

**A**  
**Report on**  
**“RAPIDMINER”**

In partial fulfillment of requirements for the degree of

**Bachelor of Technology (B.Tech.)**  
in  
**Computer Science and Engineering**



**Submitted by:**

Ms. Aashi Barodiya(180009)  
Ms. Akshita Kanther(180161)  
Ms. Harshita Das(180043)  
Ms. Muskan Tayal(180068)  
Ms. Prashi Jain(180186)  
Ms. Priyanka Motiani(180093)  
Ms. Shreya Sharma(180123)

**Under the Guidance of**  
**Dr. Puneet Kumar Gupta**

*Department of Computer Science and Engineering*  
**SCHOOL OF SCIENCE AND TECHNOLOGY**

**Mody University and Science and Technology**  
**Lakshmangarh, Distt. Sikar-332311**  
**May 2021**

## **ACKNOWLEDGMENT**

We would like to express our gratitude to our project mentor Dr.Puneet Kumar Gupta for his guidance and encouragement during the completion of the project. His kindness and help have been the source of encouragement for us. It is our privilege to express our sense of gratitude towards School of Engineering and Technology, Mody University, Lakshmangarh. We would like to thank Dr. A Senthil, Assistant Dean SET for providing us the golden opportunity to do this wonderful project.

Aashi Barodiya  
Akshita Kanther  
Harshita Das  
Muskan Tayal  
Prashi Jain  
Priyanka Motiani  
Shreya Sharma

## TABLE OF CONTENTS

| Sr.No. | Topics                              | Page No. |
|--------|-------------------------------------|----------|
|        | Acknowledgement                     | 2        |
| 1      | Introduction                        | 4        |
| 2      | Evolution                           | 5        |
| 3      | Detailed description of features    | 6-18     |
| 4      | Reporting and Business Intelligence | 19-21    |
| 5      | Pricing                             | 22       |
| 6      | Conclusion                          | 23       |
| 7      | My contribution in report           | 24       |
| 8      | References                          | 25       |

## **1. INTRODUCTION:**

Rapid Miner is one of the best data mining tool. It is a software program that we use in data mining, predictive analysis as well as in text mining. Here, we enter raw data through programs which includes databases and text and that we analysed on large scale. It also includes free trial where we can assess it's capabilities. Rapid Miner serves a platform to developers to create algorithms that helps in analysis of data. Rapid Miner has written in Java programming language.

Rapid Miner provides us various procedures like data mining and machine learning. It includes loading of data, transformation (ETL), visualization of data, processing of data ,predictive analysis. It includes statistical modelling, evaluation and deployment also.

Companies that uses rapid miner are:

- Stack
- QuantiModo
- Rumble
- RedHat

Rapid Miner is used for many purposes and some of them are research, education training, development of applications, fast prototyping. There are many products of rapid miner also and some of them are Rapid Miner Studio, RapidMiner Go, Rapid Miner Server.

Processes that can be done by rapidminer handle missing values, normalization and outlier detection , pivoting and renaming, macros and sampling , looping, branching and appending and writing data. Some of the processes are operators and processes , modeling, accessing data, filtering and sorting , merging and grouping . creating and removing columns, changing types and roles.

## **2. EVOLUTION:**

15 years ago, Rapidminer was initially released in 2006 and 14 months ago, it was released stabelly-9.6/2 march. The company named rapidminer IDE developed rapidminer for preparation of data, machine learning, text mining, deep learning and predictive analysis. The rapidminer's operating system is Cross-Platform .

- **History:**

RapidMiner also knowns as YALE(Yet Another Learning Environment) and was started developing in 2001 by Ralf Klinkenberg, Ingo Mierswa, and Simon Fischer at the Artifical Intelligence Unit of the Technical University of Dortmund. In the starting of 2006,the development was guided by Rapid-I, a company founded by Ingo Mierswa and Ralf Klinkenberg in the same year.

The software's name was changed from YALE to RapidMiner in 2007.

The company rebranded the Rapid-I to RapidMiner in 2013.

- **Adoption:**

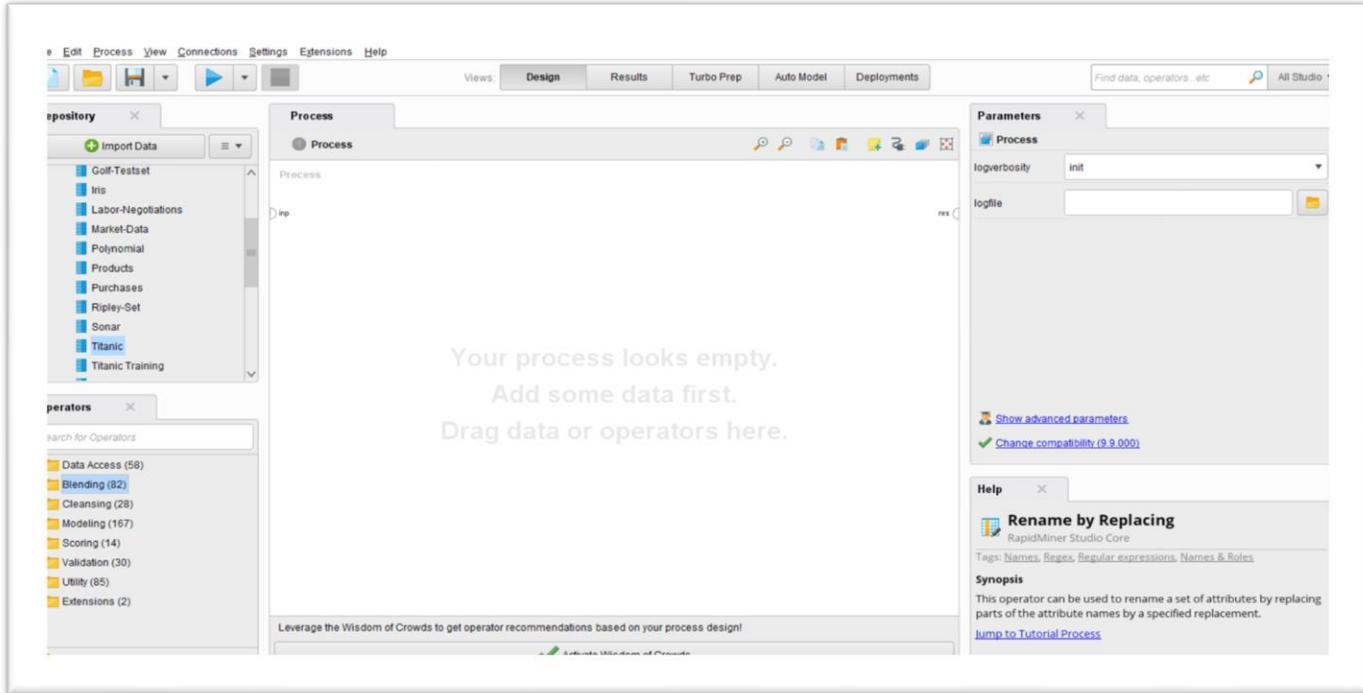
In 2019,RapidMiner was settled in the leader quadrant of it's Magic Quadrant for Data Science and machine learning platforms by Gartner. Rapid Miner serves deep and broad modelling capabilities for automated end to end development of model.

In the year of 2018 annual software poll, RapidMiner was selected as one of the most popular software for analysis of data. RapidMiner has over 400,000 users and it got downloads in millions. BMW, Intel, Cisco, GE and Samsungs are it's paying customers. RapidMiner declared as a market leader against SAS and IBM in the software for data science platforms.

- **Developer:**

Developers world wide participated in developing in open source and majority of the developers are employees of RapidMiner. It got 1.6 crores Series C funding with partification from venture capital firms Nokia Growth partners, Ascent Venture Partners,Longworth Venture Partners ,Earlybird Venture Capital and Open Ocean. OpenOcean partner “Monty” Widenius is a founder of MYSQL.

### 3. RAPIDMINER INTERFACE-



#### Importing dataset

Repository → Import Data → Select where your data is (My Computer/Database) → Select the Excel file → Select the cells to import → Format the columns if required (renaming, changing label, removing a column etc)

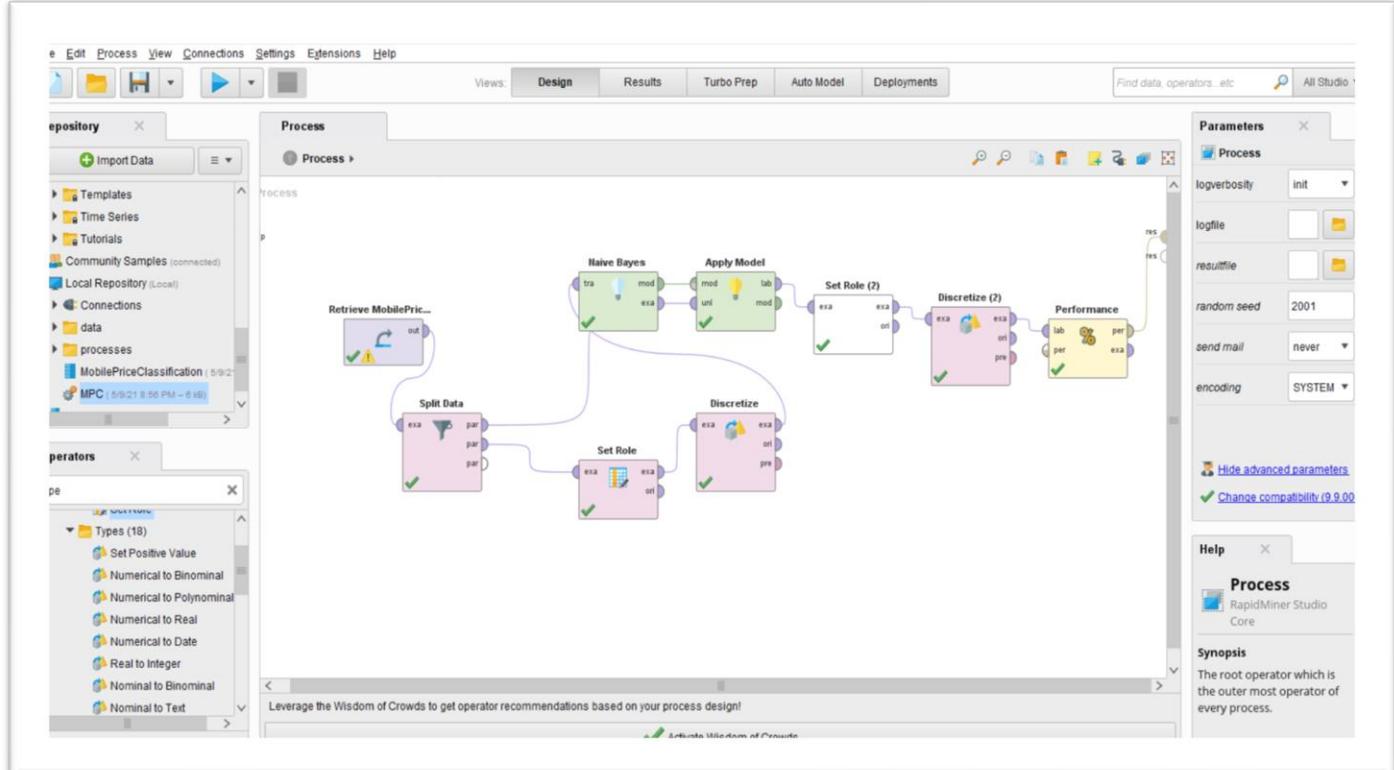
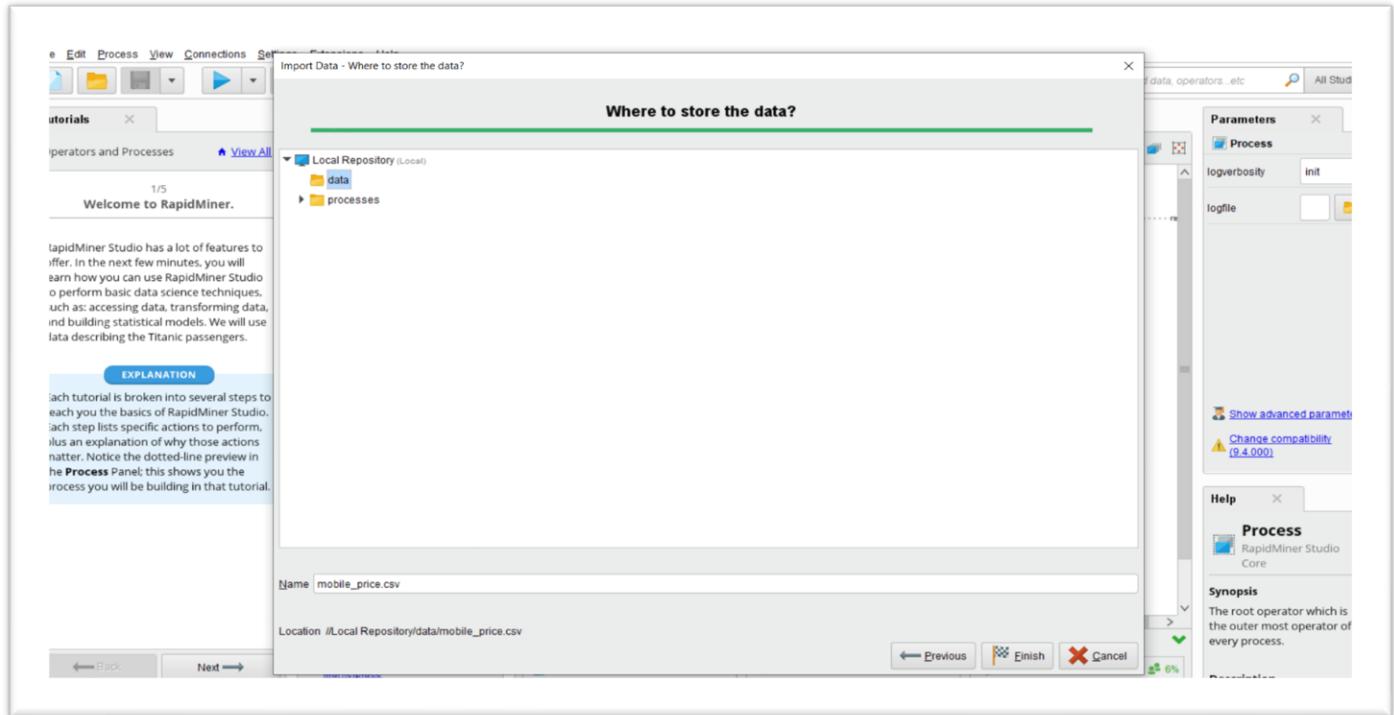
The first screenshot shows the 'Import Data - Format your columns.' dialog for a dataset with columns: battery\_power, blue, clock\_speed, dual\_sim, fc, four\_g, int\_memory, m\_dep, and rear\_view\_camera. The second screenshot shows the same dialog for a dataset with columns: screen\_height, screen\_width, talk\_time, three\_g, touch\_screen, wifi, and price\_range.

|   | battery_power | blue | clock_speed | dual_sim | fc | four_g | int_memory | m_dep |
|---|---------------|------|-------------|----------|----|--------|------------|-------|
| 1 | 842           | 0    | 2.200       | 0        | 1  | 0      | 7          | 0.600 |
| 2 | 1021          | 1    | 0.500       | 1        | 0  | 1      | 53         | 0.700 |
| 3 | 563           | 1    | 0.500       | 1        | 2  | 1      | 41         | 0.900 |
| 4 | 615           | 1    | 2.500       | 0        | 0  | 0      | 10         | 0.800 |
| 5 | 1821          | 1    | 1.200       | 0        | 13 | 1      | 44         | 0.600 |
| 6 | 1859          | 0    | 0.500       | 1        | 3  | 0      | 22         | 0.700 |
| 7 | 1821          | 0    | 1.700       | 0        | 4  | 1      | 10         | 0.800 |

|   | screen_height | screen_width | talk_time | three_g | touch_screen | wifi | price_range |
|---|---------------|--------------|-----------|---------|--------------|------|-------------|
| 1 | 9             | 7            | 19        | 0       | 0            | 1    | 1           |
| 2 | 17            | 3            | 7         | 1       | 1            | 0    | 2           |
| 3 | 11            | 2            | 9         | 1       | 1            | 0    | 2           |
| 4 | 16            | 8            | 11        | 1       | 0            | 0    | 2           |
| 5 | 8             | 2            | 15        | 1       | 1            | 0    | 1           |

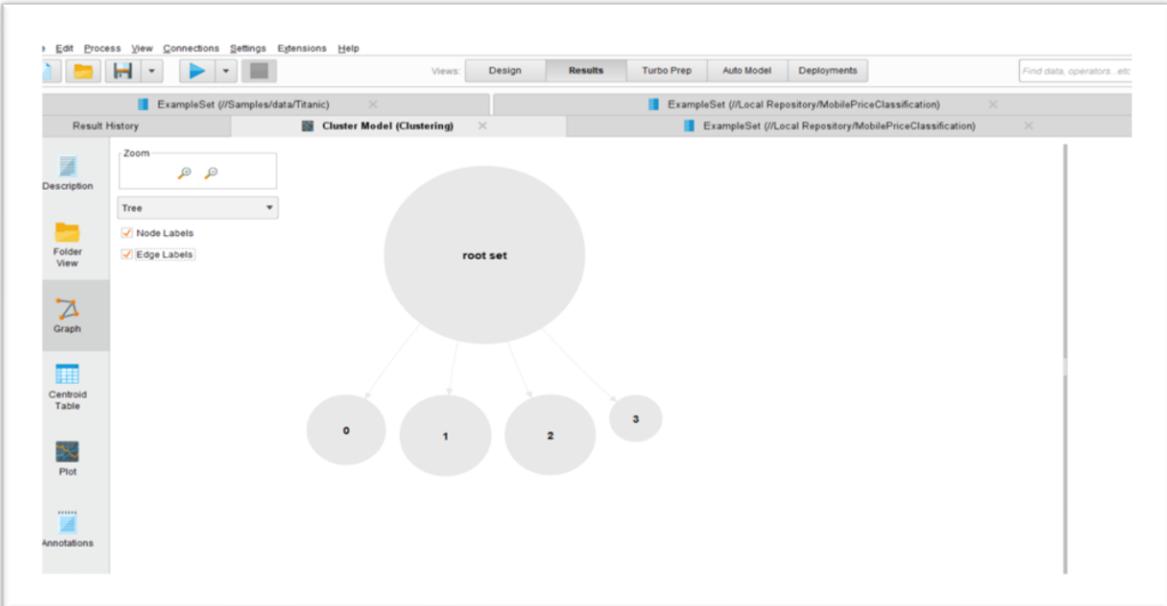
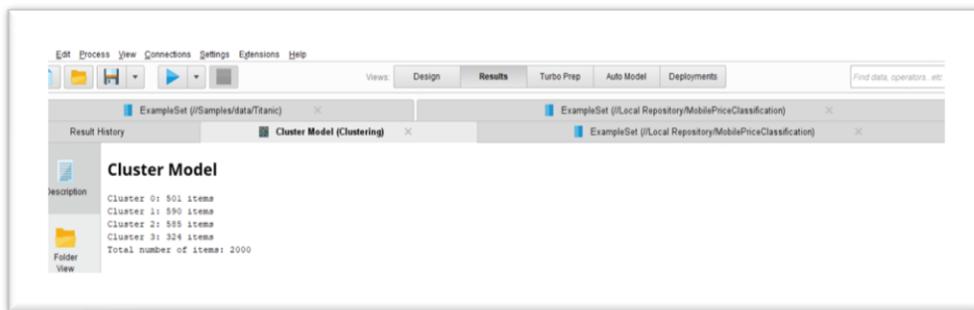
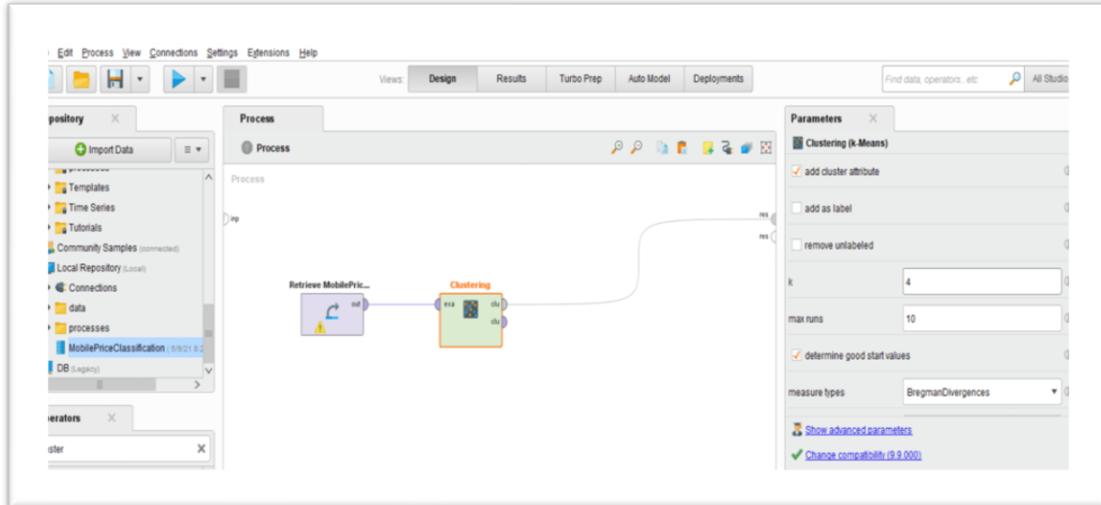
## Select Local Repository to store your data



## Detailed description of features:-

Main features of RapidMiner are:

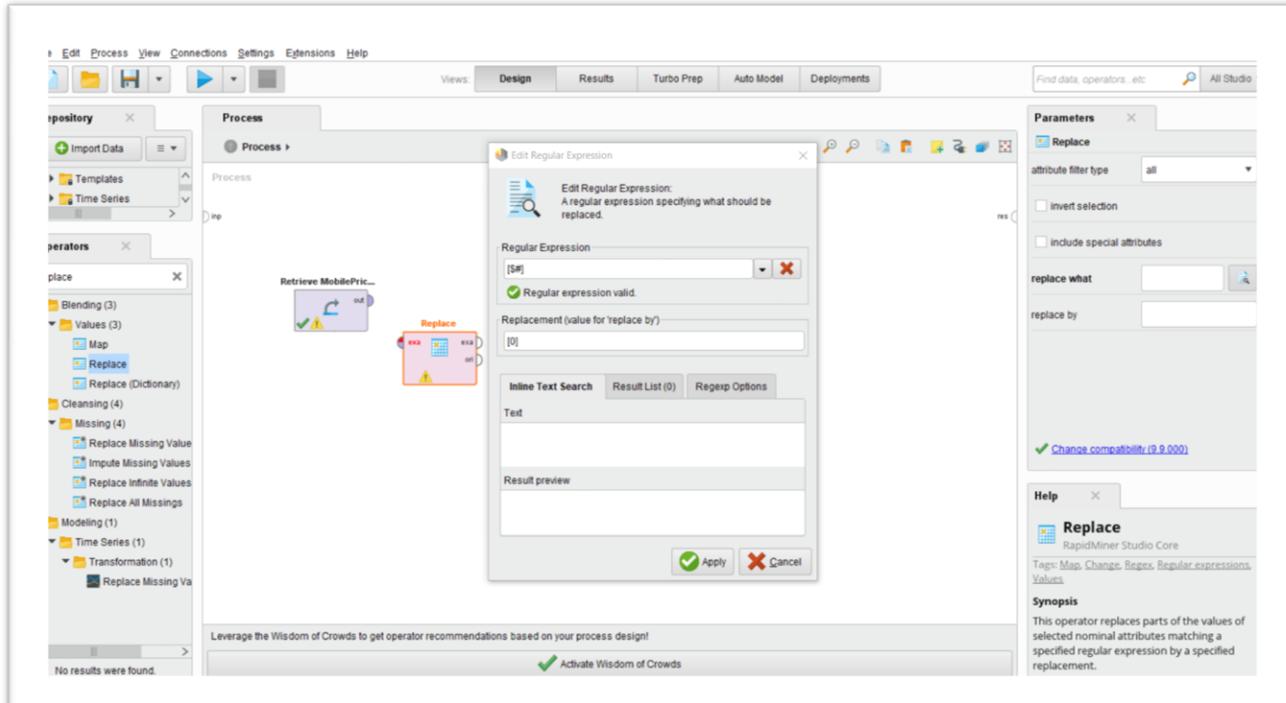
- **Clustering:** The process of grouping similar objects together in a cluster is done by **Clustering** operator. The number of clusters to be formed is to be mentioned in the Parameter section.

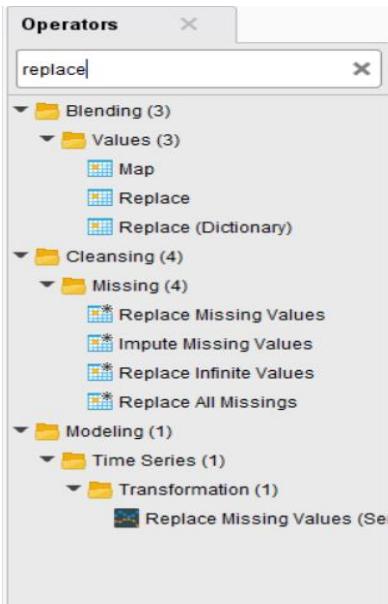


The screenshot shows the RapidMiner interface with a central table view titled "Centroid Table". The table has columns for "Attribute" and four clusters: "cluster\_0", "cluster\_1", "cluster\_2", and "cluster\_3". The rows represent different mobile phone features. A vertical sidebar on the left contains icons for "Description", "Folder View", "Graph", "Centroid Table", "Plot", and "Annotations".

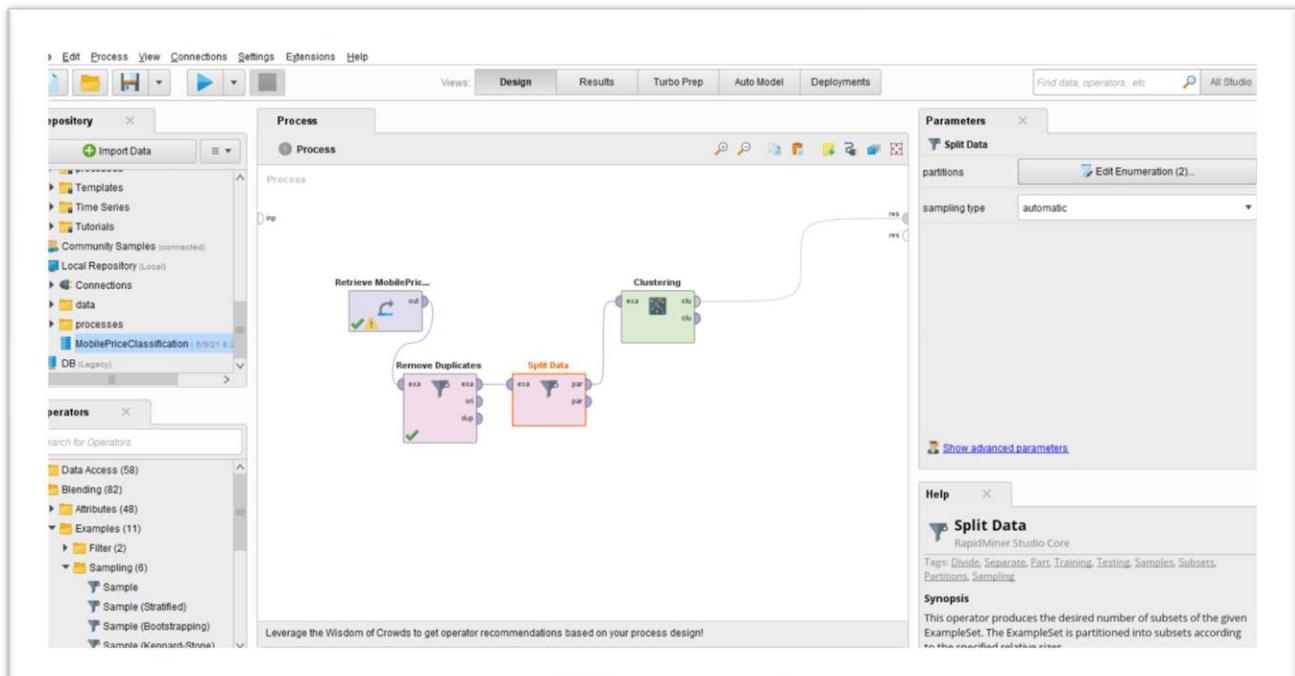
| Attribute       | cluster_0 | cluster_1 | cluster_2 | cluster_3 |
|-----------------|-----------|-----------|-----------|-----------|
| battery_power   | 1236.299  | 1232.995  | 1239.232  | 1250.719  |
| bluetooth       | 0.505     | 0.480     | 0.518     | 0.466     |
| clock_speed     | 1.505     | 1.559     | 1.533     | 1.463     |
| dual_sim        | 0.505     | 0.481     | 0.554     | 0.488     |
| front_camera    | 4.483     | 4.171     | 4.378     | 4.170     |
| four_g          | 0.533     | 0.519     | 0.528     | 0.497     |
| internal_memory | 31.671    | 31.119    | 32.974    | 32.642    |
| m_dep           | 0.483     | 0.494     | 0.504     | 0.540     |
| mobile_wt       | 138.230   | 140.724   | 141.091   | 140.988   |
| n_cores         | 4.615     | 4.531     | 4.446     | 4.491     |
| pc              | 10.335    | 9.641     | 10.065    | 9.503     |
| px_height       | 421.737   | 541.388   | 658.058   | 1155.997  |
| px_width        | 1029.098  | 1170.454  | 1296.460  | 1661.901  |
| ram             | 2322.591  | 842.398   | 3450.610  | 1756.747  |
| screen_height   | 12.214    | 12.242    | 12.376    | 12.441    |
| screen_width    | 5.826     | 5.542     | 5.904     | 5.836     |
| three_g         | 0.788     | 0.742     | 0.756     | 0.765     |

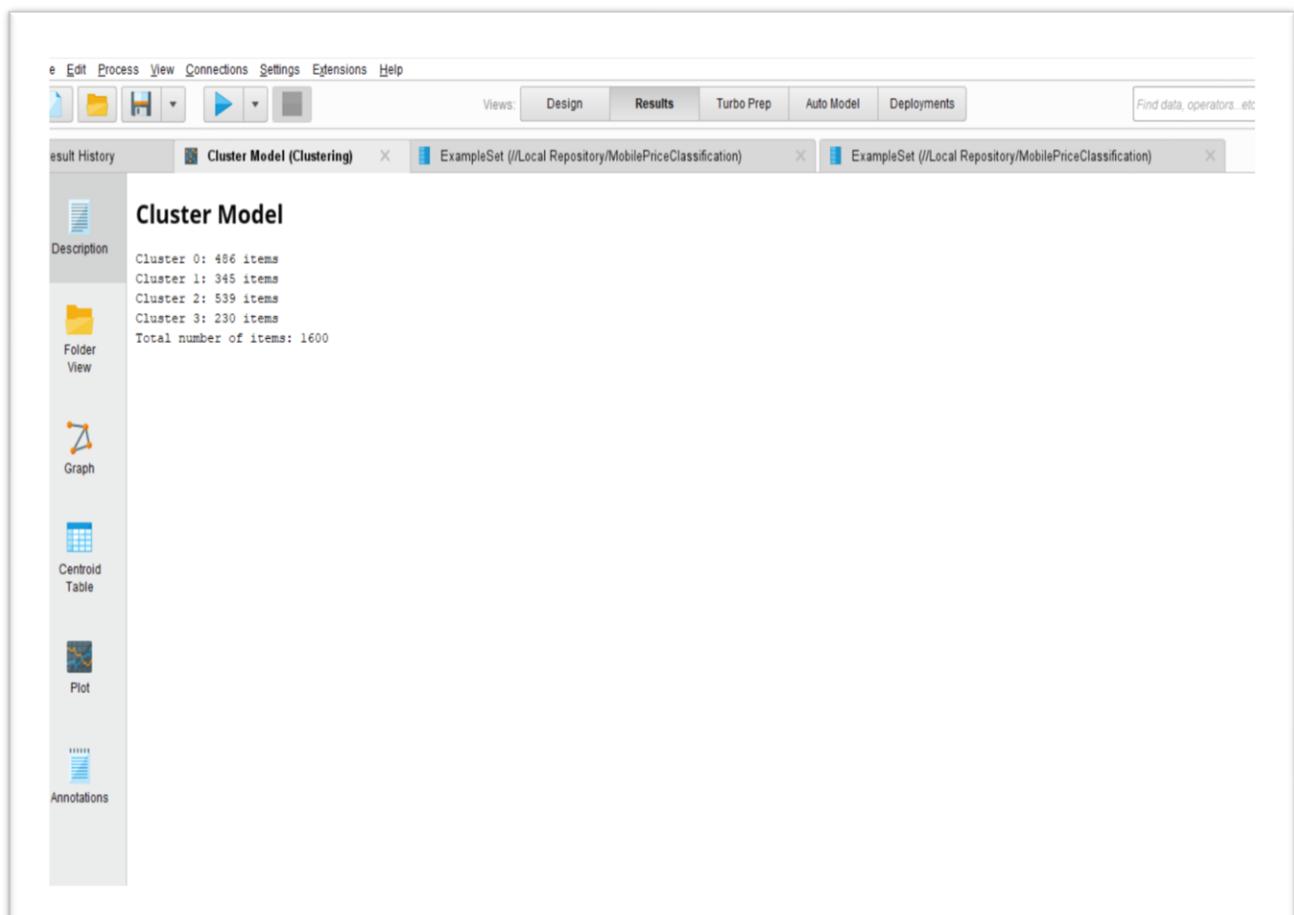
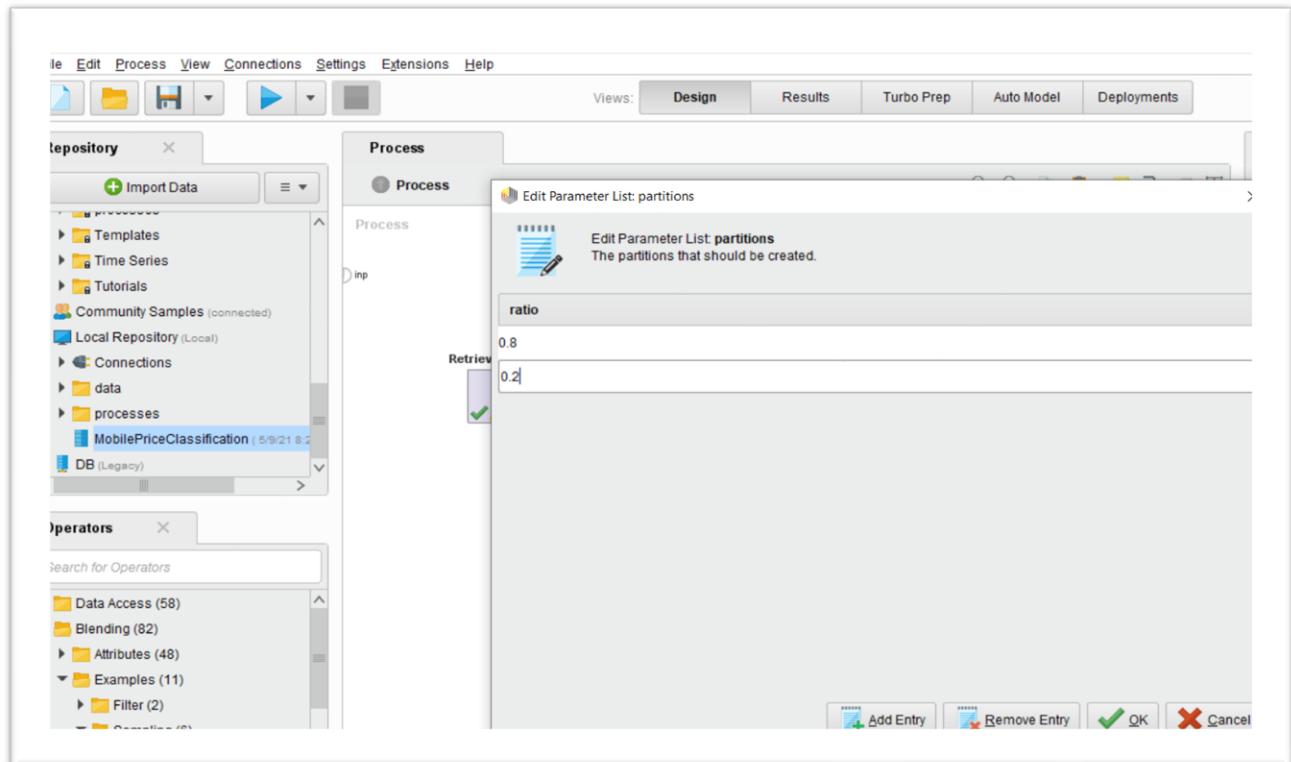
- **Data Replacement:** To process our dataset without any error, data replacement (a preprocessing step) can be done. For this step, knowledge of Regular Expression is required. But RapidMiner being a user-friendly platform helps with a list of rules for the same.



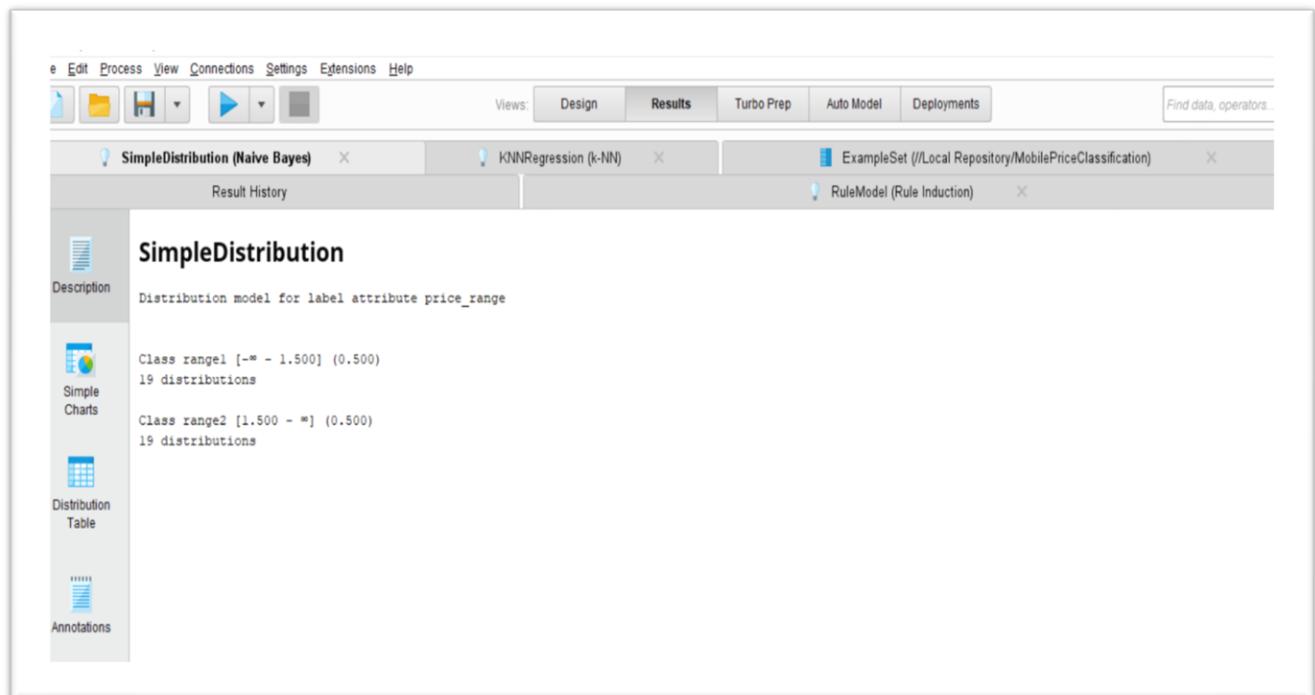
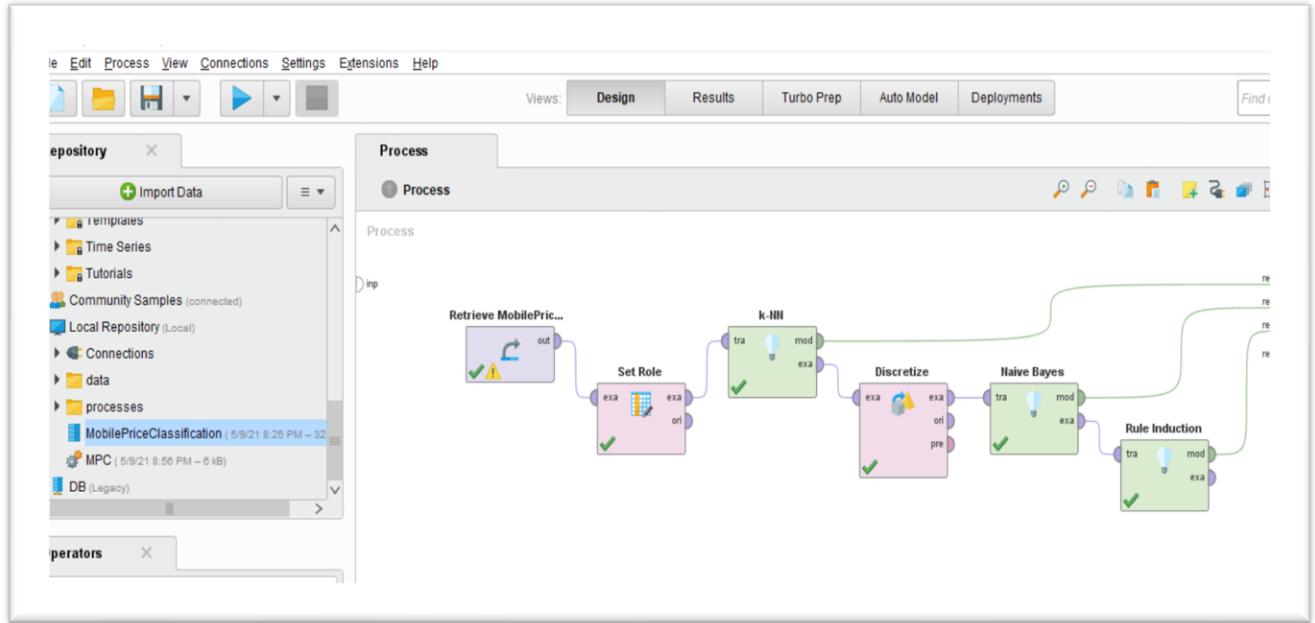


- **Data Partitioning:** For any modeling, we need separate training and testing dataset. Many times we lack the latter and hence split our dataset into the two. This is done using **Split data** operator where we first mention our training set ratio, followed by test-set ratio.





- **Automation and Process Control:** Automating important tasks are done using a variety of scripting languages. There are a number of Process Control operators to create workflows that perform any task repetitively and access system resources.
- **Data Access and Management:** RapidMiner Studio lets us access, load and analyze data (structured as well as unstructured). Unstructured data can be easily transformed to structured data.
- **Bayesian Modeling:** Bayesian modeling is used for calculating dependencies between features. In RapidMiner, this is done using the **Naive Bayes** operator. This operation is useful with classification problems.

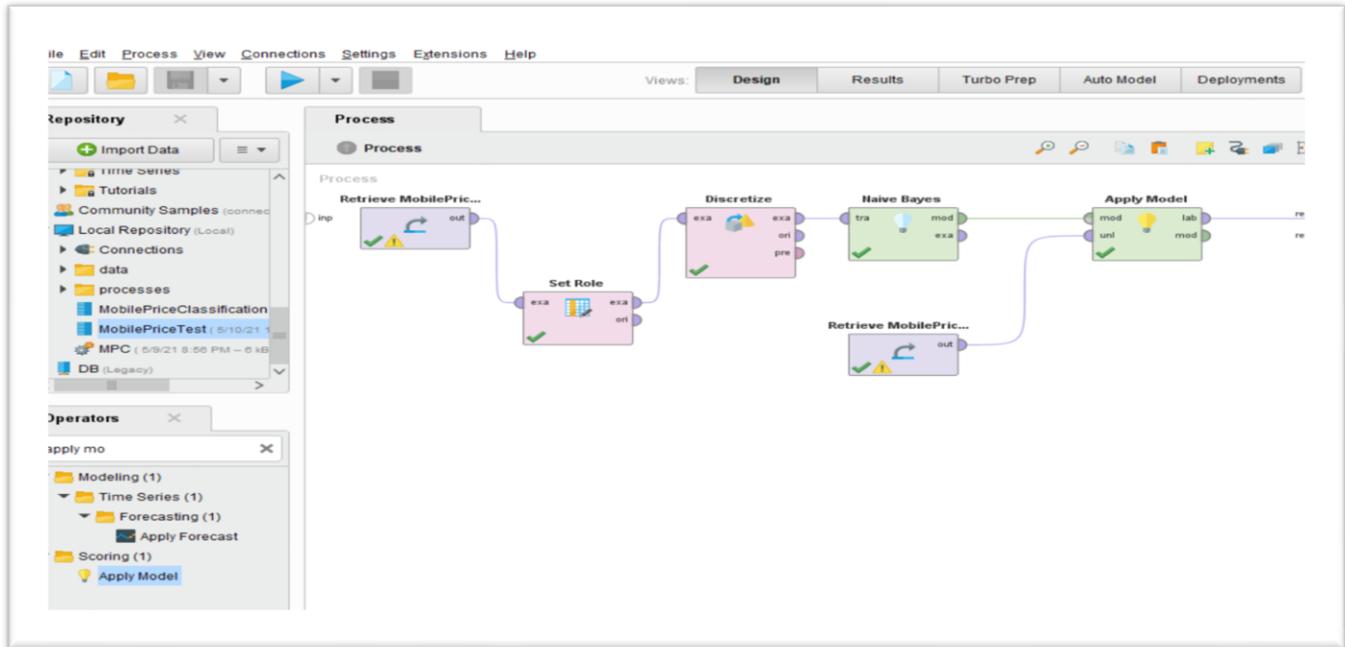


- **Data Exploration:** Data exploration is method which is similar to data analysis, where we uses visual exploration to understand what is in a dataset and the characteristics of the data, rather than through traditional data management systems.
- **Visual Workflow Designer:**
  - Speed up and automate the creation of predictive models in a drag + drop visual interface.
  - Rich library of 1,500+ algorithms and functions ensures the best model for any use case.
- **Descriptive Statistics:** The statistical overview of the dataset can be seen in just a few clicks. The Statistics button in the Result tab offers us information such as data type, number of missing values and other statistical information such as mean, median, mode etc.

| Name            | Type    | Missing | Statistics                             |
|-----------------|---------|---------|--|
| battery_power   | Integer | 0       | Min: 501, Max: 1998, Average: 1238.518 |
| bluetooth       | Integer | 0       | Min: 0, Max: 1, Average: 0.495         |
| clock_speed     | Real    | 0       | Min: 0.500, Max: 3, Average: 1.522     |
| dual_sim        | Integer | 0       | Min: 0, Max: 1, Average: 0.509         |
| front_camera    | Integer | 0       | Min: 0, Max: 19, Average: 4.309        |
| four_g          | Integer | 0       | Min: 0, Max: 1, Average: 0.521         |
| internal_memory | Integer | 0       | Min: 2, Max: 64, Average: 32.047       |
| m_dep           | Real    | 0       | Min: 0.100, Max: 1, Average: 0.502     |
| mobile_wt       | Integer | 0       | Min: 80, Max: 200, Average: 140.249    |

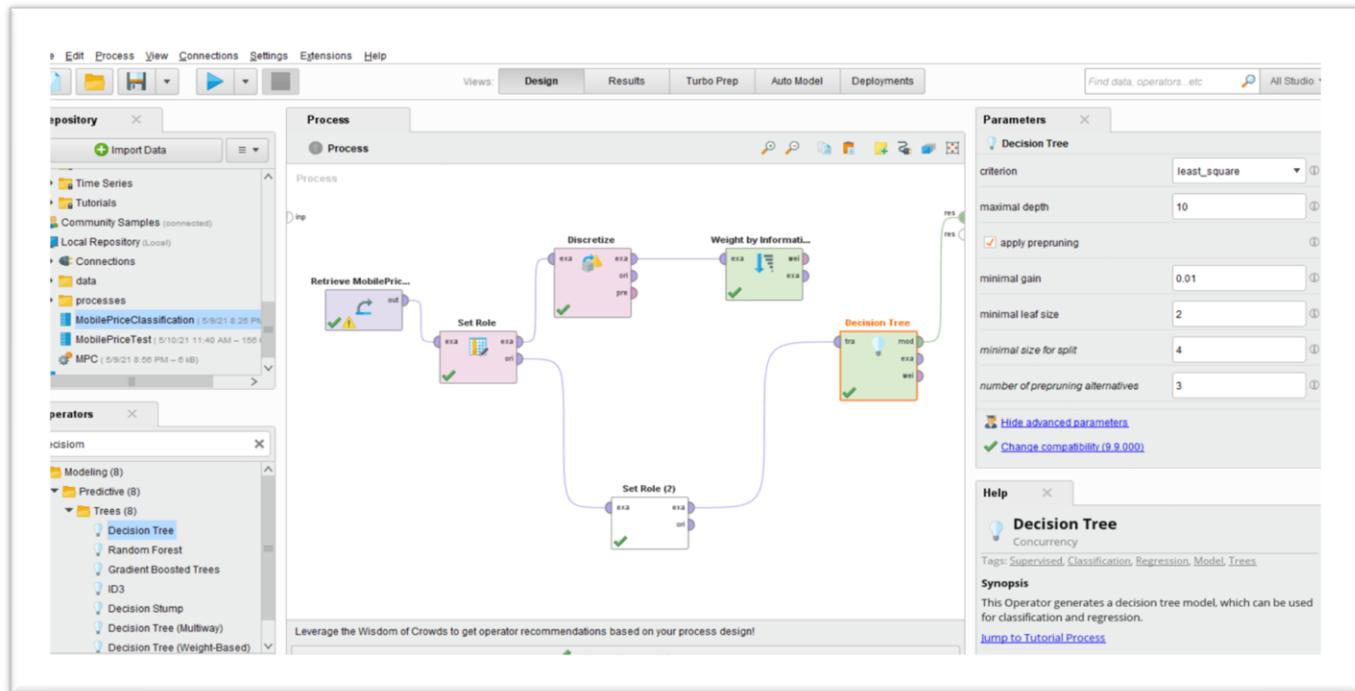
| Name          | Type    | Missing | Statistics |
|---------------|---------|---------|------------|
| battery_power | Integer | 0       | Min: 501   |
| bluetooth     | Integer | 0       | Min: 0     |
| clock_speed   | Real    | 0       | Min: 0.500 |
| dual_sim      | Integer | 0       | Min: 0     |
| front_camera  | Integer | 0       | Min: 0     |

- **Scoring:** Operator **Apply Model** is used to apply the trained model on another unlabelled dataset (testing set).

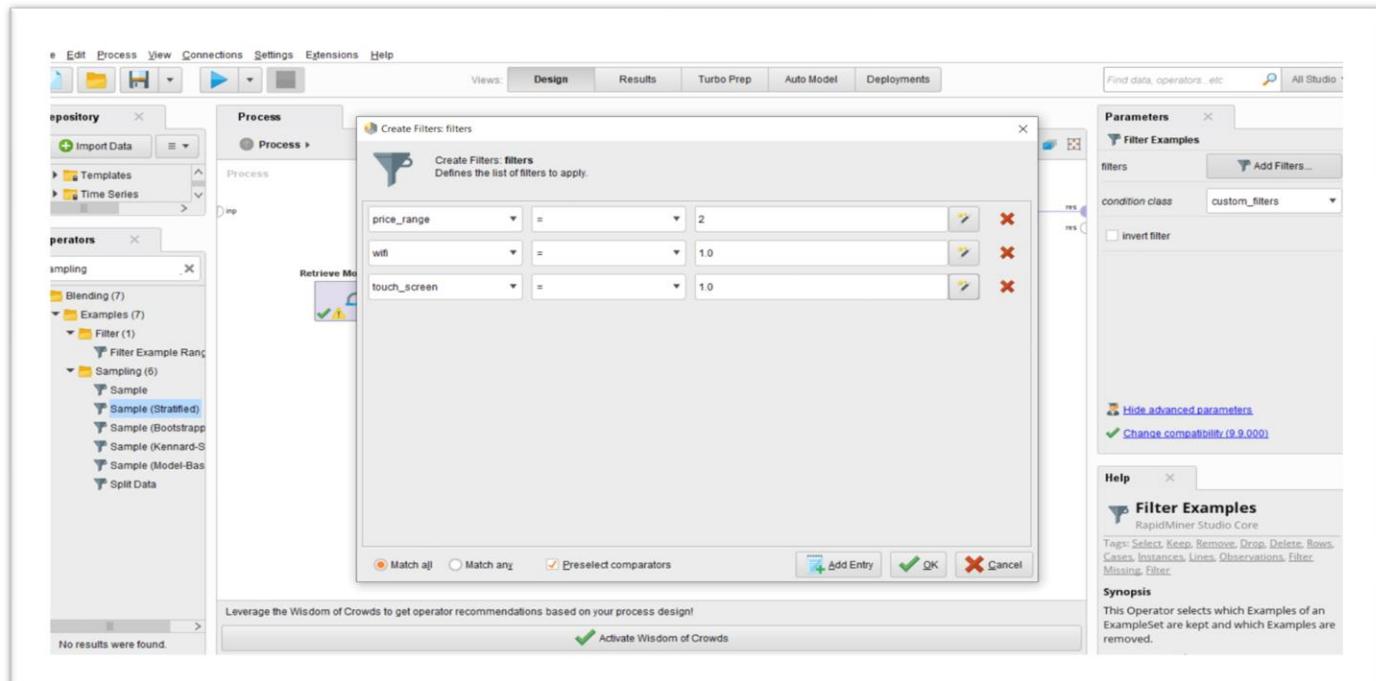


| Row No. | prediction(price_rang... | confidence(range1 [.. - 1.500]) | confidence(range2 [1.500 - ..]) | battery_pow... | bluetooth | clock_speed | dual_sim | front_camera |
|---------|--------------------------|---------------------------------|---------------------------------|----------------|-----------|-------------|----------|--------------|
| 1       | range2 [1.500 - ..]      | 0.003                           | 0.997                           | 1043           | 1         | 1.800       | 1        | 14           |
| 2       | range2 [1.500 - ..]      | 0.000                           | 1.000                           | 841            | 1         | 0.500       | 1        | 4            |
| 3       | range2 [1.500 - ..]      | 0.145                           | 0.855                           | 1807           | 1         | 2.800       | 0        | 1            |
| 4       | range2 [1.500 - ..]      | 0.000                           | 1.000                           | 1546           | 0         | 0.500       | 1        | 18           |
| 5       | range1 [.. - 1.500]      | 0.808                           | 0.192                           | 1434           | 0         | 1.400       | 0        | 11           |
| 6       | range2 [1.500 - ..]      | 0.002                           | 0.998                           | 1464           | 1         | 2.900       | 1        | 5            |
| 7       | range2 [1.500 - ..]      | 0.000                           | 1.000                           | 1718           | 0         | 2.400       | 0        | 1            |
| 8       | range1 [.. - 1.500]      | 0.945                           | 0.055                           | 833            | 0         | 2.400       | 1        | 0            |
| 9       | range2 [1.500 - ..]      | 0.002                           | 0.998                           | 1111           | 1         | 2.900       | 1        | 9            |
| 10      | range1 [.. - 1.500]      | 0.999                           | 0.001                           | 1520           | 0         | 0.500       | 0        | 1            |
| 11      | range2 [1.500 - ..]      | 0.000                           | 1.000                           | 1500           | 0         | 2.200       | 0        | 2            |
| 12      | range2 [1.500 - ..]      | 0.000                           | 1.000                           | 1343           | 0         | 2.900       | 0        | 2            |
| 13      | range1 [.. - 1.500]      | 1.000                           | 0.000                           | 900            | 1         | 1.400       | 1        | 0            |
| 14      | range1 [.. - 1.500]      | 0.994                           | 0.006                           | 1190           | 1         | 2.200       | 1        | 5            |
| 15      | range2 [1.500 - ..]      | 0.046                           | 0.954                           | 630            | 0         | 1.800       | 0        | 8            |
| 16      | range1 [.. - 1.500]      | 0.998                           | 0.002                           | 1846           | 1         | 1           | 0        | 5            |
| 17      | range2 [1.500 - ..]      | 0.443                           | 0.557                           | 1985           | 0         | 0.500       | 1        | 14           |
| 18      | range1 [.. - 1.500]      | 0.605                           | 0.395                           | 4022           | 0         | 2.000       | 0        | 6            |

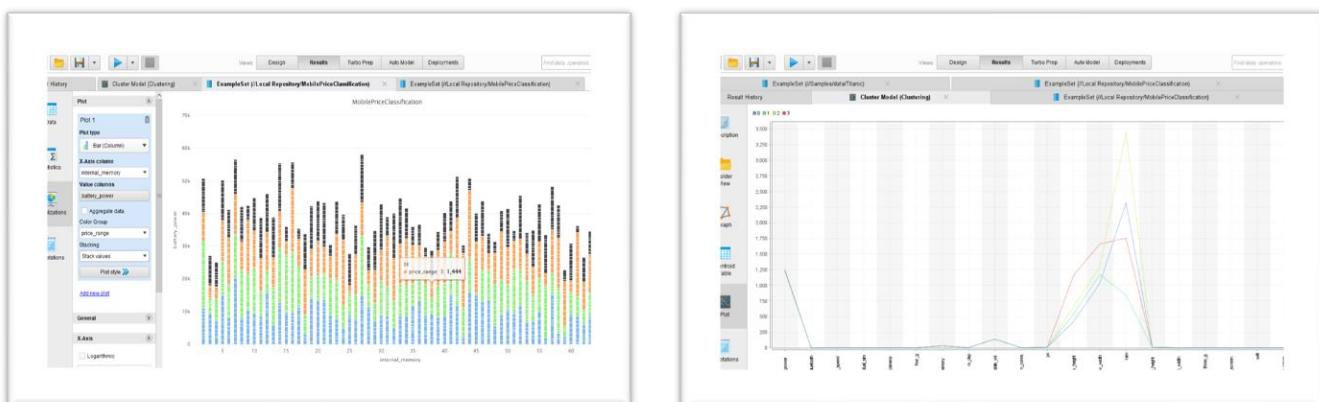
- **Weighting and Selection:** This method uses a variety of operators through which data is selected according to the weight as per the mentioned criteria.



- **Market Basket Analysis:** RapidMiner uses **FP-Growth** operator to create frequent item-set which is used by **Create Association Rule** operator and then finally these rules are applied to our model using **Apply Association Rule** operator.
- **Data Sampling:** To reduce the cost of data collection, data sampling is done, i.e., selecting a subset from the same dataset and then manipulating and applying algorithms to it. This is done using **Sample** operation in RapidMiner.



- **Similarity Calculation:** Calculating similarity and dissimilarity is yet another important step for data mining (used in Recommendation Systems) since it enables us to find how equal two or more vectors are between users.
- **Modeling Evaluation:** Every machine learning algorithm has to be evaluated for its future implementation. With **Performance** operators, this step can be completed in a few seconds.
- **Graphs and Visualization:** Visual representation is the best way to start our data analysis with. RapidMiner provides a variety of plot options and the user-friendly interface makes it easier to plot them.



## Turbo-prep and Auto-model:

All the above operations performed takes up time and effort, i.e., preprocessing, splitting, applying models etc. To save our time, we can use Turbo-prep and Auto-model to speed up the process and apply multiple algorithms at once.

| m_dep<br>Number | mobile_wt<br>Number | n_cores<br>Number | pc<br>Number | px_height<br>Number | px_width<br>Number | ram<br>Number | screen_height<br>Number | screen_width<br>Number | three_g<br>Number | touch_screen<br>Number | wifi<br>Number | price_range<br>Number |
|-----------------|---------------------|-------------------|--------------|---------------------|--------------------|---------------|-------------------------|------------------------|-------------------|------------------------|----------------|-----------------------|
| 0.600           | 188                 | 2                 | 2            | 20                  | 756                | 2549          | 9                       | 7                      | 0                 | 0                      | 1              | 1                     |
| 0.700           | 136                 | 3                 | 6            | 905                 | 1988               | 2631          | 17                      | 3                      | 1                 | 1                      | 0              | 2                     |
| 0.900           | 145                 | 5                 | 6            | 1263                | 1716               | 2603          | 11                      | 2                      | 1                 | 1                      | 0              | 2                     |
| 0.800           | 131                 | 6                 | 9            | 1216                | 1786               | 2769          | 16                      | 8                      | 1                 | 0                      | 0              | 2                     |
| 0.600           | 141                 | 2                 | 14           | 1208                | 1212               | 1411          | 8                       | 2                      | 1                 | 1                      | 0              | 1                     |
| 0.700           | 164                 | 1                 | 7            | 1004                | 1654               | 1057          | 17                      | 1                      | 1                 | 0                      | 0              | 1                     |
| 0.800           | 139                 | 8                 | 10           | 381                 | 1018               | 3220          | 13                      | 8                      | 1                 | 0                      | 1              | 3                     |
| 0.800           | 187                 | 4                 | 0            | 512                 | 1149               | 700           | 16                      | 3                      | 1                 | 1                      | 1              | 0                     |
| 0.700           | 174                 | 7                 | 14           | 386                 | 836                | 1099          | 17                      | 1                      | 1                 | 0                      | 0              | 0                     |
| 0.100           | 93                  | 5                 | 15           | 1137                | 1224               | 513           | 19                      | 10                     | 1                 | 0                      | 0              | 0                     |
| 0.100           | 102                 | 6                 | 1            | 240                 | 274                | 2046          | 6                       | 2                      | 0                 | 0                      | 0              | 2                     |

| Selected                            | Status ↑                              | Quality   | Name          | Correlation | ID-ness | Stability | Missing | Text-ness |
|-------------------------------------|---------------------------------------|---|---------------|-------------|---------|-----------|---------|-----------|
| <input checked="" type="checkbox"/> | <span style="color: yellow;">●</span> | <span style="color: orange;">█</span><br>Correlation: 50.81%<br>ID-ness: 78.10%<br>Stability: 0.20%<br>Missing: 0.00%<br>Text-ness: 0.00% | ram           | 50.81%      | 78.10%  | 0.20%     | 0.00%   | 0.00%     |
| <input checked="" type="checkbox"/> | <span style="color: green;">●</span>  | <span style="color: orange;">█</span>   | battery_power | 3.46%       | 54.70%  | 0.30%     | 0.00%   | 0.00%     |
| <input checked="" type="checkbox"/> | <span style="color: green;">●</span>  | <span style="color: grey;">█</span>   | bluetooth     | 0.07%       | 0.10%   | 50.50%    | 0.00%   | 0.00%     |
| <input checked="" type="checkbox"/> | <span style="color: green;">●</span>  | <span style="color: grey;">█</span>   | clock_speed   | 0.06%       | ?       | 20.65%    | 0.00%   | 0.00%     |
| <input checked="" type="checkbox"/> | <span style="color: green;">●</span>  | <span style="color: grey;">█</span>   | dual_sim      | 0.06%       | 0.10%   | 50.95%    | 0.00%   | 0.00%     |

**Execution**

Execute on: Local Computer (this machine)

Queue: No queues available

Select Folder for Storing Results

The results of this run will be stored in the folder selected below. We recommend to use an empty folder in the selected AI Hub repository.

- Local Repository (Local)
- Temporary Repository (Local)

**Models**

Naive Bayes (selected)

Generalized Linear Model

Use Regularization (checked)

Calculate p-values (unchecked)

Logistic Regression (unchecked)

Fast Large Margin (unchecked)

Automatically Optimize (checked)

Deep Learning

Decision Tree

Automatically Optimize (checked) Maximal Depth: 20

Random Forest

Automatically Optimize (checked) Number of Trees: 20 Maximal Depth: 20

Gradient Boosted Trees

**Data Preparation**

Remove Columns with Too Many Values (unchecked)

Maximum Number of Values: 50

Extract Date Information (unchecked)

Extract Text Information (unchecked)

Select Text Columns (0)

Number of Extracted Features: 1,000

Automatic Feature Selection (unchecked)

Additional Minutes (Maximum): 60

Final Feature Set should be: Accurate

Automatic Feature Generation (unchecked)

Function Complexity can be: Medium

**Column Analysis**

**Results**

Comparison Overview (selected)

Naive Bayes

- Model
- Weights
- Simulator
- Performance
- Predictions
- Production Model

Generalized Linear Model

Deep Learning

Decision Tree

Random Forest

Gradient Boost.

Support Vect.

**Overview**

Classification Error

Number of Models: 2

Runtimes (ms)

| Model                  | Classification Error | Standard Deviation | Gains             | Total Time        | Training Time (1,000 ...) | Scoring Time |
|------------------------|----------------------|--------------------|-------------------|-------------------|---------------------------|--------------|
| Deep Learning          | No results yet...    | No results yet...  | No results yet... | No results yet... | No results yet...         | No result    |
| Decision Tree          | No results yet...    | No results yet...  | No results yet... | No results yet... | No results yet...         | No result    |
| Random Forest          | No results yet...    | No results yet...  | No results yet... | No results yet... | No results yet...         | No result    |
| Gradient Boosted Trees | No results yet...    | No results yet...  | No results yet... | No results yet... | No results yet...         | No result    |
| Support Vector Machine | No results yet...    | No results yet...  | No results yet... | No results yet... | No results yet...         | No result    |

## 4. REPORTING AND BUSINESS INTELLIGENCE:-

### RapidMiner Radoop:

It is the predictive analysis tool used by RapidMiner, that governs group of business intelligence tools created to empower organizations with data. This tool is slightly more complex than the other tools in the market and can require technical knowledge. Also, there are many different resources available on their website for those who want to get start with predictive analysis. It uses Hadoop(open-source Java-based programming framework) which is used to convert intelligence into workflows and also enables the user to transfer through big data.

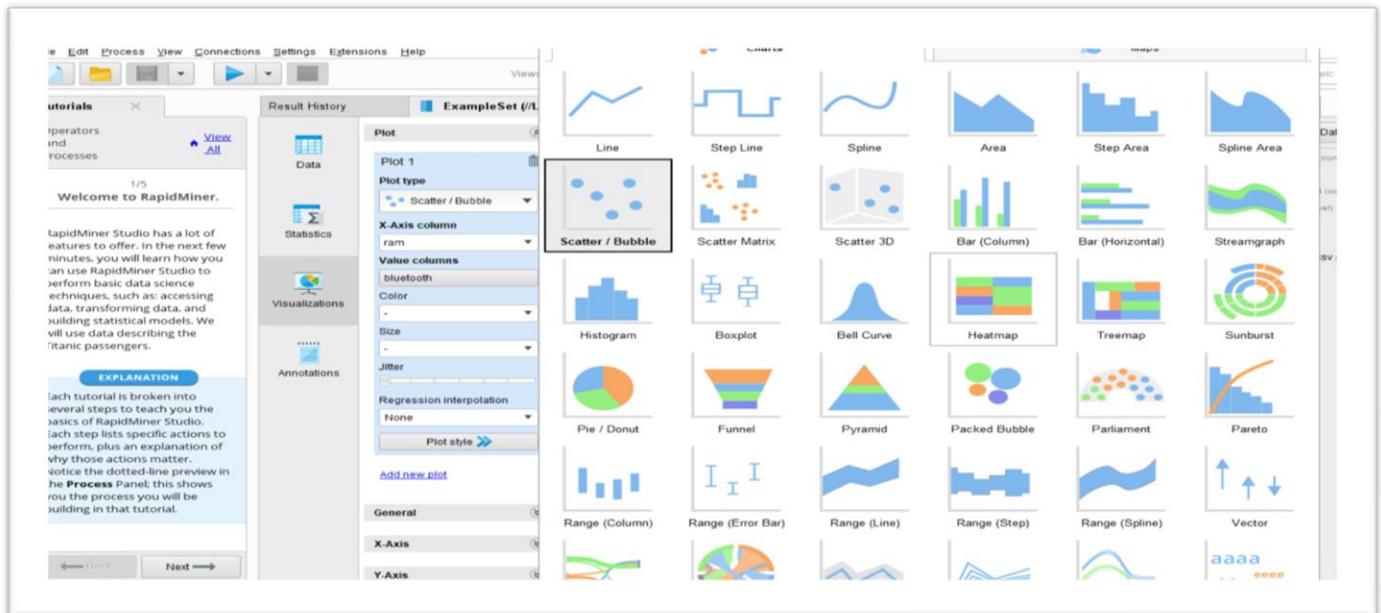
RapidMiner impart software, solutions & services in the department of predictive analytics, data mining & text mining. We eventually and intelligently analyze data (both structured and unstructured) also includes multimedia and text on a great scale.

We can also create model in RapidMiner and use a server module for us to do the Web services, and this Web services can get integrated as a source for Qlik. this also supports bidirectional integration also. lets illustrate with an example, if we make changes in the modeling parameters in Qlik, We can have the results back almost real-time.

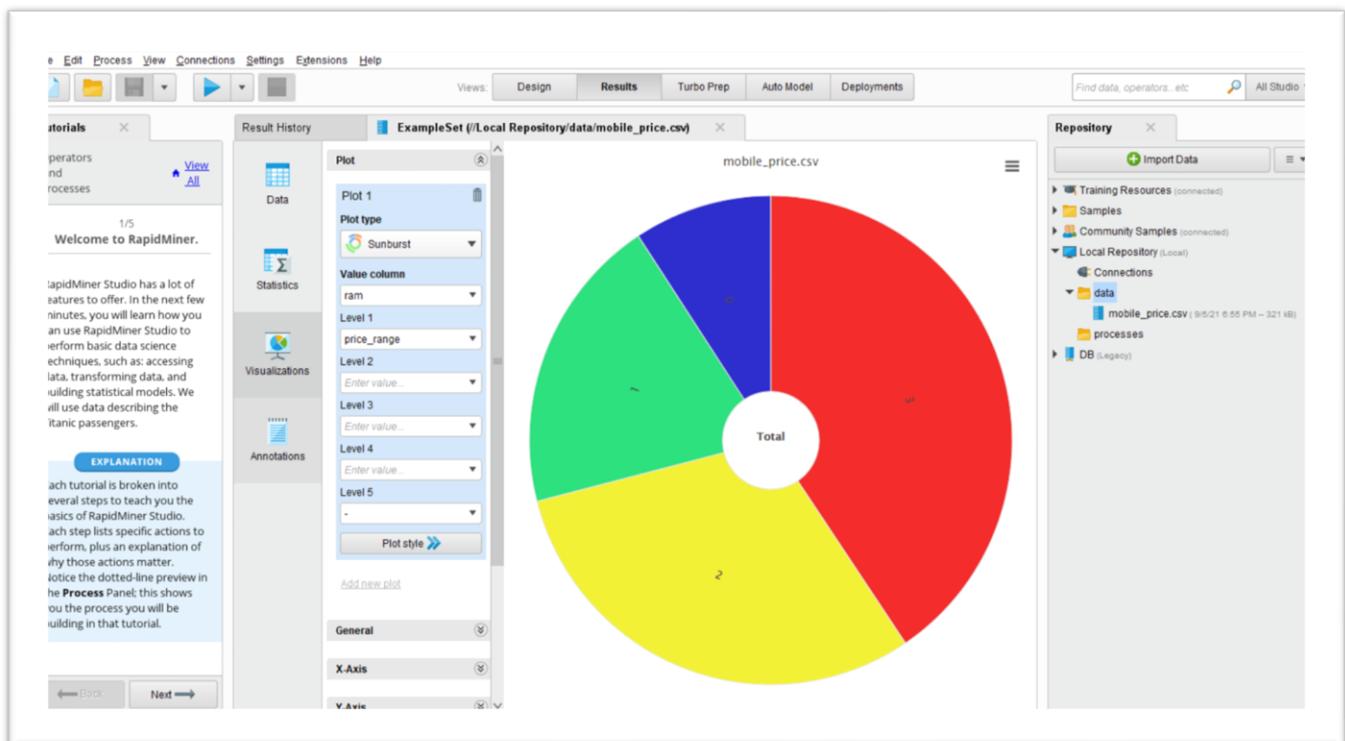
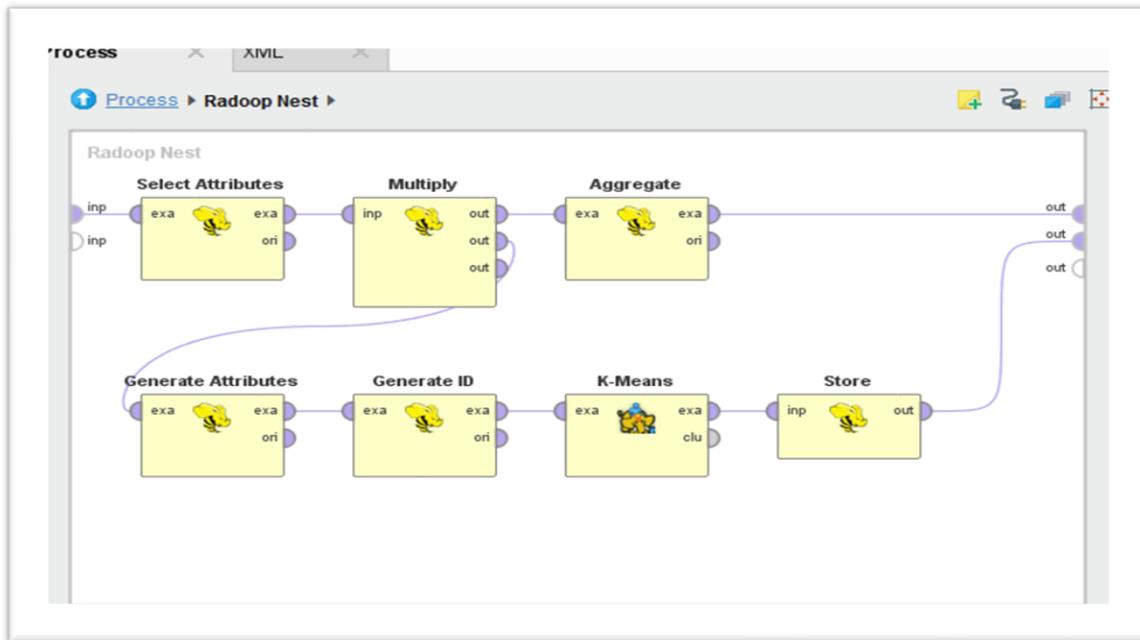
“RapidMiner combine decently by Tableau, Qlik

RapidMiner brings data mining and machine learning steps that also including: data loading and transformation (ETL), data preprocessing and visualization, predictive analytics and statistical modeling, evaluation and deployment.

RapidMiner Studio provides data analysis and merges along with custom business deployment. This code-optional software supports automate reporting on the basis of time intervals else having events that trigger changes in the visualizations.



RapidMiner is a product that is used by quarter million users for the use of data science . In the most datascience projects the final mile is to deliver the results to a broader audience, the audience that can act on based on predictions. In RapidMiner we call this operationalization of Predictive analytics. As part of operationalization a very common practice is delivering the results via Dashboards, Reports or Visualizations.



## **1) Rapidminer Web Apps**

RapidMiner also assigned dashboard, reporting and Business Intelligence functionalities that can be performed with RapidMiner workflows to create a really optimum alternative. Advanced analytics solutions can be distributed in browser based on richHTML5 based solution. All of this can be performed using configurations and can be customized in accordance with advanced Javascript and CSS control.

## **2) Integration with Tableau**

Rapidminer has the below three ways to combine with Tableau.

- a) **Tableau Data Extract** One way is use the "Tableau Data Writer" operator, this method supports creating Tableau Data extract file using RapidMiner workflows. these processes can be scheduled to push updates on Tableau data files thereby updating the Dashboards.
- b) **Odata Service** : Tableau give ability of reading data from an OData webservice. RapidMiner server ability of displaying RapidMiner workflows as a webservice and provide data using OData format.
- c) **Tableau Web Data Connector** : Tableau provides flexibility with Web Data Connector. If this data source is preferred you can still use the webservices exposed from RapidMiner , by handling the data with web data connector API.

## **3) Integration with Qlikview & QlikSense**

RapidMiner provides atleast two ways to integrate with Qlikview.

- a) **QVX (Qlikview Extract)** - This methods allows writing data out to Qlikview extract formats that can be used in both Qlikview & QlikSense applications. Such process can then be scheduled to push updates to Qlik Extract files, thereby uploading the dashboards.
- b) **Web Service - RapidMiner** provides ability to expose any processes as a webservice. Each webservices can return data in various formats including XML, JSON, HTML table etc. Using this technique we can pull data directly from RapidMiner server based process into a Qlikview or QlikSense dashboard.

## **4) Other Applications**

If webservices are not possible to be invoked from your BI application, you can still rely on RapidMiner to write its output to databases or files as needed. Your BI Applications can then get the data from these intermediate stores as needed.

## 5. PRICING:

- **Rapidminer Enterprise Pricing:**

Pricing: plan created specific to teams needs.

- **Rapidminer Go:**

Pricing: \$10 per month

- **Rapidminer Studio (free version):**

Pricing: Free

- **Pricing plans:**

Free: \$0

Small: \$2,500/user/year

Medium: \$5,000/user/year

Large: \$10,000/user/year

## **6. CONCLUSION**

As Rapid Miner provides an integrated environment. It supports all steps of Data mining like data filtering, merging, joining, aggregation. It helps to build, train and validate predictive models. It runs on every major platform and operating system. It has wide range of applications in our daily life. Graphical design environment makes it simple and fast to design better models.

## **7. INDIVIDUAL CONTRIBUTION**

### **Aashi Barodiya(180009)**

As, she worked on Rapidminer Evolution and also on dashboard working. Appreciates team work and efficiently worked on this project.

### **Muskan Tayal (180068)**

She worked on the working of RapidMiner and benefits of RapidMiner. She has been a keen learner and co-operative throughout.

### **Shreya Sharma(180123)**

She worked on the Introduction part and also done the editing part. The report was depicted with an information script and was planned and led before time.

### **Prashi Jain(180186)**

She efficiently worked on Reporting and operators. She also worked on the impact in Business intelligence through RapidMiner.

### **Harshita Das(180043)**

She worked on presentation and the features of RapidMiner. Her decision nature ceased us to a calculative mathematical problem which requires a model for optimal output and can be used for daily life problems.

### **Priyanka Motiani(180093)**

She worked on features of RapidMiner and also worked on presentation. Her logical and analytical skills supported us.

### **Akshita Kanther(180161)**

She worked on TurboPrep like how to use turboprep in RapidMiner and also she worked on the visualizations part. Her time particular definition to work drafted the project before time.

## **8. REFERENCES**

- [1] RapidMiner . Accessed on: 10 May, 2021.[Online]. Available: [rapidminer.com](http://rapidminer.com).
- [2] Vijay Kotu and Bala Deshpande, Data Science, 2nd edition. Accessed on: 10 May, 2021. [Online]. Available : <https://www.sciencedirect.com/science/article/pii/B9780128147610000150>
- [3] [RapidMiner Embraces its Community and Open Source Culture Delivering Get-More-Open-Core Predictive Analytics, September 1, 2015.](#)