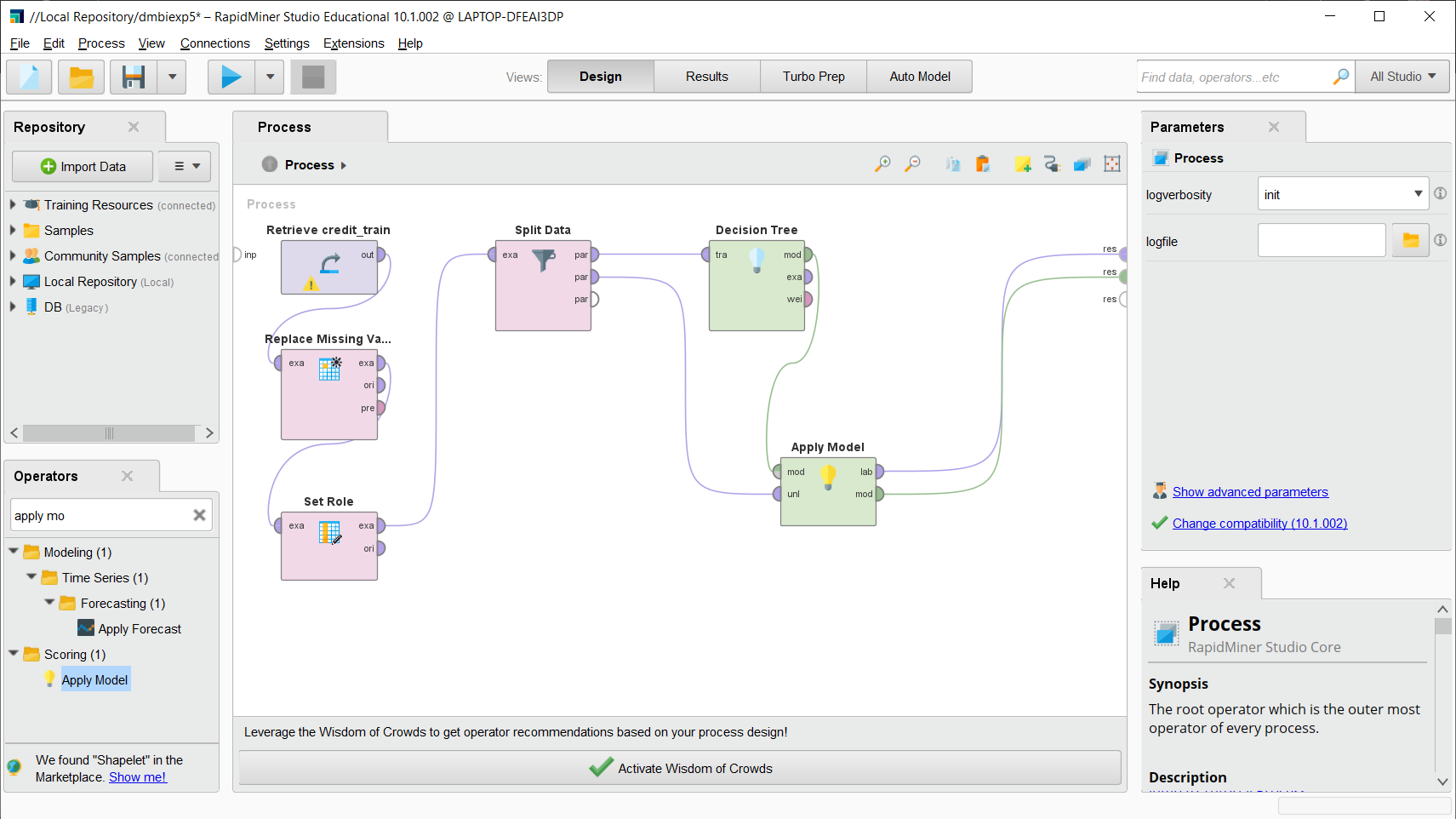
Step 6: Splitting the data in the ratio of 8:2



Step 7: Now search for Decision tree operator and drag and drop it

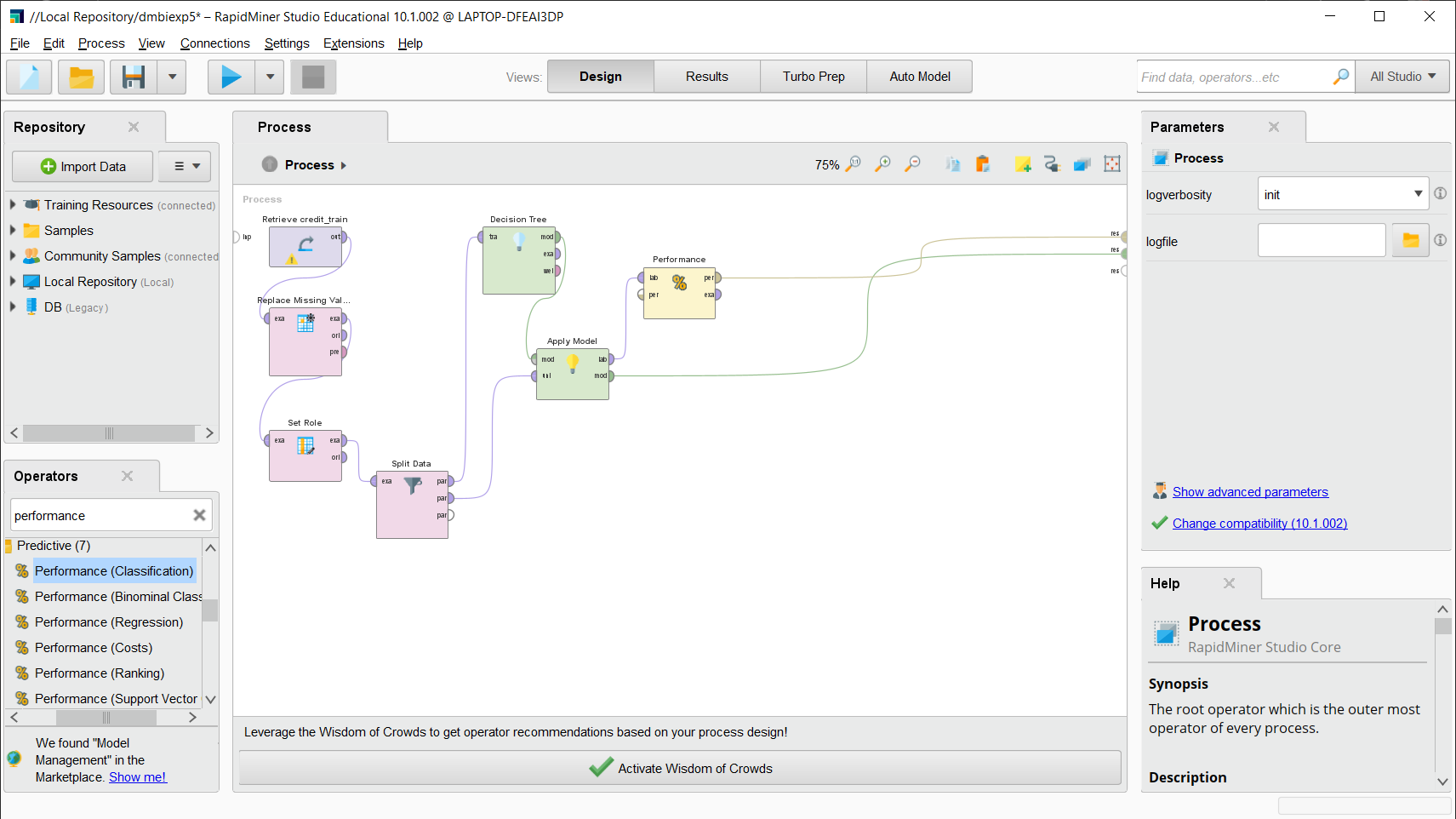


Now add apply model operator as shown below

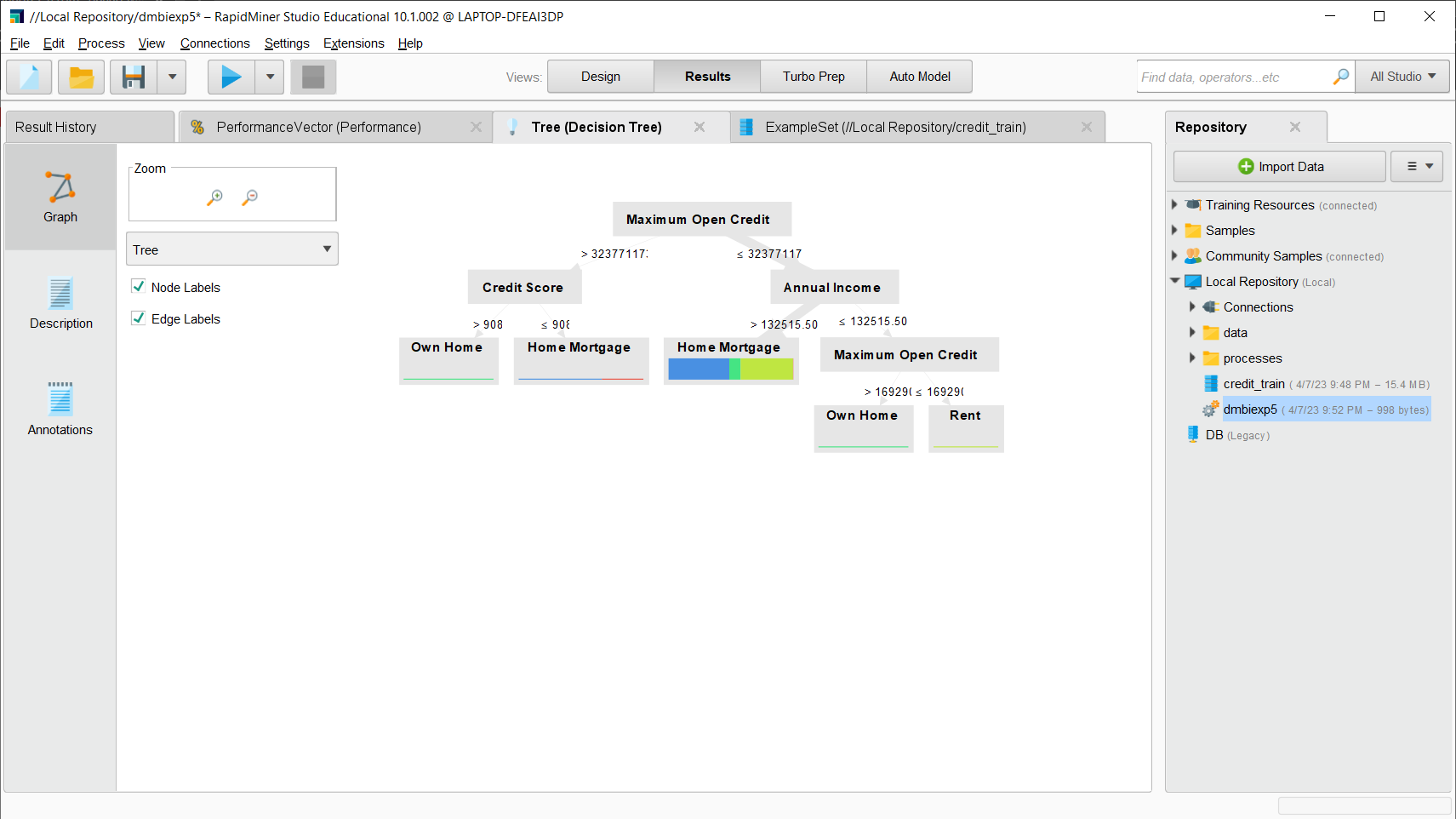


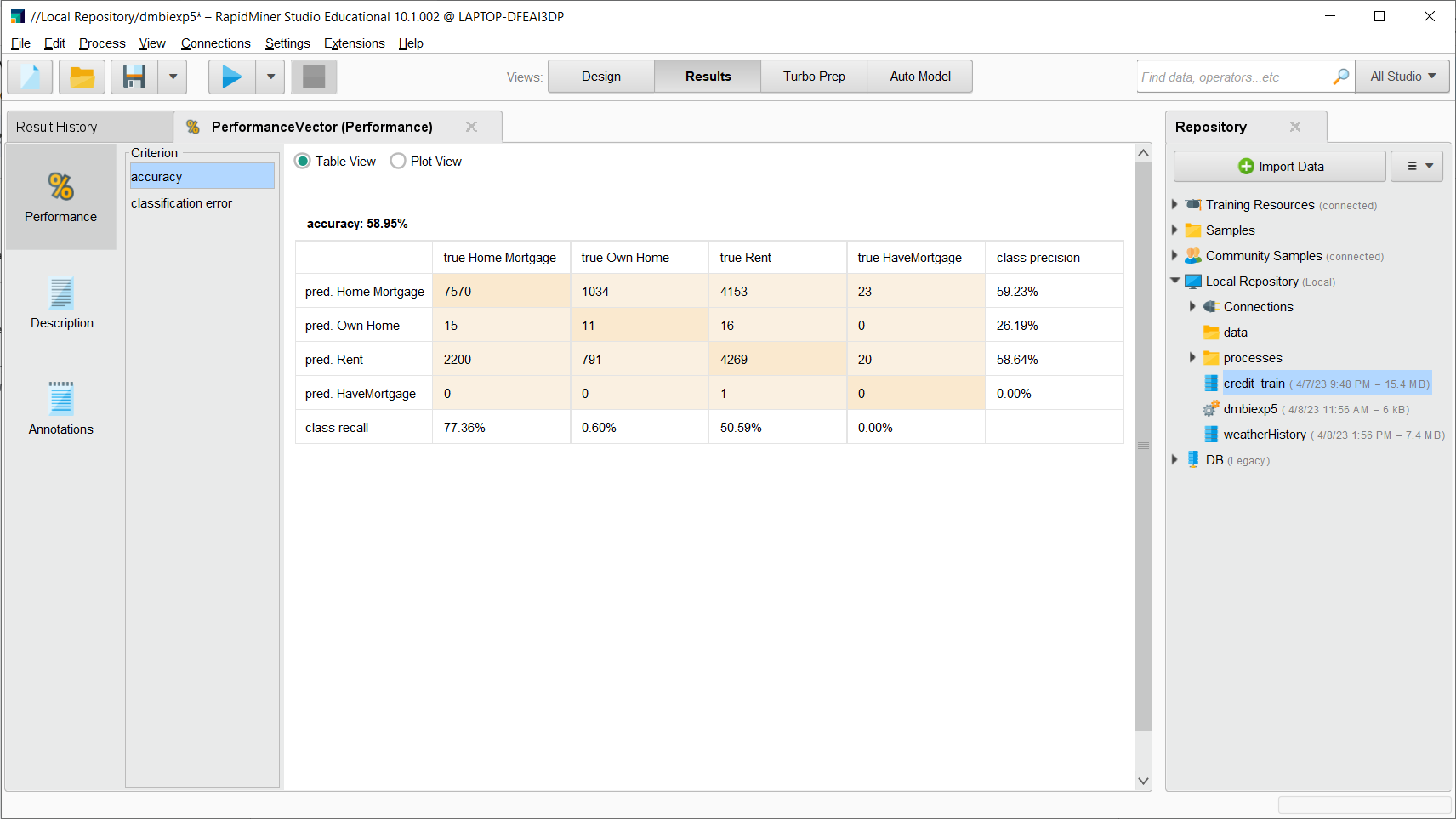
Step 8: Now add the Performance measure operator and then connect it as shown below.

Also do the connection for the model as shown below.



Step 9: Run



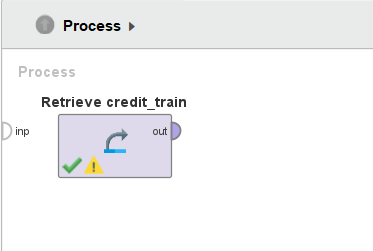


This is the decision tree formed by the given dataset

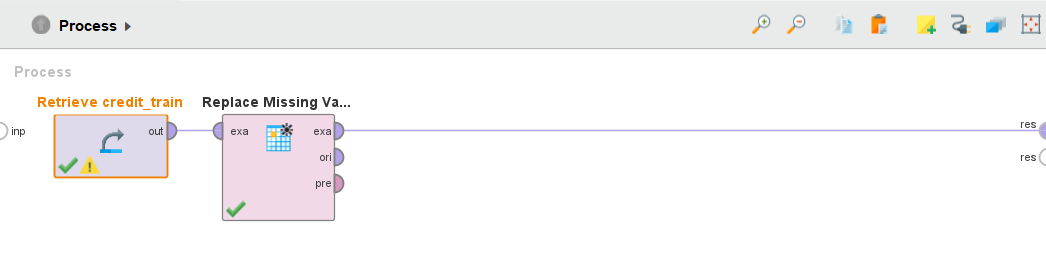
The overall accuracy of Decision Tree is **58.95%**

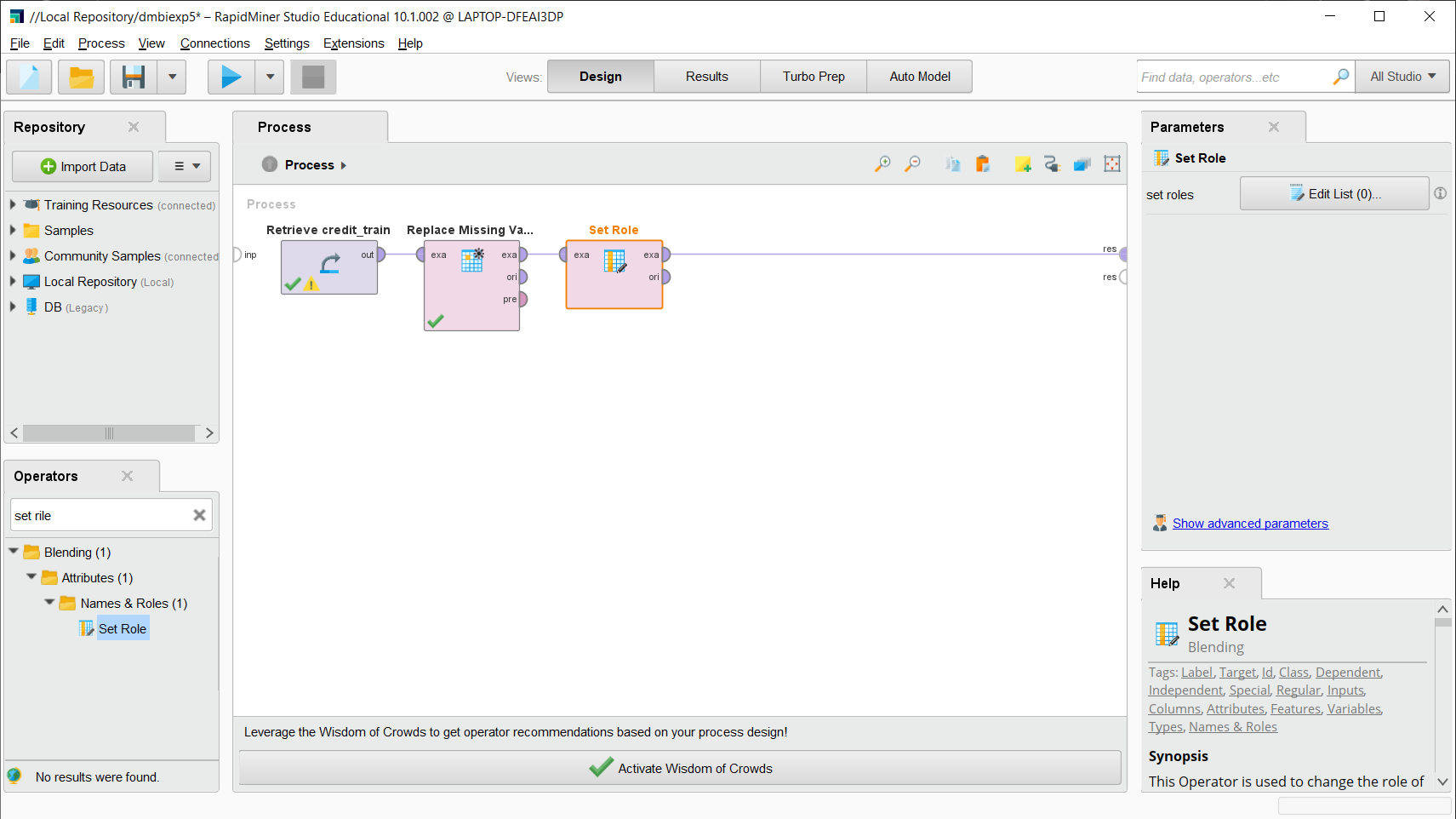
**Implementation of Naive Bayes in RapidMiner**

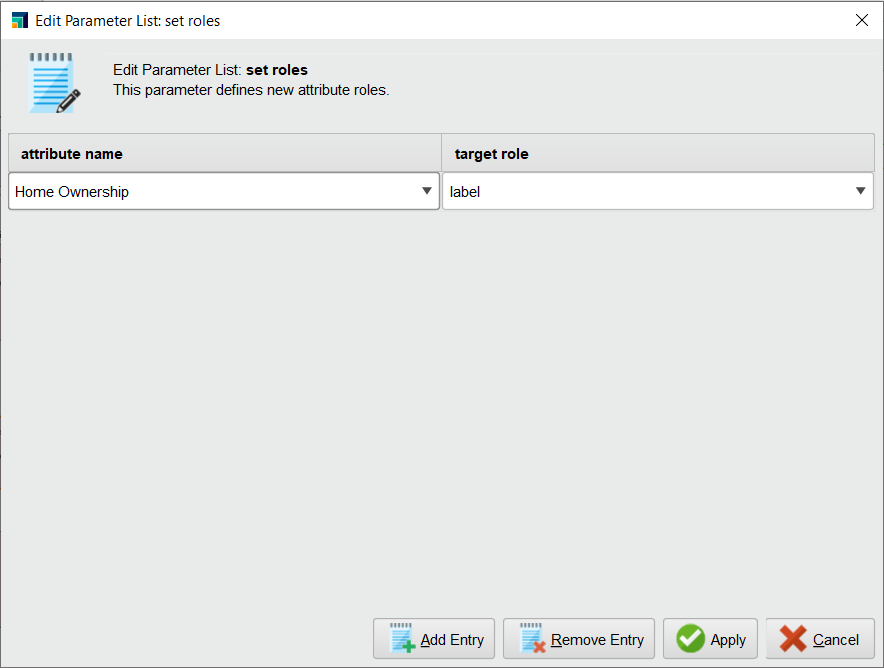
Step 1: Drag and drop the desired data set



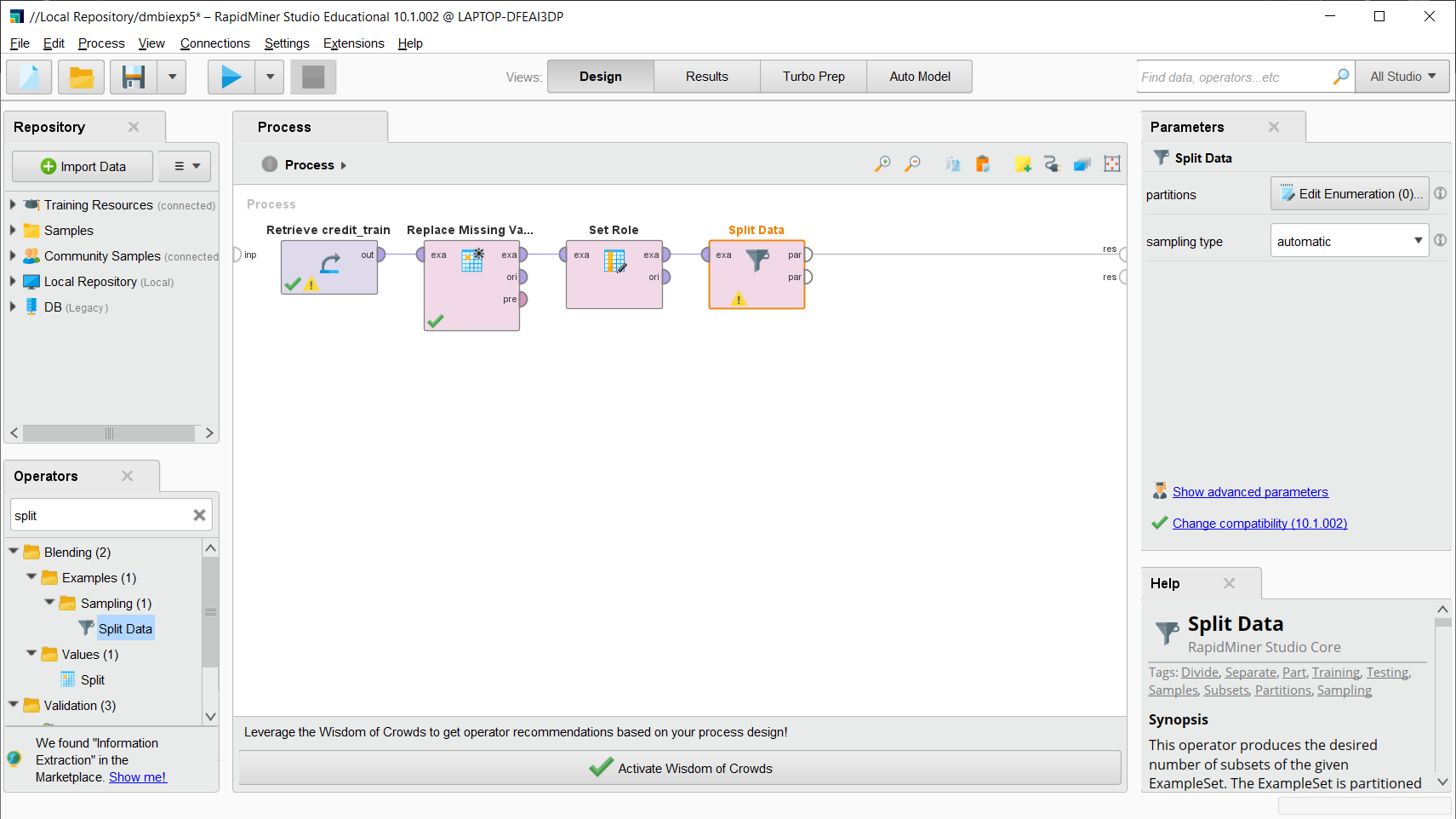
Step 2: Search ‘Replace Missing Values’ to deal with the question marks in the data(missing values).



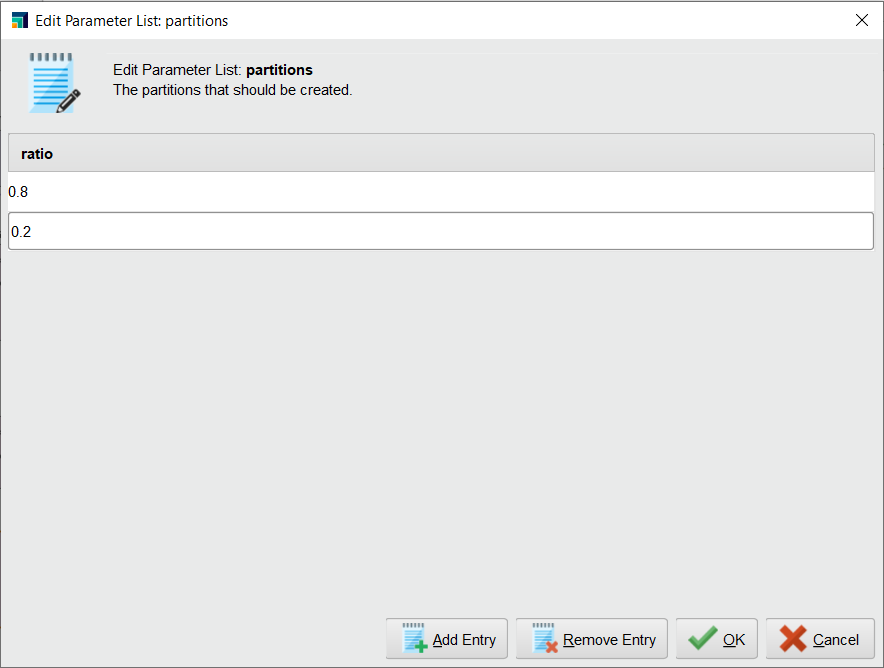
Step 3: Now search Set role and then drag and drop i



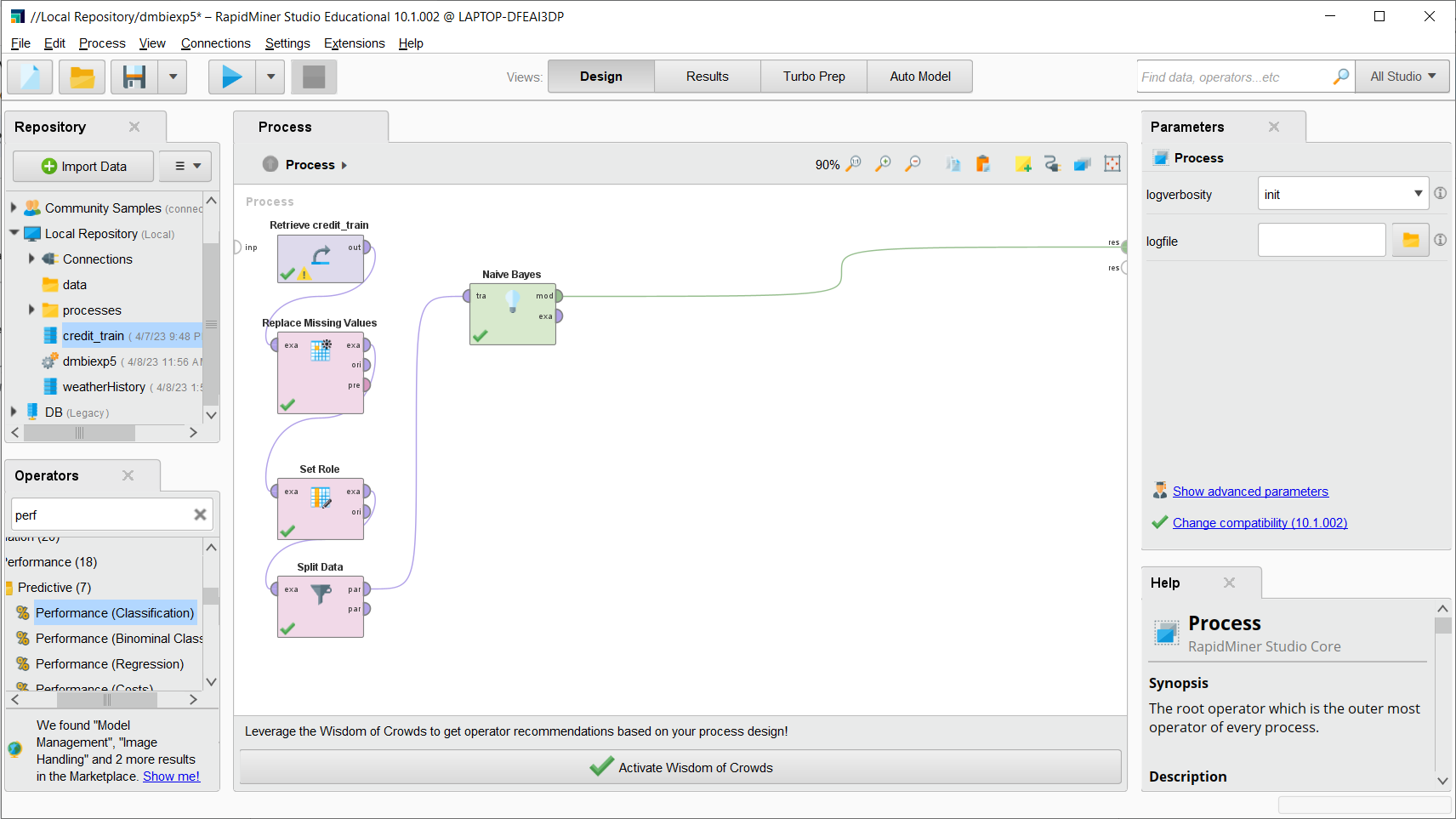
Step 4: As now we want to split the data we will use the split operator



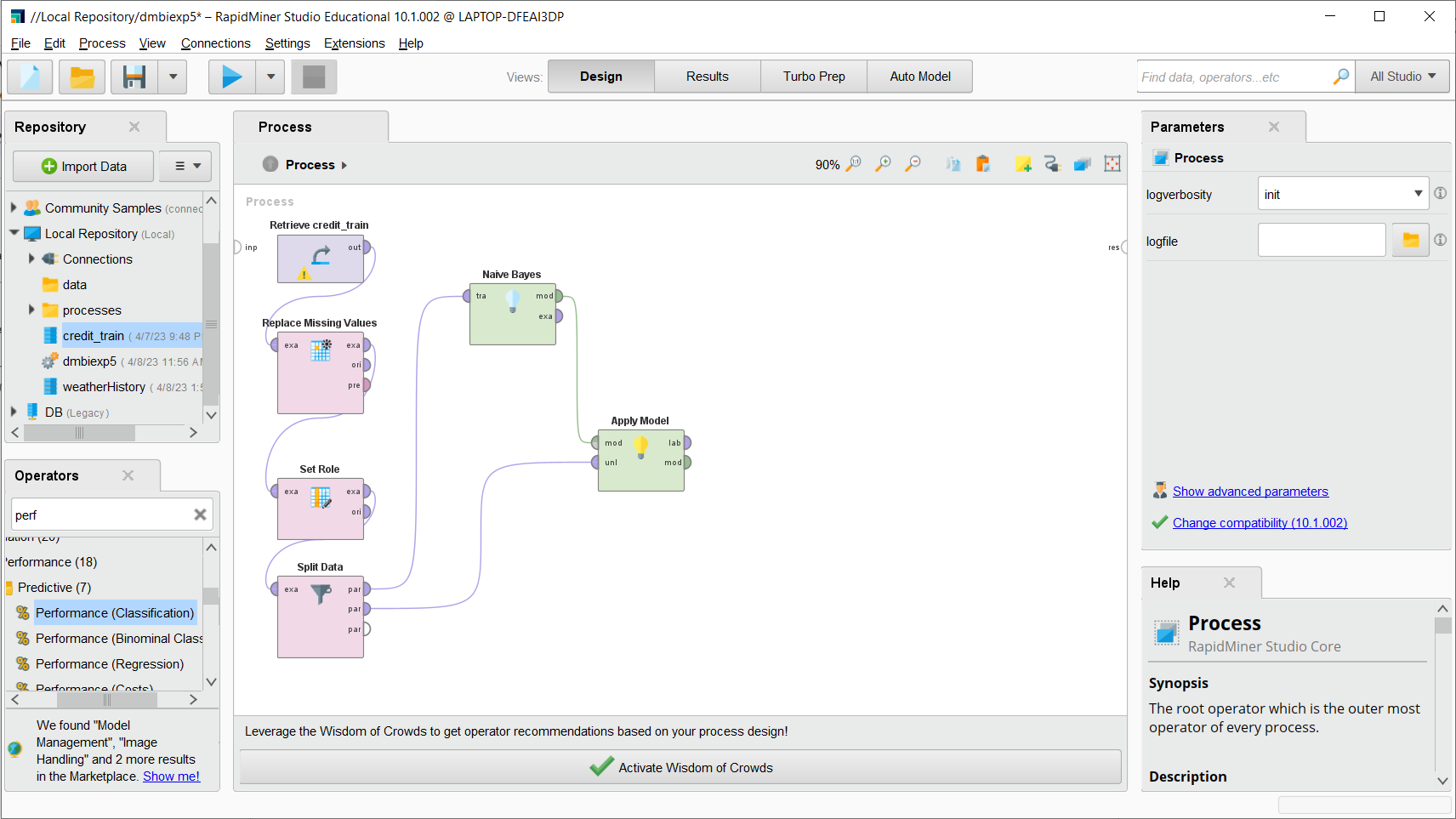
Step 5: Splitting the data in the ratio of 8:2



Step 6: Now search for Naive Bayes operator and drag and drop it

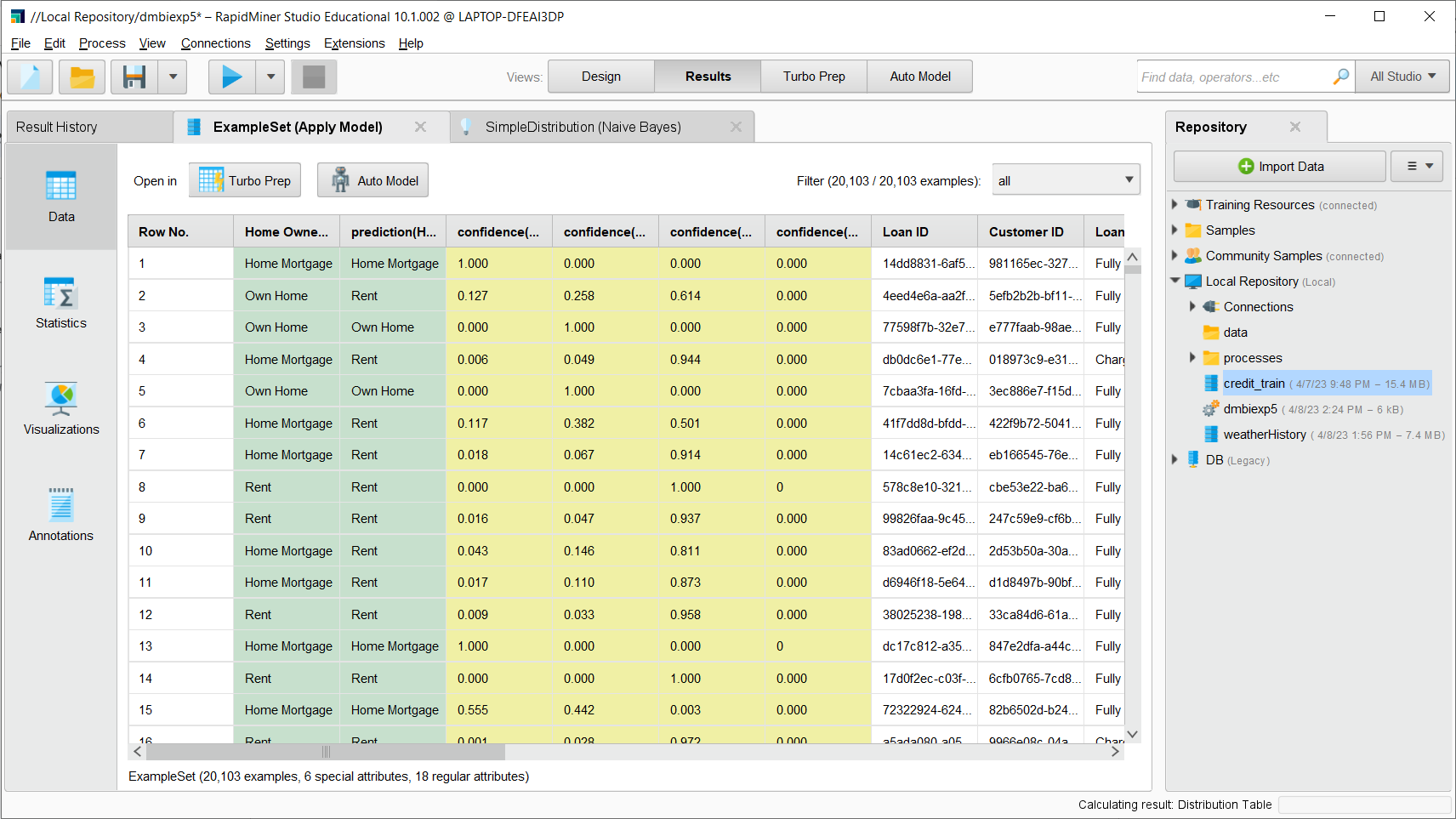


Now add apply model operator as shown below



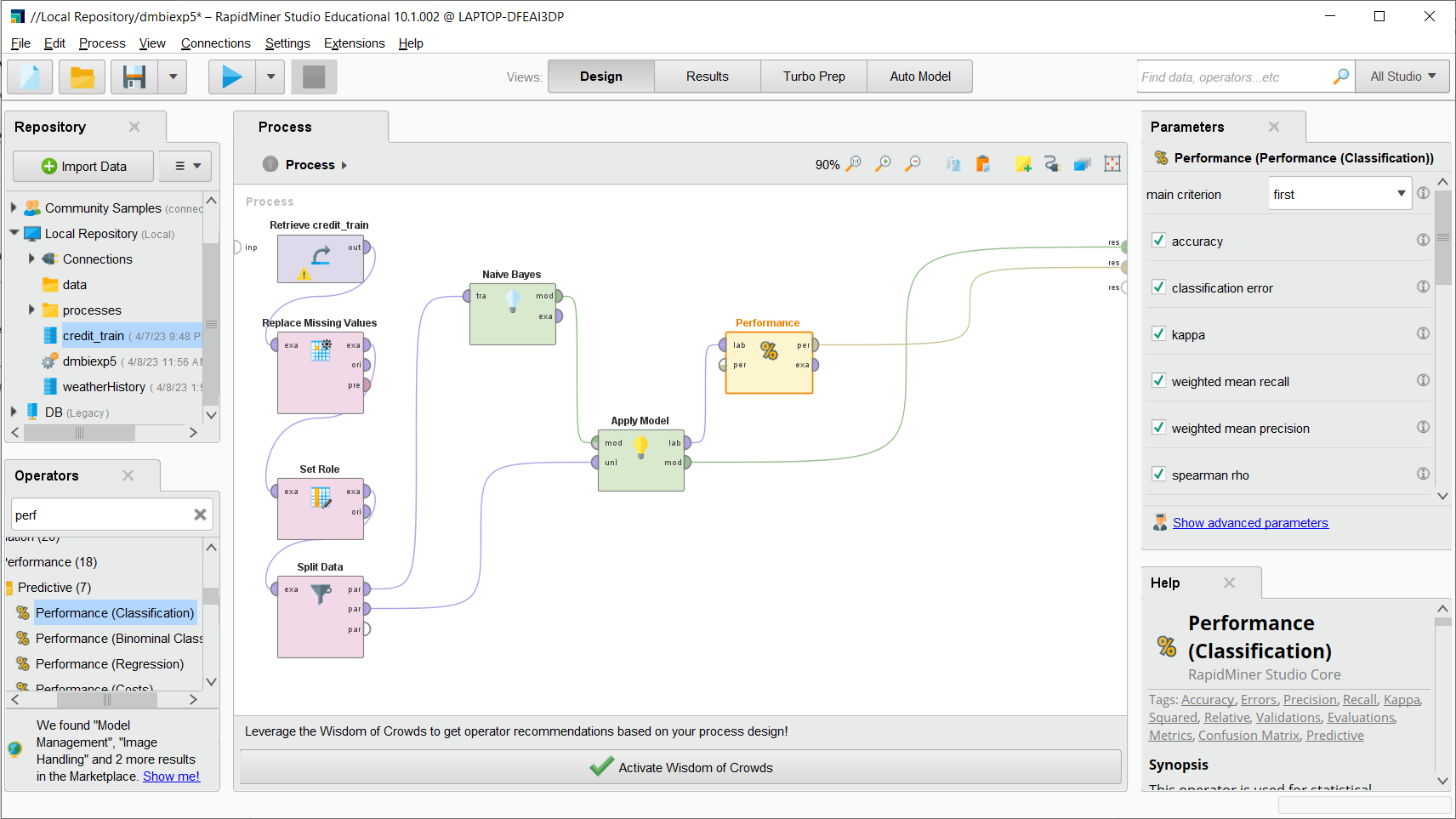
Step 7: After running this model we will get the following output:

The following output shows the actual value and predicted value with confidence



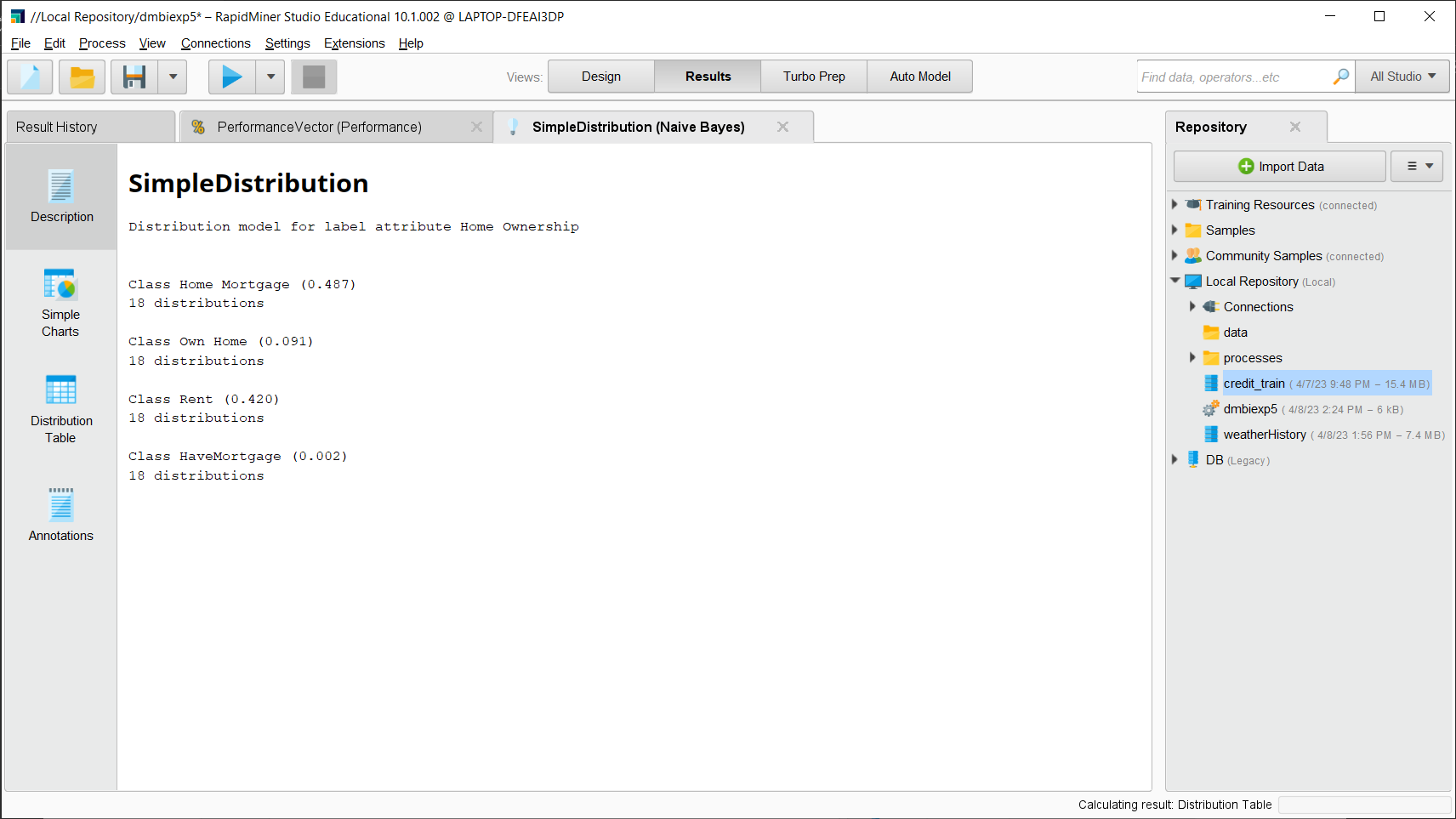
Step 8: Now add Performance measure operator and then connect it as shown below

And also do the connection for model as shown below

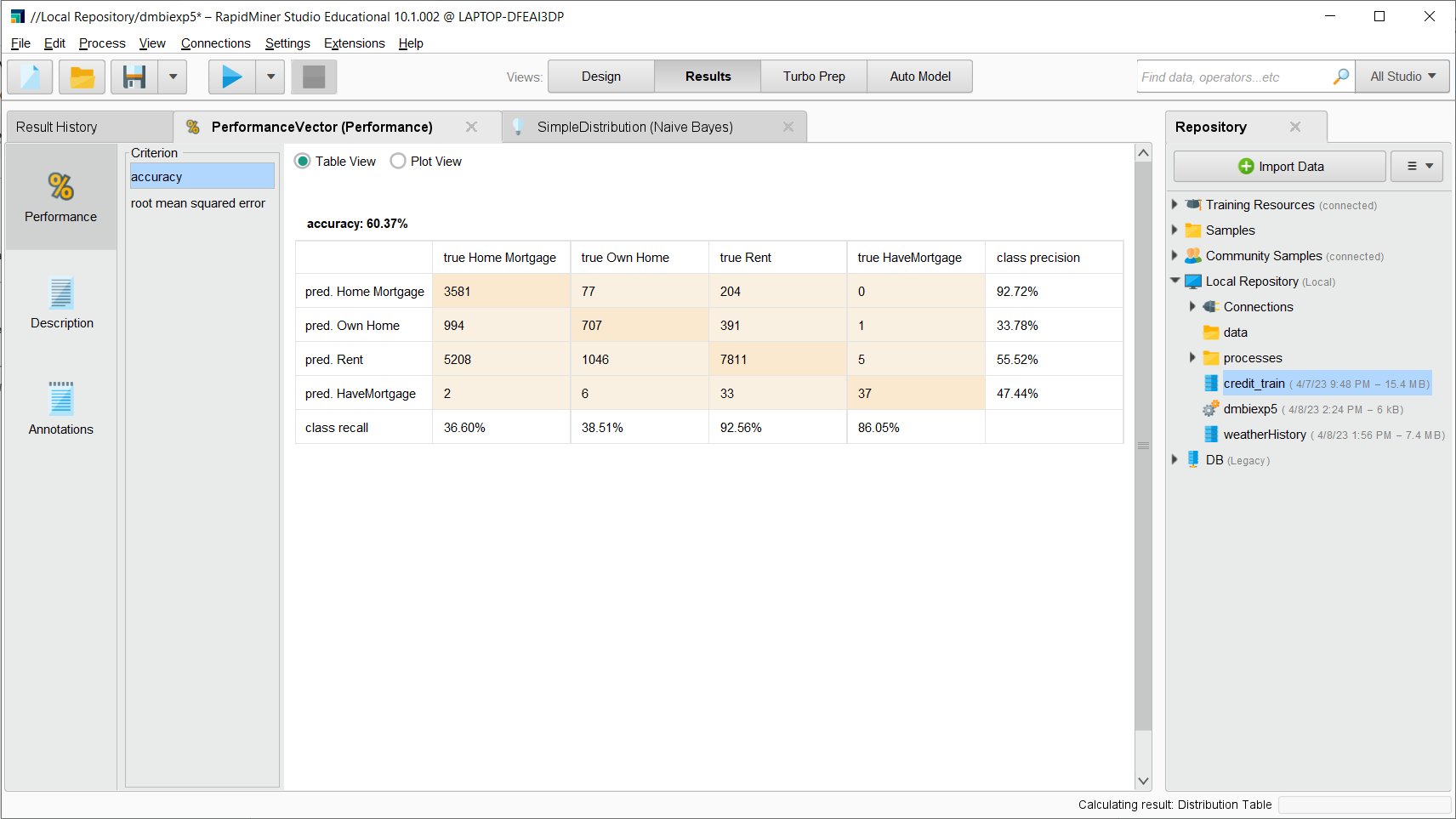


Step 9: After running this model we will get:

This will show the Simple Distribution of Dataset



From the performance Vector we can see the accuracy as follows:



The overall Accuracy of Naive Bayes is **60.37%**

From Given conditions we can clearly conclude that **Naive Bayes** has performed better

than Decision Tree.

**CONCLUSION:**

In this Experiment we saw two classification algorithms that are Decision Tree and Naive

Bayes. We found their accuracies based on the provided dataset of Loan Status.

The performance of Naive Bayes is better than Decision Tree in RapidMiner. We have

successfully implemented Experiment to explore Rapid Miner and implement

classification models like Decision Tree and Naive Bayes etc