



L OVELY
P ROFESSIONAL
U NIVERSITY

School of computer application

CA third
of

DATA WAREHOUSING AND DATA MINING - LAB(CAP447)

Session 2022-2024

Submitted to:

Name: Ms. Satwinder Kaur
(Dept of. Computer Science)

Submitted by:

Name: Sanjeev Kumar
Roll no: RDOC14 A10
Course: MCA

Department of computer Science
Lovely Professional University Jalandhar Punjab (144401)
India.

DATA WAREHOUSING AND DATA MINING LABORATORY (CAP447)

SET-B

Q1. Construct a decision tree modal on dataset and show performance.

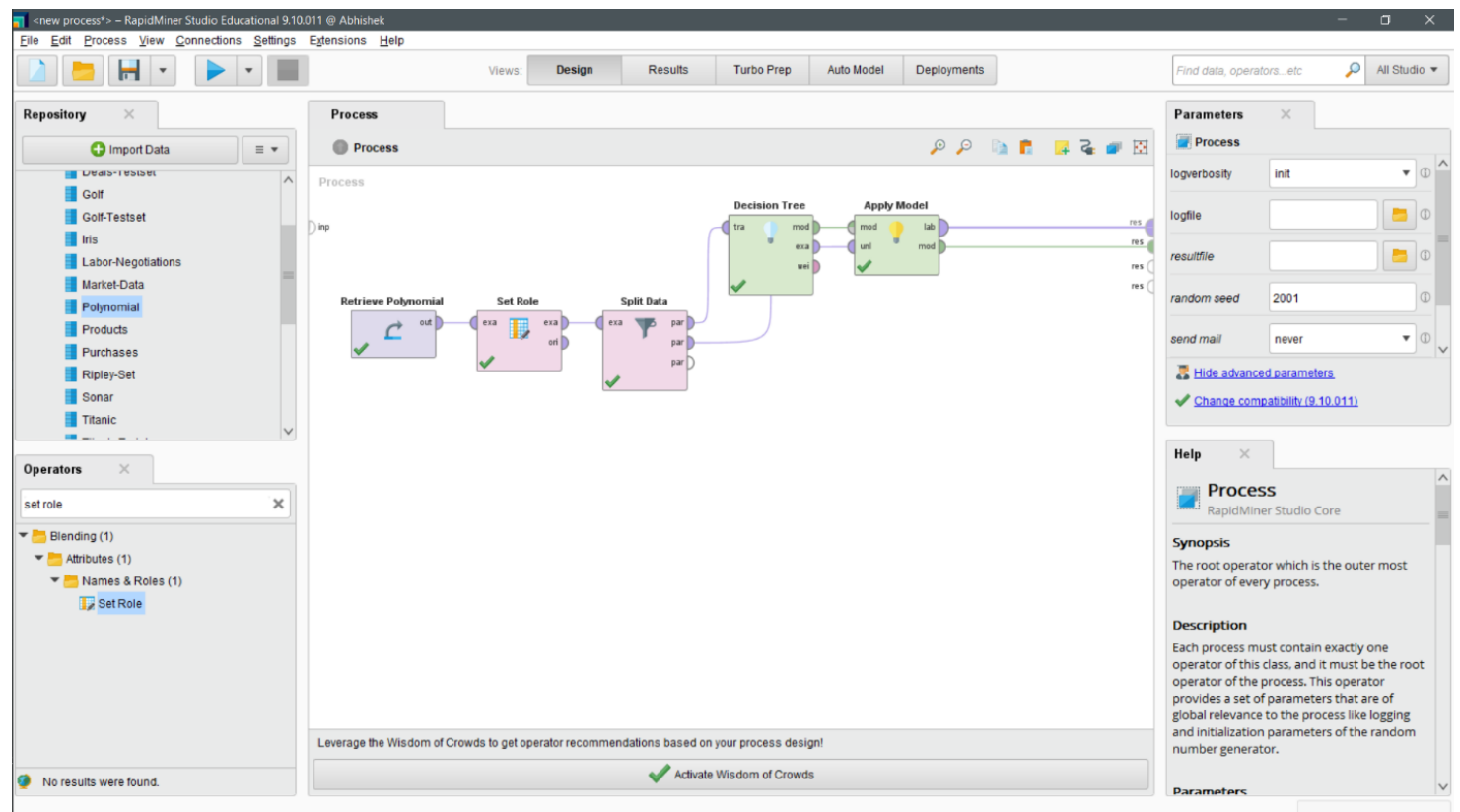
Q2. Implement following Operators on any data :

1. **Rename by replacing**
2. **Split data**
3. **Set role**
4. **Read CSV**

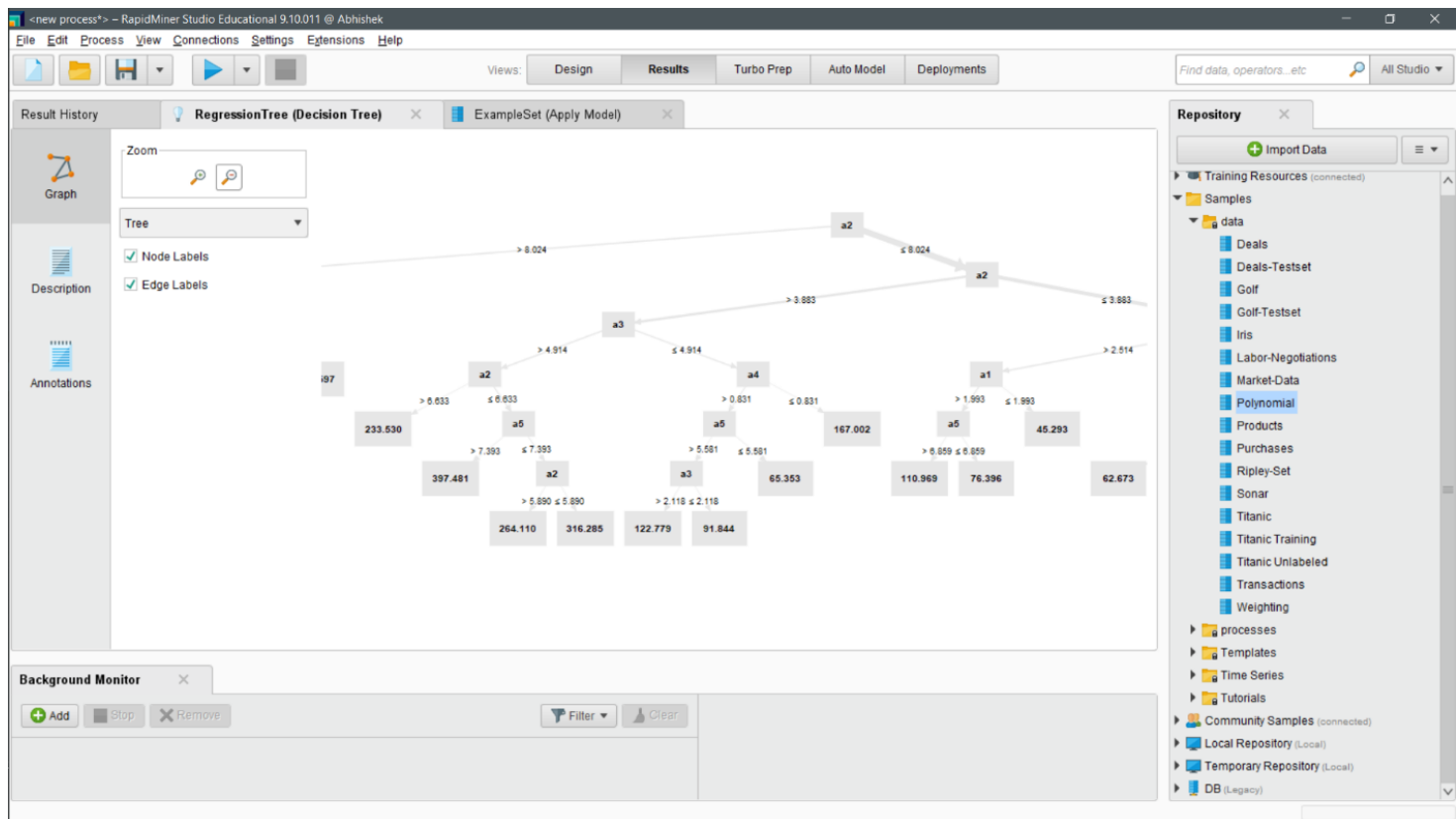
Q3. What do you mean by sampling ? Implement it on any data set.

Que-1

Step 1: To import data from a local computer, firstly we will click on data and then drag the polynomial dataset to the blank process .



Output:



Que- 2

From operators drag Read CSV operator. This Operator reads an ExampleSet from the specified CSV file.

The screenshot shows the RapidMiner Studio interface with a workflow in the 'Design' view. The workflow consists of four operators: 'Read CSV', 'Rename by Replacing', 'Set Role', and 'Split Data'. The 'Read CSV' operator is connected to the 'Rename by Replacing' operator, which is connected to the 'Set Role' operator, which is finally connected to the 'Split Data' operator. The 'Split Data' operator has three output ports labeled 'res', 'res', and 'res'. The 'Parameters' panel on the right shows the 'Split Data' operator's settings, including 'partitions' set to 'Edit Enumeration (2)...' and 'sampling type' set to 'automatic'. The 'Help' panel on the right provides information about the 'Split Data' operator, including its synopsis and description.

Repository

- Temporary Repository (Local)
- Connections
- aggregate (11/28/22 1:57 PM - 3 kB)
- all attr (11/28/22 12:55 PM - 6 kB)
- ca1 (11/7/22 3:03 PM - 4 kB)
- decision (12/12/22 12:43 PM - 8 kB)
- multiply (11/28/22 1:39 PM - 4 kB)
- netflix_titles (12/12/22 12:36 PM - 10.2 MB)
- read split (12/12/22 12:51 PM - 5 kB)
- Retrieve Iris Temp 95048 (10/17/22 12:40 P
- titanic (10/31/22 12:16 PM - 7 kB)

Operators

- split
- Blending (2)
- Examples (1)
- Sampling (1)
- Split Data
- Values (1)
- Split
- Validation (3)
- Cross Validation
- Split Validation
- Wrapper Split Validation

Process

Process

Read CSV → Rename by Replacing → Set Role → Split Data

Parameters

Split Data

partitions Edit Enumeration (2)...

sampling type automatic

☐ use local random seed

[Hide advanced parameters](#)

Help

Split Data

RapidMiner Studio Core

Tags: Divide, Separate, Part, Training, Testing, Samples, Subsets, Partitions, Sampling

Synopsis

This operator produces the desired number of subsets of the given ExampleSet. The ExampleSet is partitioned into subsets according to the specified relative sizes.

[Jump to Tutorial Process](#)

Description

The Split Data operator takes an ExampleSet as its input and delivers the subsets of that ExampleSet through its output ports. The number of subsets (or partitions) and the

Output:

The screenshot shows the 'Results' view of the RapidMiner Studio interface. The 'ExampleSet (Split Data)' operator is selected, and its output is displayed as a table. The table has 14 rows and 11 columns: 'Row No.', 'release_year', 'show_id', 'type', 'title', 'director', 'casted by', 'country', 'date_added', 'rating', and 'duration'. The data includes movies like 'Dick Johnson...', 'Sankofa', 'The Starling', and TV shows like 'Blood & Water', 'Jailbirds New...', 'Kota Factory', 'Midnight Mass', 'Vendetta: Tru...', 'Bangkok Bre...', 'Je Suis Karl', 'Confessions...', and 'Crime Story'.

Result History

- ExampleSet (Split Data)
- ExampleSet (Split Data)

Open in Turbo Prep Auto Model

Filter (7,028 / 7,028 examples): all

Row No.	release_year	show_id	type	title	director	casted by	country	date_added	rating	duration
1	2020	s1	Movie	Dick Johnson...	Kirsten John...	?	United States	September 2...	PG-13	90 min
2	2021	s2	TV Show	Blood & Water	?	Ama Qamata...	South Africa	September 2...	TV-MA	2 Seasons
3	2021	s4	TV Show	Jailbirds New...	?	?	?	September 2...	TV-MA	1 Season
4	2021	s5	TV Show	Kota Factory	?	Mayur More, J...	India	September 2...	TV-MA	2 Seasons
5	2021	s6	TV Show	Midnight Mass	Mike Flanagan	Kate Siegel, ...	?	September 2...	TV-MA	1 Season
6	2021	s7	Movie	My Little Pony...	Robert Culle...	Vanessa Hud...	?	September 2...	PG	91 min
7	1993	s8	Movie	Sankofa	Haile Gerima	Kofi Ghanaba...	United States...	September 2...	TV-MA	125 min
8	2021	s9	TV Show	The Great Brit...	Andy Devons...	Mel Giedroyc...	United Kingd...	September 2...	TV-14	9 Seasons
9	2021	s10	Movie	The Starling	Theodore Melfi	Melissa McC...	United States	September 2...	PG-13	104 min
10	2021	s11	TV Show	Vendetta: Tru...	?	?	?	September 2...	TV-MA	1 Season
11	2021	s12	TV Show	Bangkok Bre...	Kongkiat Ko...	Sukollawat K...	?	September 2...	TV-MA	1 Season
12	2021	s13	Movie	Je Suis Karl	Christian Sch...	Luna Wedler...	Germany, Cz...	September 2...	TV-MA	127 min
13	2021	s14	Movie	Confessions ...	Bruno Garotti	Klara Castan...	?	September 2...	TV-PG	91 min
14	2021	s15	TV Show	Crime Story	?	?	?	September 2...	TV-MA	1 Season

ExampleSet (7,028 examples, 1 special attribute, 11 regular attributes)

Background Monitor

Add Stop Remove Filter Clear

Repository

- Import Data
- Products
- Purchases
- Ripley-Set
- Sonar
- Titanic
- Titanic Training
- Titanic Unlabeled
- Transactions
- Weighting
- processes
- Templates
- Time Series
- Tutorials
- Community Samples (connected)
- Local Repository (Local)
- Connections
- data
- processes
- rapid miner
- binning (12/5/22 1:59 PM - 2 kB)
- ca2 (11/11/22 11:19 AM - 981 bytes)
- dude (11/21/22 1:24 PM - 2 kB)
- netflix_titles (12/12/22 12:22 PM - 10.2 MB)
- normalize (12/5/22 12:43 PM - 7 kB)
- tit (10/17/22 1:54 PM - 1 kB)
- titanic (10/31/22 1:51 PM - 986 bytes)
- Temporary Repository (Local)

Que-3

The screenshot shows the RapidMiner Studio interface. On the left, the 'Repository' pane lists various datasets, including 'Titanic'. The 'Operators' pane shows the 'Sample' operator selected under 'Sampling (6)'. The central 'Process' pane displays a workflow: 'Retrieve Ripley-Set' and 'Retrieve Titanic' operators are connected to a 'Sample (2)' operator. The 'Sample' operator is configured with 'absolute' sampling and a 'sample size' of 20. The right-hand 'Parameters' pane shows the 'Sample' operator's settings, including 'sample' (absolute), 'balance data' (unchecked), 'sample size' (20), and 'use local random seed' (unchecked). The 'Help' pane provides a synopsis and description of the 'Sample' operator.

Process Design:

- Retrieve Ripley-Set
- Retrieve Titanic
- Sample (2)

Sample Operator Parameters:

- sample: absolute
- balance data: ☐
- sample size: 20
- use local random seed: ☐

Help - Sample

RapidMiner Studio Core

Tags: Subsets, Random, Ratio, Stratified, Stratification, Bootstrap, Population, Downsample, Sampling

Synopsis

This operator creates a sample from an ExampleSet by selecting examples randomly. The size of a sample can be specified on absolute, relative and probability basis.

[Jump to Tutorial Process](#)

Description

This operator is similar to the Filter Examples operator in principle that it takes an ExampleSet as input and delivers a subset of

Output:

The screenshot shows the 'Results' view of RapidMiner Studio. The 'ExampleSet (Sample 2)' is displayed as a table with 20 rows and 11 columns. The columns are: Row No., Passenger..., Name, Sex, Age, No of Sibling..., No of Parent..., Ticket Num..., Passenger F..., Cabin, and Port of Emi. The table contains data for 20 examples. The 'Background Monitor' pane at the bottom shows the process is running.

ExampleSet (Sample 2)

Row No.	Passenger...	Name	Sex	Age	No of Sibling...	No of Parent...	Ticket Num...	Passenger F...	Cabin	Port of Emi
1	First	Carter, Mrs. ...	Female	36	1	2	113760	120	B96 B98	Southampt
2	First	Hilliard, Mr. H...	Male	?	0	0	17463	51.862	E46	Southampt
3	First	Hipkins, Mr. ...	Male	55	0	0	680	50	C39	Southampt
4	First	Ward, Miss. A...	Female	35	0	0	PC 17755	512.329	?	Cherbourg
5	Second	Brown, Mrs. T...	Female	40	1	1	29750	39	?	Southampt
6	Second	Collander, Mr...	Male	28	0	0	248740	13	?	Southampt
7	Second	Louch, Mr. Ch...	Male	50	1	0	SC/4H 3085	26	?	Southampt
8	Second	Reeves, Mr. ...	Male	36	0	0	C.A. 17248	10.500	?	Southampt
9	Second	Reynaldo, Ms...	Female	28	0	0	230434	13	?	Southampt
10	Second	Sincock, Miss...	Female	20	0	0	C.A. 33112	36.750	?	Southampt
11	Second	Weisz, Mrs. L...	Female	29	1	0	228414	26	?	Southampt
12	Third	Burns, Miss. ...	Female	18	0	0	330963	7.879	?	Queenstow
13	Third	Leent, Mr. Fa...	Male	22	0	0	2620	7.225	?	Cherbourg
14	Third	Lennon, Miss	Female	?	1	0	370371	15.500	?	Queenstow

ExampleSet (20 examples, 0 special attributes, 12 regular attributes)