Movies Recommendations Based on Collaborative Filtering.

ROLL NO	NAME
BCA-30	Kshiteeja Jadhav
BCA-69	Jyoti
BCA-70	Ajinkya Gaikwad

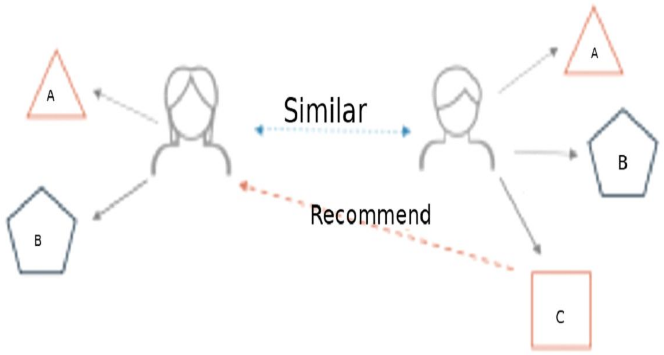
What is a recommender system?

A recommender system is a simple algorithm whose aim is to provide the most relevant information to a user by discovering patterns in a dataset. The algorithm rates the items and shows the user the items that they would rate highly. An example of recommendation in action is when you visit Amazon and you notice that some items are being recommended to you or when Netflix recommends certain movies to you. They are also used by Music streaming applications such as Spotify and Deezer to recommend music that you might like.

Below is a very simple illustration of how recommender systems work in the context of an e-commerce site.

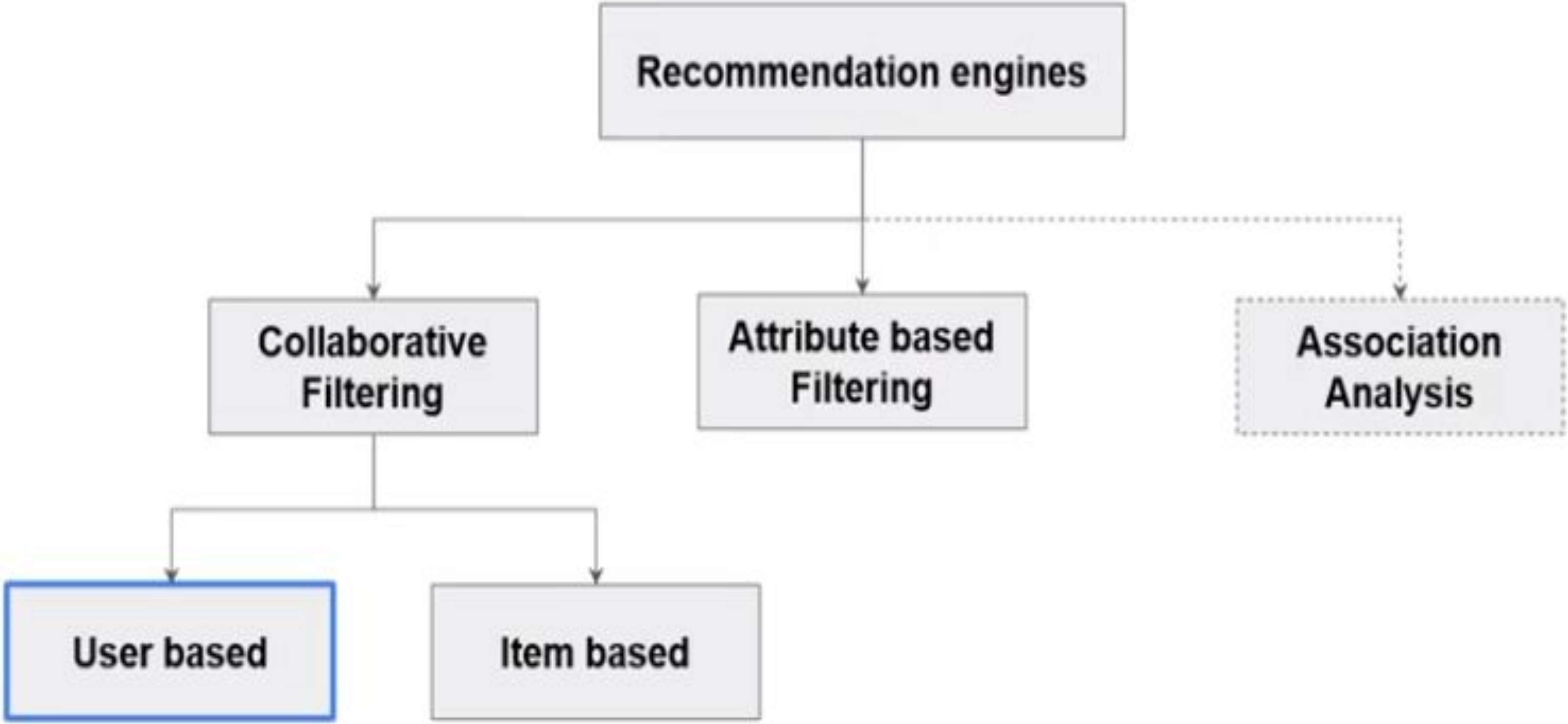
What is a recommender system?

Two users buy the same items A and B from an e-commerce store. When this happens the similarity index of these two users is computed. Depending on the score the system can recommend item C to the other user because it detects that those two users are similar in terms of the items they purchase.



Different types of recommendation engines.

The most common types of recommendation systems are content-based and collaborative filtering recommender systems. In collaborative filtering, the behavior of a group of users is used to make recommendations to other users. The recommendation is based on the preference of other users. A simple example would be recommending a movie to a user based on the fact that their friend liked the movie. There are two types of collaborative models Memory-based methods and Model-based methods. The advantage of memory-based techniques is that they are simple to implement and the resulting recommendations are often easy to explain. They are divided into two:


Recommendation engines
Collaborative Filtering
Attribute based Filtering
Association Analysis
User based
Item based

Different types of recommendation engines:

User-based collaborative filtering: In this model, products are recommended to a user based on the fact that the products have been liked by users similar to the user. For example, if Derrick and Dennis like the same movies and a new movie come out that Derick like, then we can recommend that movie to Dennis because Derrick and Dennis seem to like the same movies.

•**Item-based collaborative filtering**: These systems identify similar items based on users' previous ratings. For example, if users A, B, and C gave a 5-star rating to books X and Y then when a user D buys book Y they also get a recommendation to purchase book X because the system identifies book X and Y as similar based on the ratings of users A, B, and C.

Different types of recommendation engines:

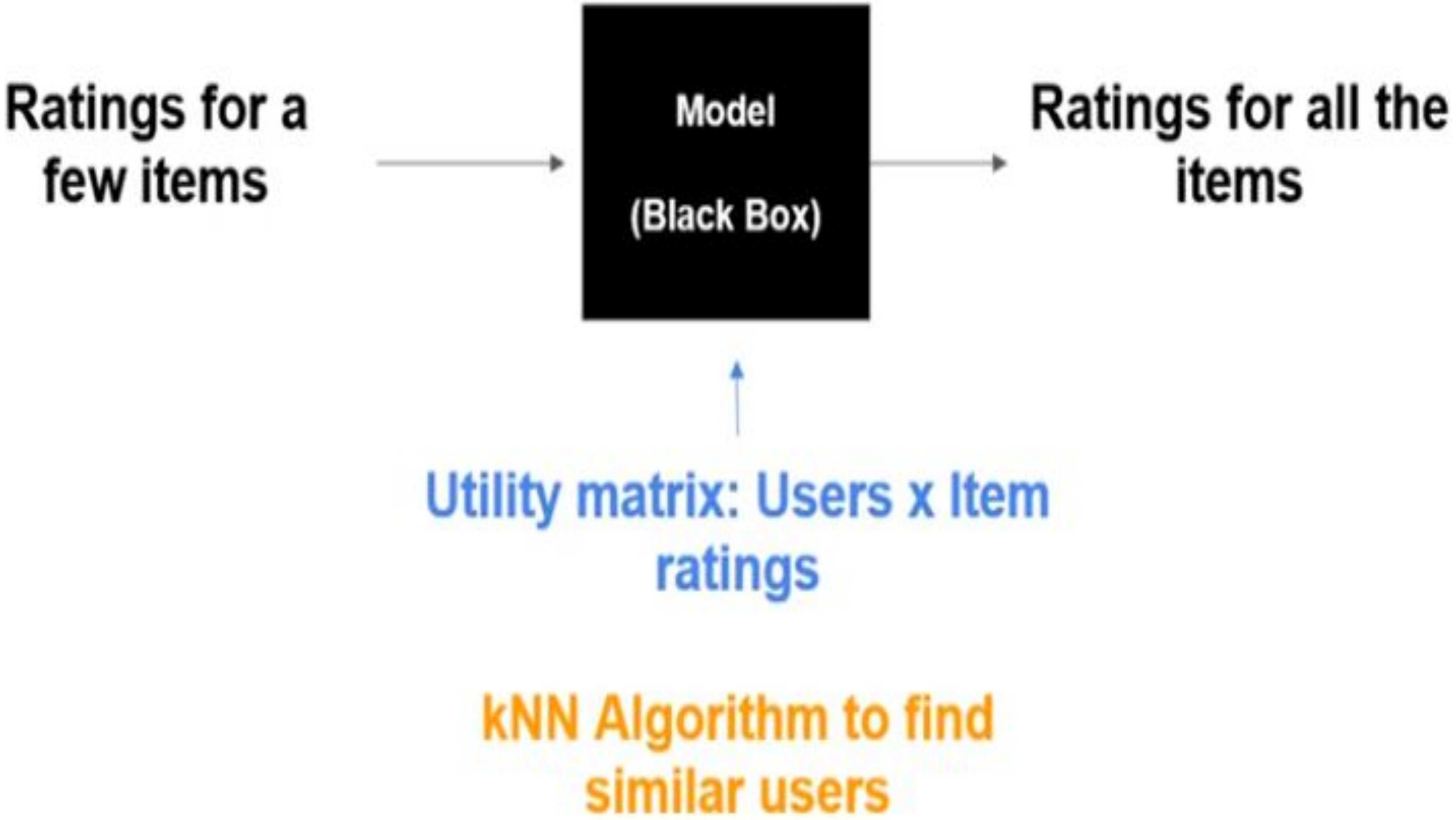
Content-based systems use metadata such as genre, producer, actor, musician to recommend items say movies or music. Such a recommendation would be for instance recommending Infinity War that featured Vin Diesel because someone watched and liked The Fate of the Furious. Similarly, you can get music recommendations from certain artists because you liked their music. Content-based systems are based on the idea that if you liked a certain item you are most likely to like something that is similar to it.

Datasets Used:

We are going to use the Movie Lens Data Set. This dataset was put together by the Group lens research group at the University of Minnesota. It contains 1, 10, and 20 million ratings. Movie lens also has a website where you can sign up, contribute reviews and get movie recommendations.

The Datasets are as follows:

- movies.csv
  - movieid
  - title
  - genre
- ratings.csv
  - userId
  - movieId
  - Ratings
  - timestamp



Ratings for a few items

Model

(Black Box)

Ratings for all the items

Utility matrix: Users x Item ratings

kNN Algorithm to find similar users

K-Nearest Neighbours Algorithm:

In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method proposed by Thomas Cover used for classification and regression.[1] In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

In k-NN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

K-Nearest Neighbours Algorithm:

In k-NN regression, the output is the property value for the object. This value is the average of the values of k nearest neighbors.

k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until function evaluation. Since this algorithm relies on distance for classification, normalizing the training data can improve its accuracy dramatically.

Dataset

MovieLens
100,000 ratings from 1,000 users and 1,700 movies

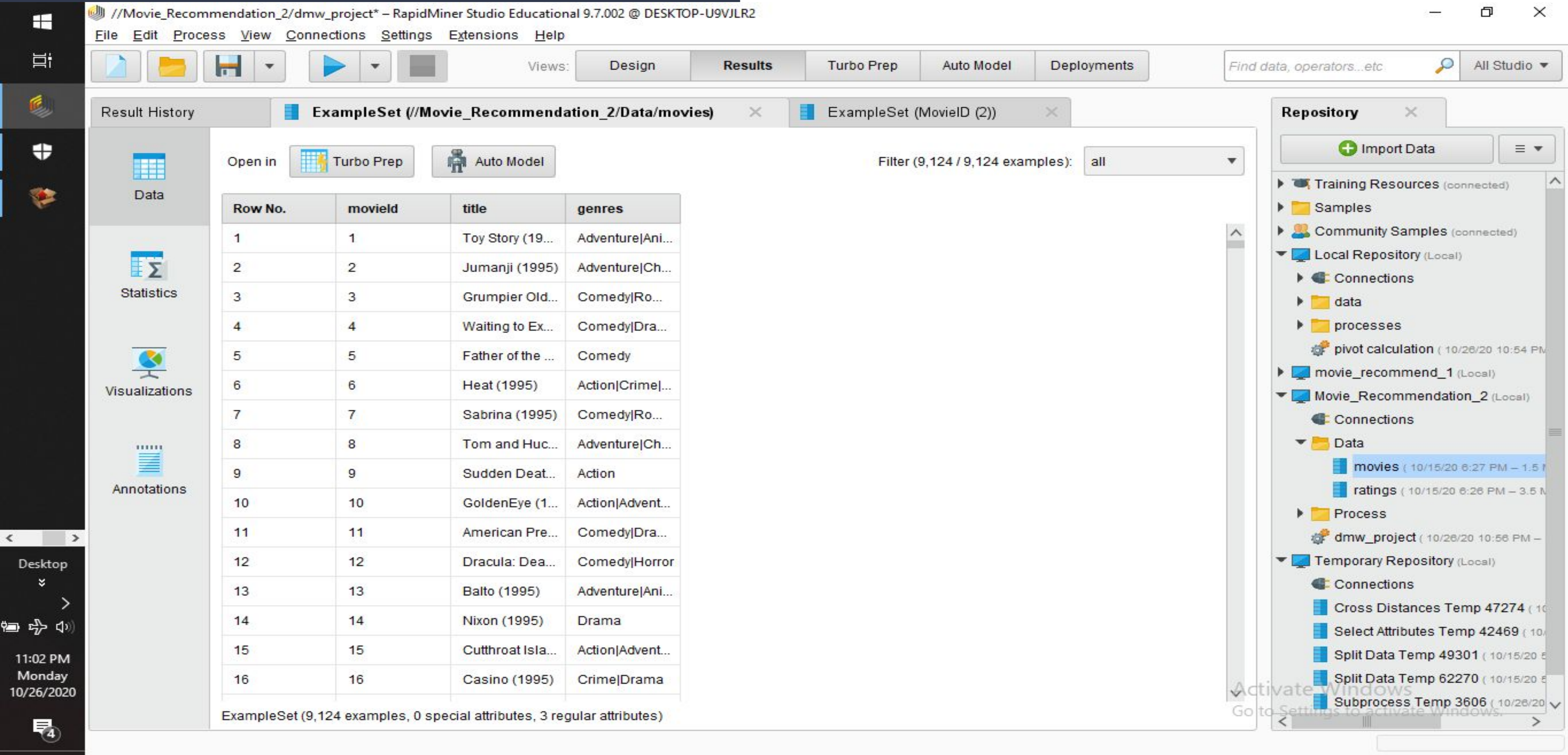
Parameters

k in kNN:
80

Tool

rapidminer

Movies.csv:



//Movie_Recommendation_2/dmw_project* – RapidMiner Studio Educational 9.7.002 @ DESKTOP-U9VJLR2

ExampleSet (//Movie_Recommendation_2/Data/movies)

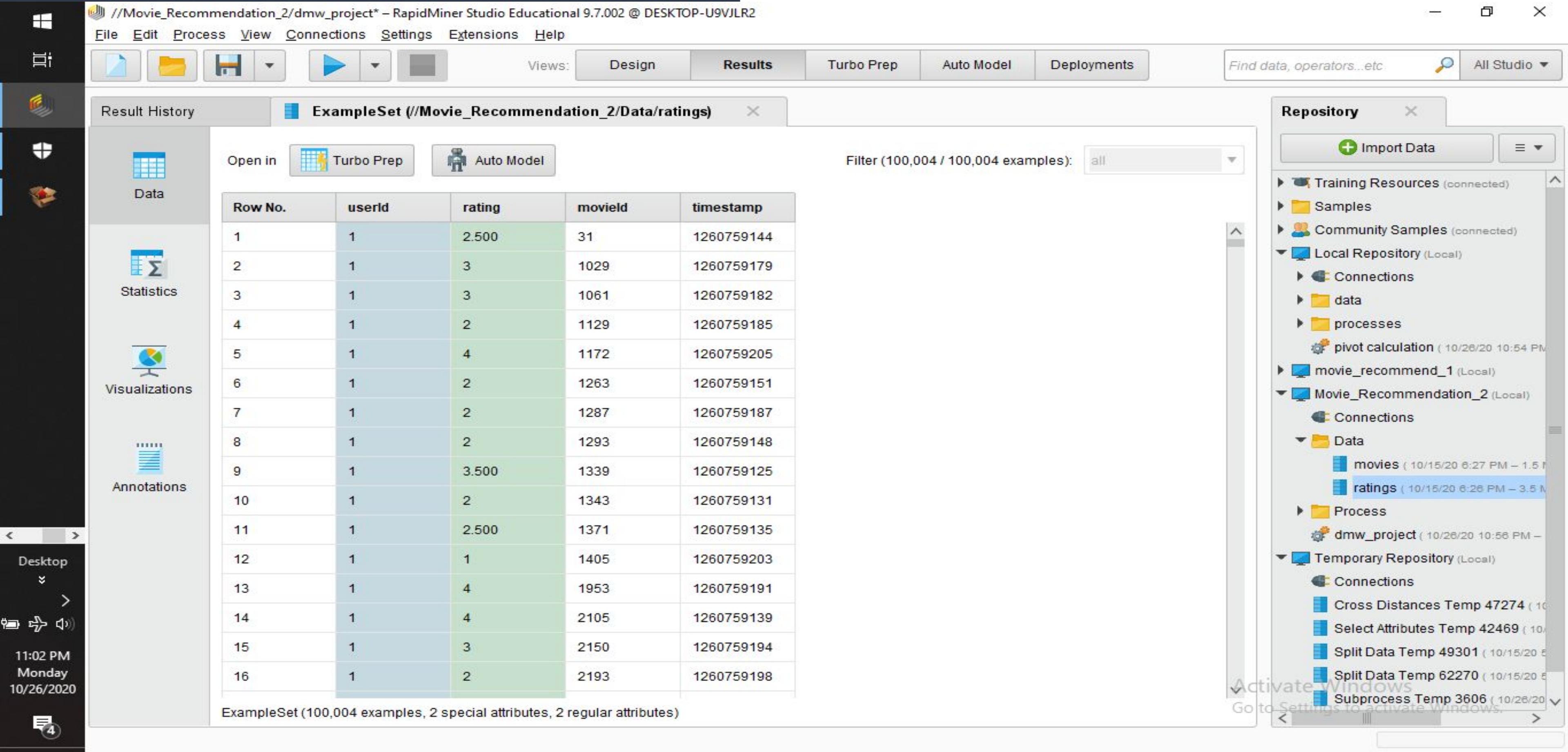
Filter (9,124 / 9,124 examples): all

Row No.	movieId	title	genres
1	1	Toy Story (19...	Adventure|Ani...
2	2	Jumanji (1995)	Adventure|Ch...
3	3	Grumpier Old...	Comedy|Ro...
4	4	Waiting to Ex...	Comedy|Dra...
5	5	Father of the ...	Comedy
6	6	Heat (1995)	Action|Crime|...
7	7	Sabrina (1995)	Comedy|Ro...
8	8	Tom and Huc...	Adventure|Ch...
9	9	Sudden Deat...	Action
10	10	GoldenEye (1...	Action|Advent...
11	11	American Pre...	Comedy|Dra...
12	12	Dracula: Dea...	Comedy|Horror
13	13	Balto (1995)	Adventure|Ani...
14	14	Nixon (1995)	Drama
15	15	Cutthroat Isla...	Action|Advent...
16	16	Casino (1995)	Crime|Drama

ExampleSet (9,124 examples, 0 special attributes, 3 regular attributes)

File
Edit
Process
View
Connections
Settings
Extensions
Help
Views:
Design
Results
Turbo Prep
Auto Model
Deployments
Find data, operators...etc
All Studio
Result History
ExampleSet (MovieID (2))
Data
Statistics
Visualizations
Annotations
Open in
Turbo Prep
Auto Model
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
Connections
data
processes
pivot calculation (10/26/20 10:54 PM
movie_recommend_1 (Local)
Movie_Recommendation_2 (Local)
Connections
Data
movies (10/15/20 6:27 PM – 1.5 M
ratings (10/15/20 6:26 PM – 3.5 M
Process
dmw_project (10/26/20 10:56 PM –
Temporary Repository (Local)
Connections
Cross Distances Temp 47274 (10
Select Attributes Temp 42469 (10/
Split Data Temp 49301 (10/15/20 5
Split Data Temp 62270 (10/15/20 5
Subprocess Temp 3606 (10/26/20
Activate Windows
Go to Settings to activate Windows.
Desktop
11:02 PM
Monday
10/26/2020

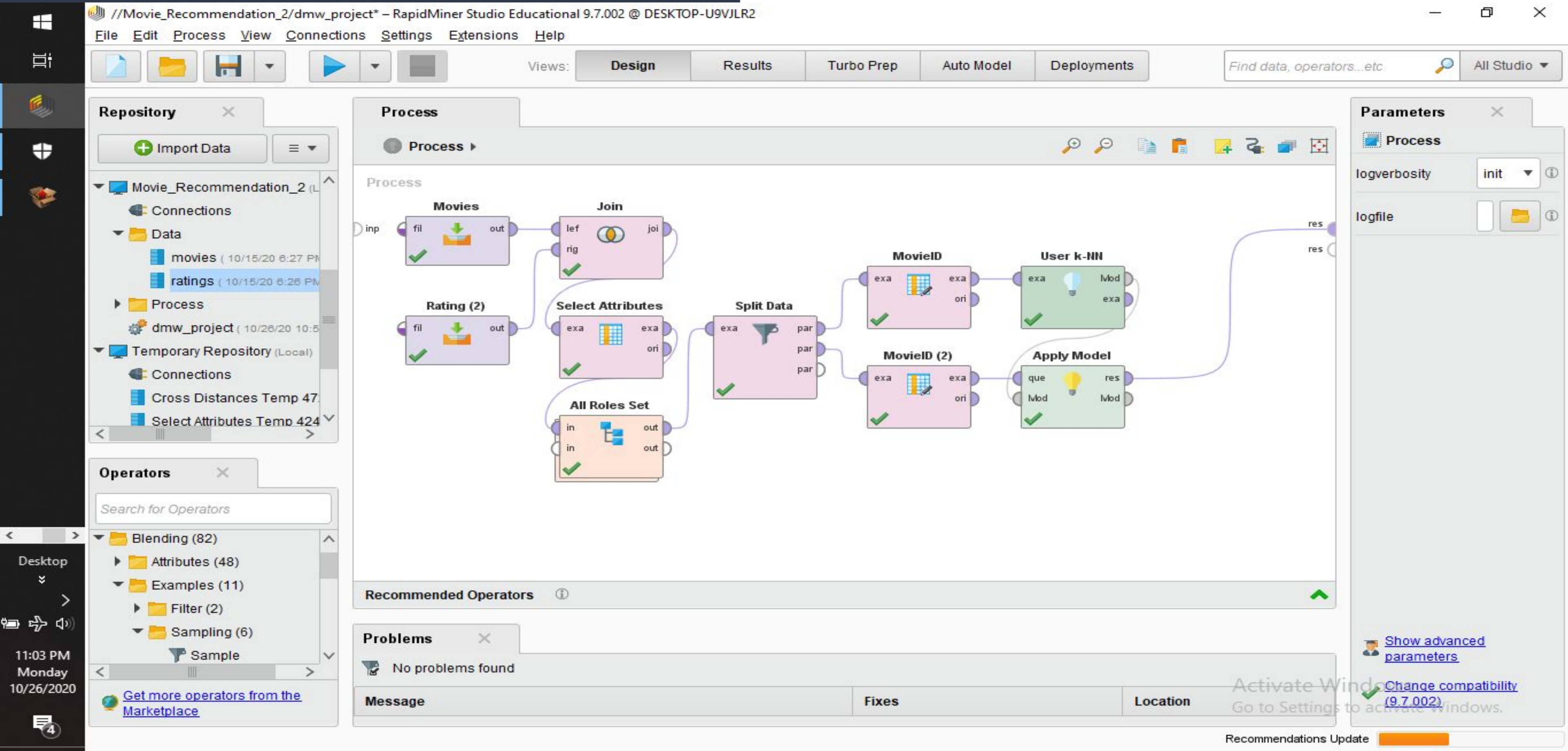
Ratings.csv:


//Movie_Recommendation_2/dmw_project* – RapidMiner Studio Educational 9.7.002 @ DESKTOP-U9VJLR2
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
All Studio
Result History
ExampleSet (//Movie_Recommendation_2/Data/ratings)
Data
Statistics
Visualizations
Annotations
Open in Turbo Prep Auto Model
Filter (100,004 / 100,004 examples): all
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
Connections
data
processes
pivot calculation (10/26/20 10:54 PM
movie_recommend_1 (Local)
Movie_Recommendation_2 (Local)
Connections
Data
movies (10/15/20 6:27 PM – 1.5
ratings (10/15/20 6:26 PM – 3.5
Process
dmw_project (10/26/20 10:56 PM –
Temporary Repository (Local)
Connections
Cross Distances Temp 47274 (10
Select Attributes Temp 42469 (10/
Split Data Temp 49301 (10/15/20
Split Data Temp 62270 (10/15/20
Subprocess Temp 3606 (10/26/20
Activate Windows
Go to Settings to activate Windows.
Desktop
11:02 PM
Monday
10/26/2020

Row No.	userId	rating	movieId	timestamp
1	1	2.500	31	1260759144
2	1	3	1029	1260759179
3	1	3	1061	1260759182
4	1	2	1129	1260759185
5	1	4	1172	1260759205
6	1	2	1263	1260759151
7	1	2	1287	1260759187
8	1	2	1293	1260759148
9	1	3.500	1339	1260759125
10	1	2	1343	1260759131
11	1	2.500	1371	1260759135
12	1	1	1405	1260759203
13	1	4	1953	1260759191
14	1	4	2105	1260759139
15	1	3	2150	1260759194
16	1	2	2193	1260759198

ExampleSet (100,004 examples, 2 special attributes, 2 regular attributes)

Process:



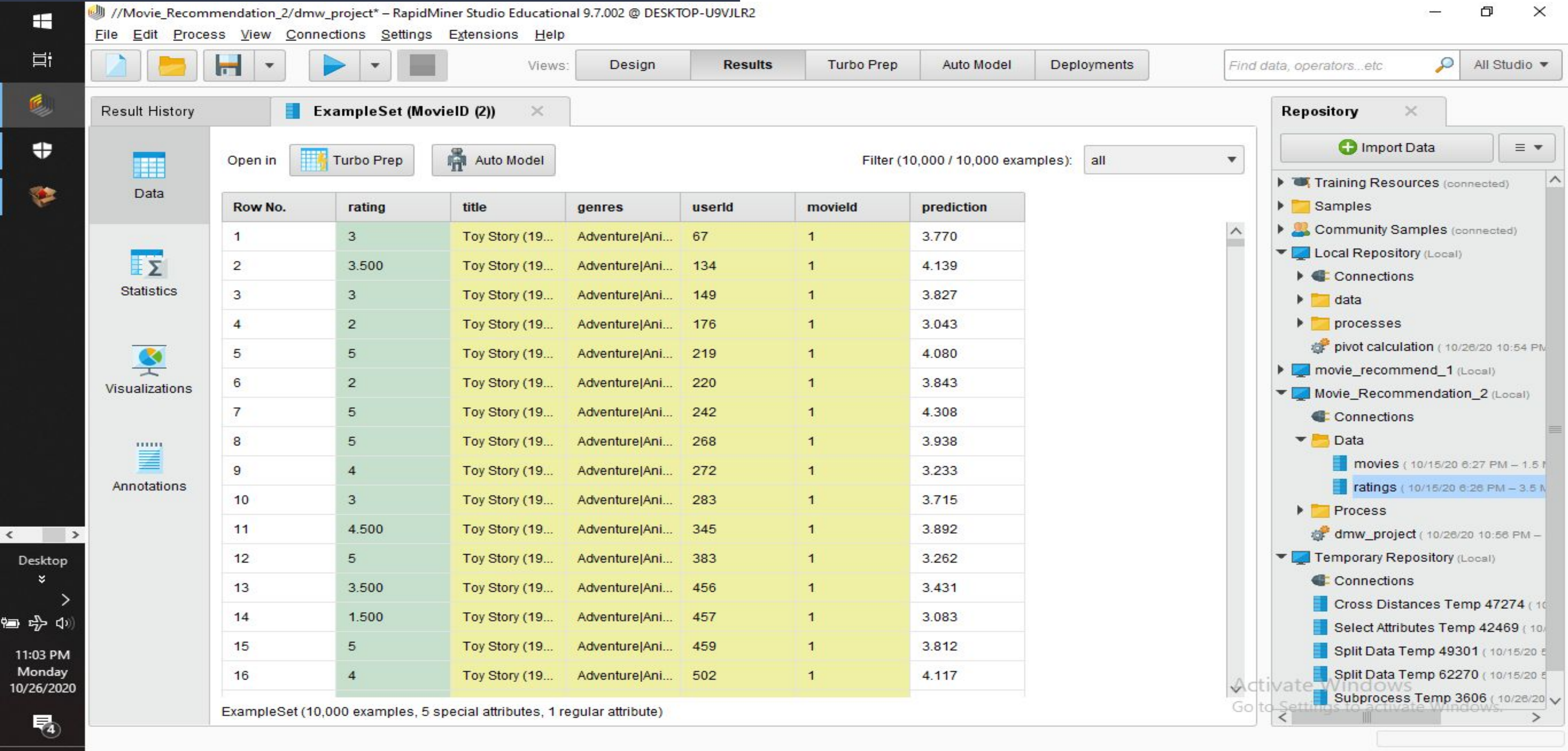
//Movie_Recommendation_2/dmw_project* – RapidMiner Studio Educational 9.7.002 @ DESKTOP-U9VJLR2

logverbosity: init

logfile:

No problems found

Process Output:



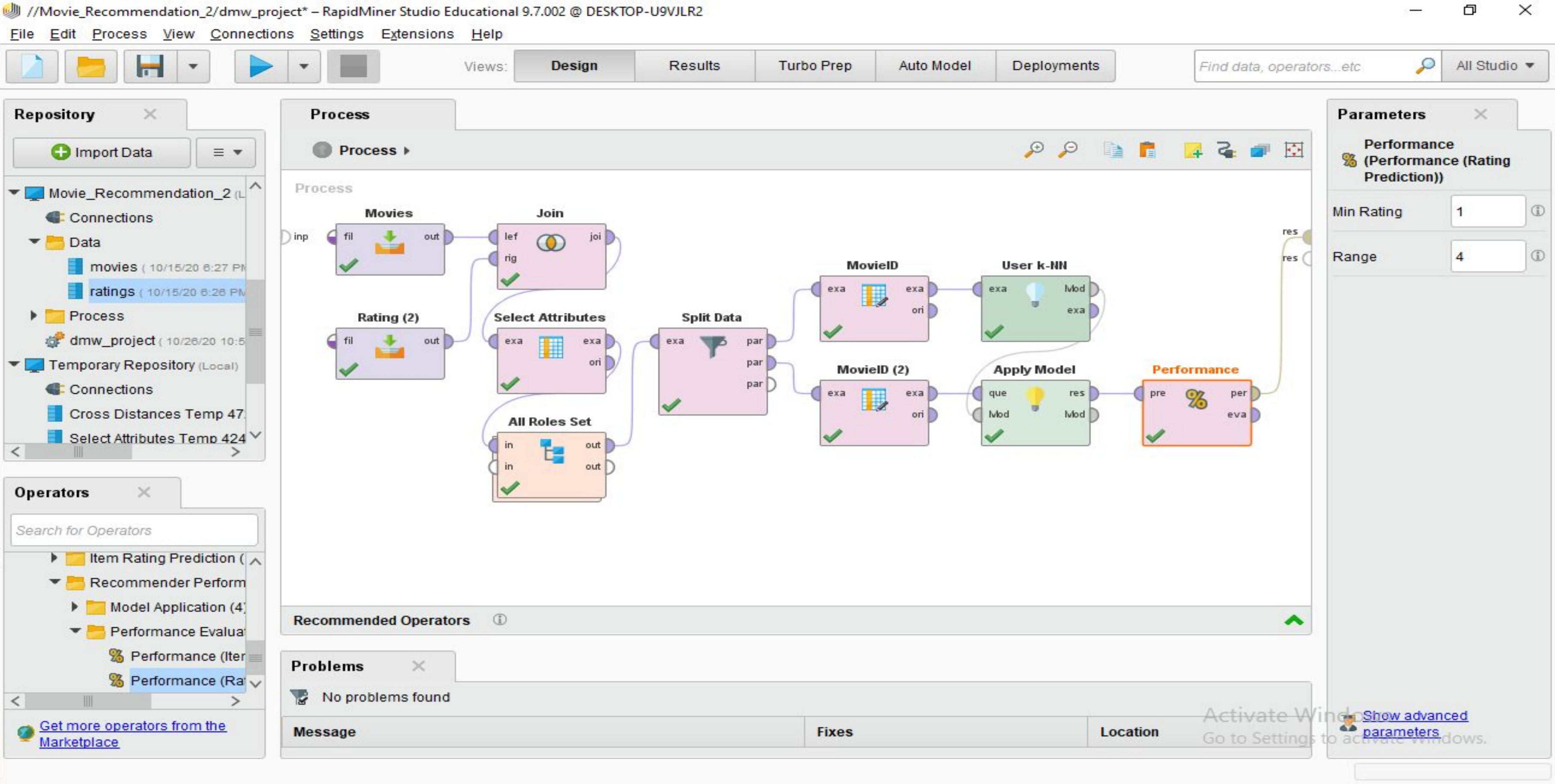
//Movie_Recommendation_2/dmw_project* – RapidMiner Studio Educational 9.7.002 @ DESKTOP-U9VJLR2
File
Edit
Process
View
Connections
Settings
Extensions
Help
Views:
Design
Results
Turbo Prep
Auto Model
Deployments
Find data, operators...etc
All Studio
Result History
ExampleSet (MovieID (2))
Data
Statistics
Visualizations
Annotations
Open in
Turbo Prep
Auto Model
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
Connections
data
processes
pivot calculation (10/26/20 10:54 PM
movie_recommend_1 (Local)
Movie_Recommendation_2 (Local)
Connections
Data
movies (10/15/20 6:27 PM – 1.5
ratings (10/15/20 6:26 PM – 3.5
Process
dmw_project (10/26/20 10:56 PM –
Temporary Repository (Local)
Connections
Cross Distances Temp 47274 (10
Select Attributes Temp 42469 (10/
Split Data Temp 49301 (10/15/20
Split Data Temp 62270 (10/15/20
Subprocess Temp 3606 (10/26/20
Activate Windows
Go to Settings to activate Windows.
Desktop
11:03 PM
Monday
10/26/2020

Filter (10,000 / 10,000 examples): all

Row No.	rating	title	genres	userId	movieId	prediction
1	3	Toy Story (19...	Adventure|Ani...	67	1	3.770
2	3.500	Toy Story (19...	Adventure|Ani...	134	1	4.139
3	3	Toy Story (19...	Adventure|Ani...	149	1	3.827
4	2	Toy Story (19...	Adventure|Ani...	176	1	3.043
5	5	Toy Story (19...	Adventure|Ani...	219	1	4.080
6	2	Toy Story (19...	Adventure|Ani...	220	1	3.843
7	5	Toy Story (19...	Adventure|Ani...	242	1	4.308
8	5	Toy Story (19...	Adventure|Ani...	268	1	3.938
9	4	Toy Story (19...	Adventure|Ani...	272	1	3.233
10	3	Toy Story (19...	Adventure|Ani...	283	1	3.715
11	4.500	Toy Story (19...	Adventure|Ani...	345	1	3.892
12	5	Toy Story (19...	Adventure|Ani...	383	1	3.262
13	3.500	Toy Story (19...	Adventure|Ani...	456	1	3.431
14	1.500	Toy Story (19...	Adventure|Ani...	457	1	3.083
15	5	Toy Story (19...	Adventure|Ani...	459	1	3.812
16	4	Toy Story (19...	Adventure|Ani...	502	1	4.117

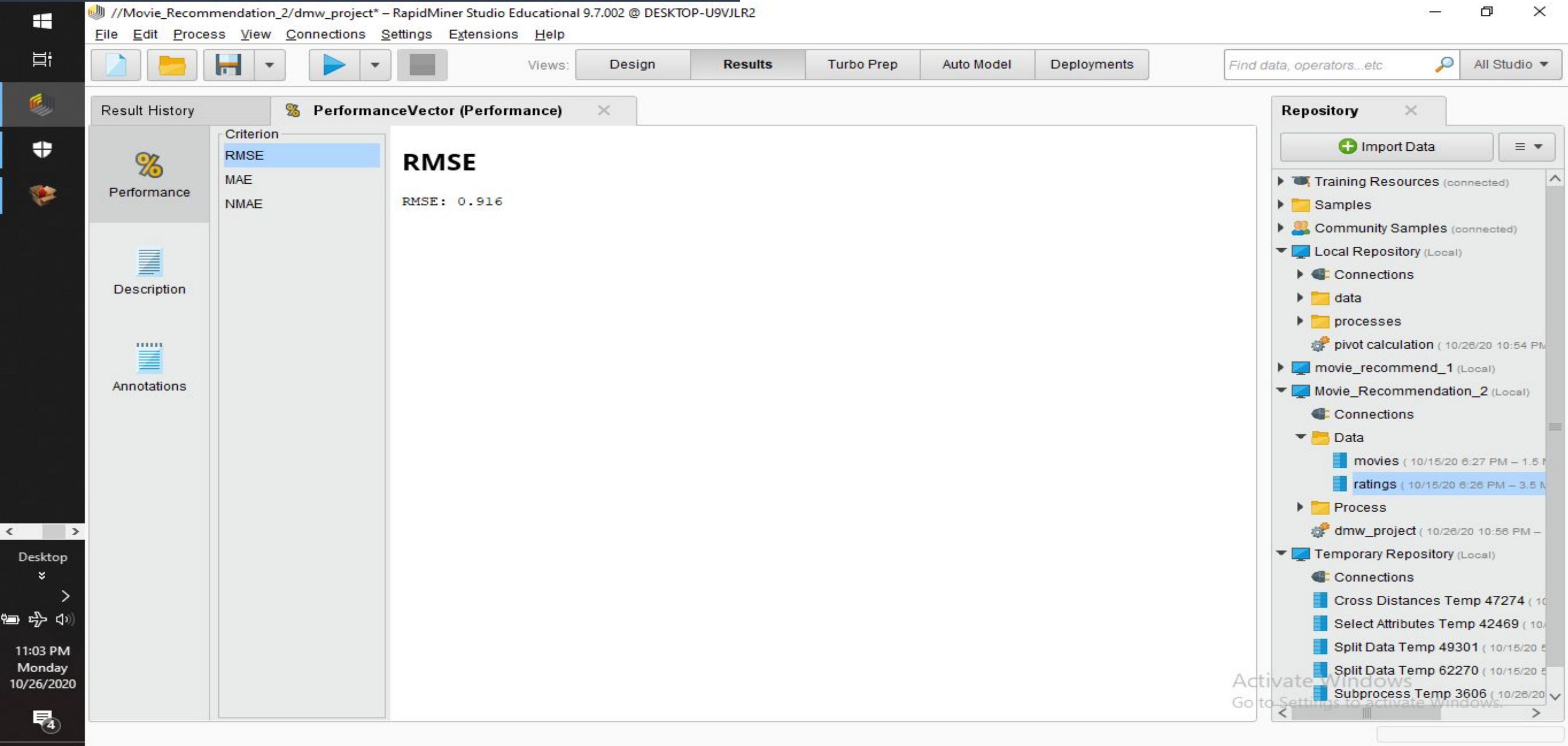
ExampleSet (10,000 examples, 5 special attributes, 1 regular attribute)

Performance:



//Movie_Recommendation_2/dmw_project* – RapidMiner Studio Educational 9.7.002 @ DESKTOP-U9VJLR2

Performance Output:

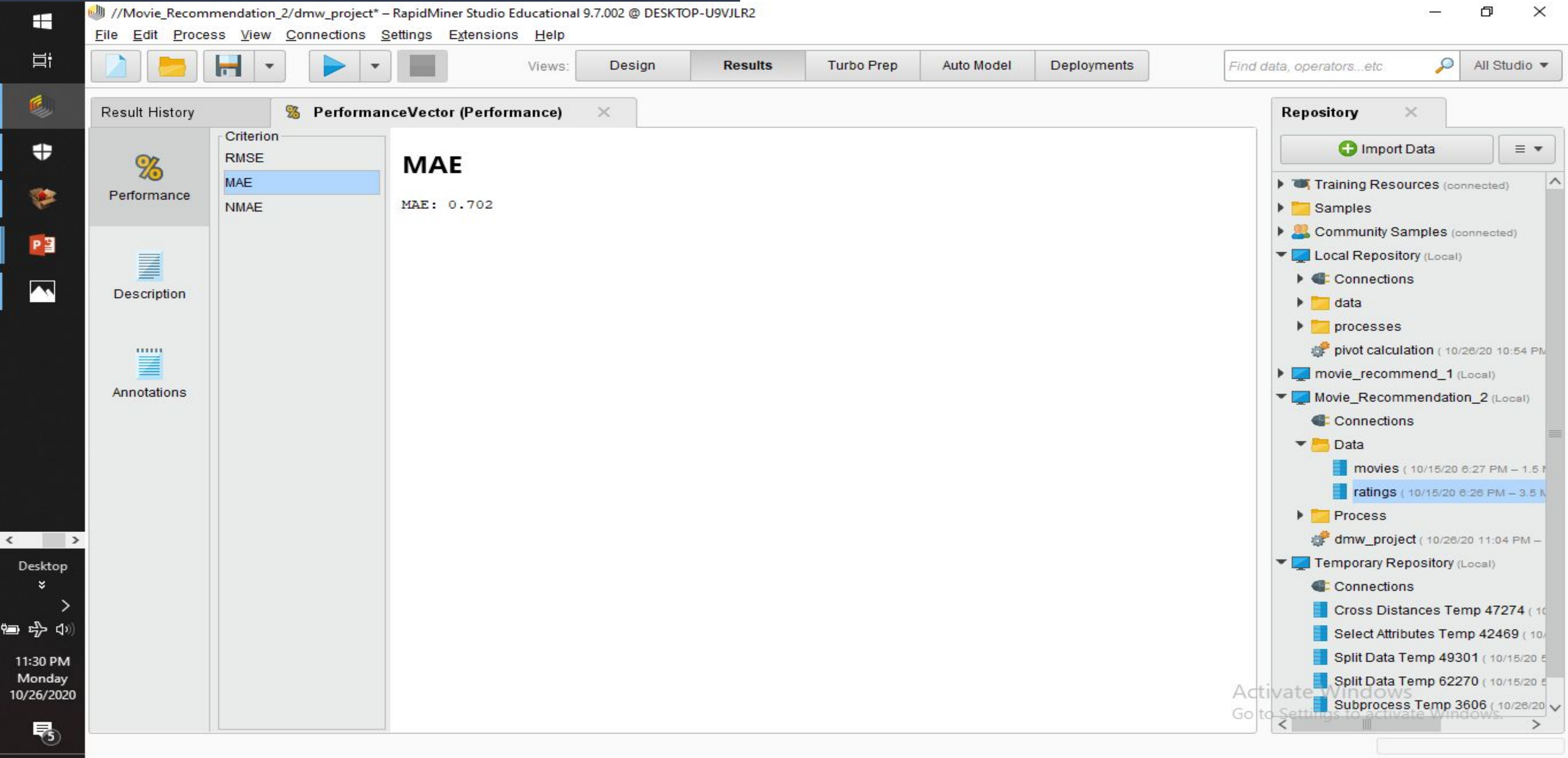


//Movie_Recommendation_2/dmw_project* – RapidMiner Studio Educational 9.7.002 @ DESKTOP-U9VJLR2
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
All Studio
Result History
PerformanceVector (Performance)
Performance
Description
Annotations
Criterion
RMSE
MAE
NMAE
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
Connections
data
processes
pivot calculation (10/26/20 10:54 PM
movie_recommend_1 (Local)
Movie_Recommendation_2 (Local)
Connections
Data
movies (10/15/20 6:27 PM – 1.5
ratings (10/15/20 6:26 PM – 3.5
Process
dmw_project (10/26/20 10:56 PM –
Temporary Repository (Local)
Connections
Cross Distances Temp 47274
Select Attributes Temp 42469
Split Data Temp 49301 (10/15/20
Split Data Temp 62270 (10/15/20
Subprocess Temp 3606 (10/26/20
Activate Windows
Go to Settings to activate Windows.
Desktop
11:03 PM
Monday
10/26/2020

RMSE

```
RMSE: 0.916
```

Performance Output:



//Movie_Recommendation_2/dmw_project* – RapidMiner Studio Educational 9.7.002 @ DESKTOP-U9VJLR2

PerformanceVector (Performance)

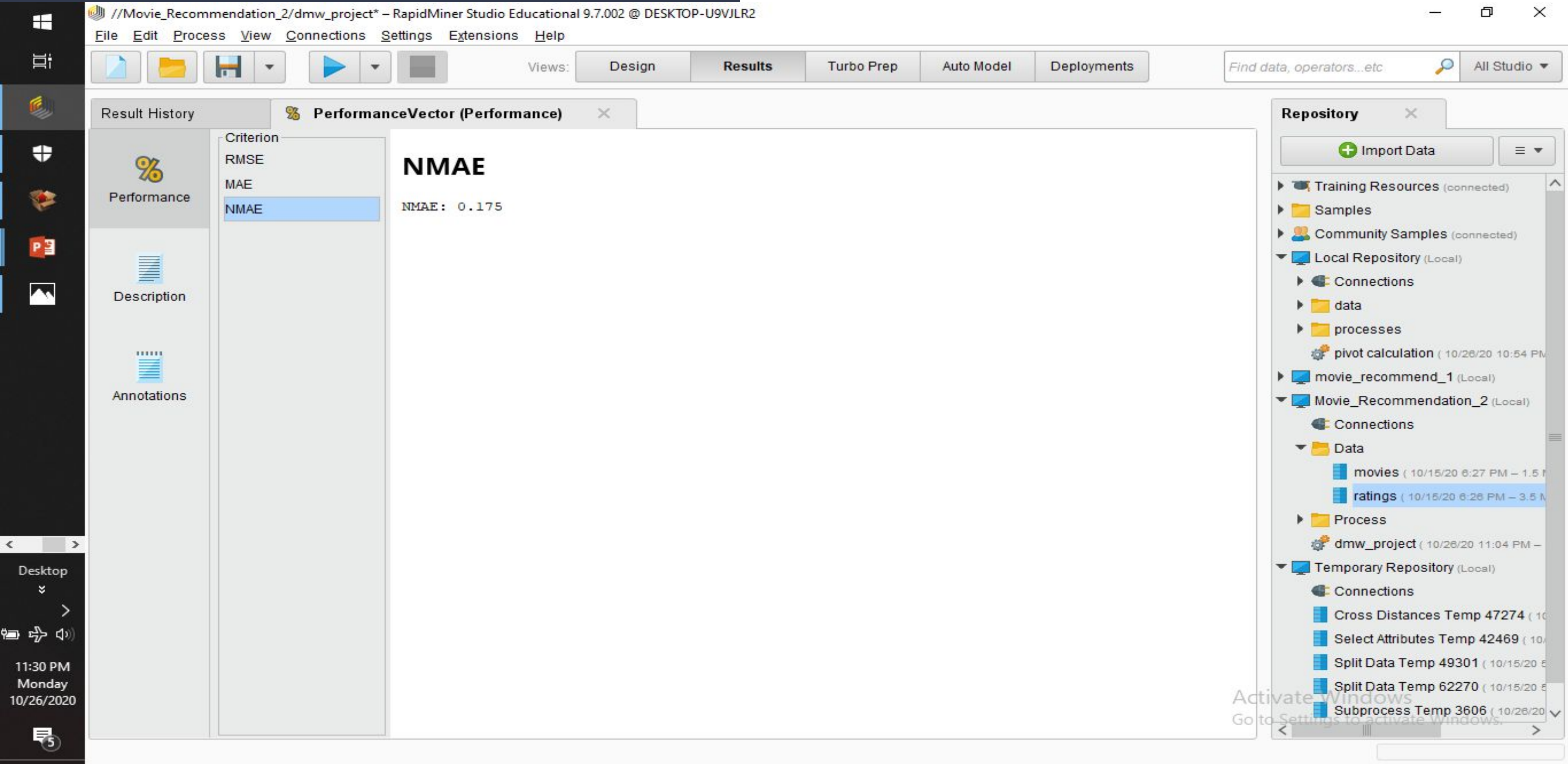
Criterion
RMSE
MAE
NMAE

MAE

```
MAE: 0.702
```

File
Edit
Process
View
Connections
Settings
Extensions
Help
Views:
Design
Results
Turbo Prep
Auto Model
Deployments
Find data, operators...etc
All Studio
Result History
Performance
Description
Annotations
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
Connections
data
processes
pivot calculation (10/26/20 10:54 PM
movie_recommend_1 (Local)
Movie_Recommendation_2 (Local)
Connections
Data
movies (10/15/20 6:27 PM – 1.5
ratings (10/15/20 6:26 PM – 3.5
Process
dmw_project (10/26/20 11:04 PM –
Temporary Repository (Local)
Connections
Cross Distances Temp 47274
Select Attributes Temp 42469
Split Data Temp 49301 (10/15/20
Split Data Temp 62270 (10/15/20
Subprocess Temp 3606 (10/26/20
Activate Windows
Go to Settings to activate Windows.
Desktop
11:30 PM
Monday
10/26/2020

Performance Output:



//Movie_Recommendation_2/dmw_project* – RapidMiner Studio Educational 9.7.002 @ DESKTOP-U9VJLR2
File Edit Process View Connections Settings Extensions Help
Views: Design Results Turbo Prep Auto Model Deployments
Find data, operators...etc
All Studio
Result History
PerformanceVector (Performance)
Performance
Description
Annotations
Criterion
RMSE
MAE
NMAE
Repository
Import Data
Training Resources (connected)
Samples
Community Samples (connected)
Local Repository (Local)
Connections
data
processes
pivot calculation (10/26/20 10:54 PM
movie_recommend_1 (Local)
Movie_Recommendation_2 (Local)
Connections
Data
movies (10/15/20 6:27 PM – 1.5
ratings (10/15/20 6:26 PM – 3.5
Process
dmw_project (10/26/20 11:04 PM –
Temporary Repository (Local)
Connections
Cross Distances Temp 47274
Select Attributes Temp 42469
Split Data Temp 49301 (10/15/20
Split Data Temp 62270 (10/15/20
Subprocess Temp 3606 (10/26/20
Activate Windows
Go to Settings to activate Windows.
Desktop
11:30 PM
Monday
10/26/2020

NMAE

```
NMAE: 0.175
```

Thank You