



CPE408 - BIG DATA AND AI

การวิเคราะห์ข้อมูลเชิงลึก

Presented by Malila

Group Member

- 64030493**
น.ส.จันทร์กานต์ เมืองนก
- 64043118**
น.ส.พัชรา เมราโชคิติมณีกุล
- 64043156**
น.ส.รัณรร ศรีษะสารถุร
- 64054662**
น.ส.นกสตร จิระศิริโชค
- 64060399**
น.ส.รัตติกาล จันทร์รัชติกาล



Presented by Malila

Business Understanding - Pain point

การพยากรณ์สภาพธุรกิจท้องถนน
เพื่อคาดการณ์ความผิดของผู้จราจรในแต่ละแยกตามช่วงเวลา



Presented by Malila

Data Understanding

Examine the data; identify problems in the data



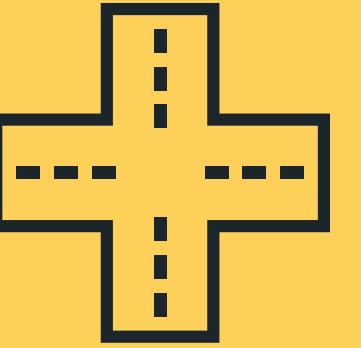
Presented by Malila

Data Understanding

Dataset Selection



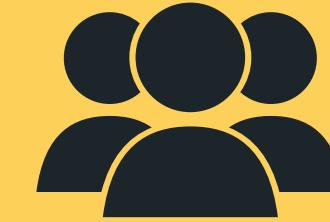
DateTime
วันเวลาที่บันทึก



Junction
หมายเลขของแยกถนน



Vehicles
ปริมาณของรถที่บันทึก



ID
เลขประจำตัว





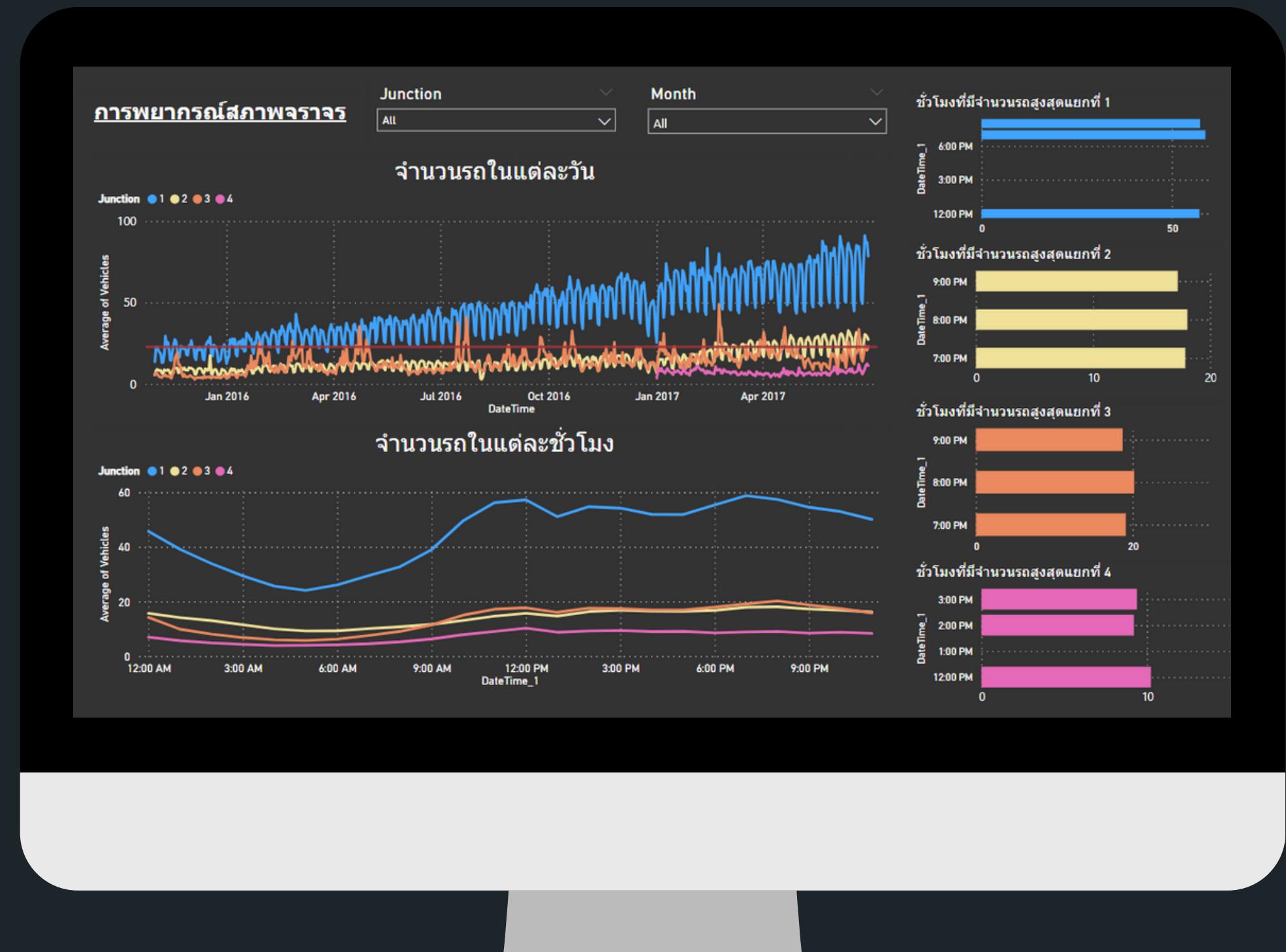
Descriptive Analytics

Data sets บอกถึงจำนวนรถ
ในแต่ละแยก (4 แยก) ตามช่วงเวลา



Presented by Malila

Diagnostics Analytics



Presented by Malila

Data Understanding

Dataset Problems

- Data redundancy
- Missing Values
- Data inconsistency
- Outlier data
- Imbalance dataset

Row No.	Date Time	Junction	Vehicles	ID
1	Nov 1, 2015 ...	1	15	20151101001
2	Nov 1, 2015 ...	1	13	20151101011
3	Nov 1, 2015 ...	1	10	20151101021
4	Nov 1, 2015 ...	1	7	20151101031
5	Nov 1, 2015 ...	1	9	20151101041
6	Nov 1, 2015 ...	1	6	20151101051
7	Nov 1, 2015 ...	1	9	20151101061
8	Nov 1, 2015 ...	1	8	20151101071
9	Nov 1, 2015 ...	1	11	20151101081
10	Nov 1, 2015 ...	1	12	20151101091
11	Nov 1, 2015 ...	1	15	20151101101
12	Nov 1, 2015 ...	1	17	20151101111
13	Nov 1, 2015 ...	1	16	20151101121
14	Nov 1, 2015 ...	1	15	20151101131

Column ID เป็นคอลัมที่ไม่จำเป็นต่อการพยากรณ์นี้



Data Preparation

Fix problems in the data; create derived variables



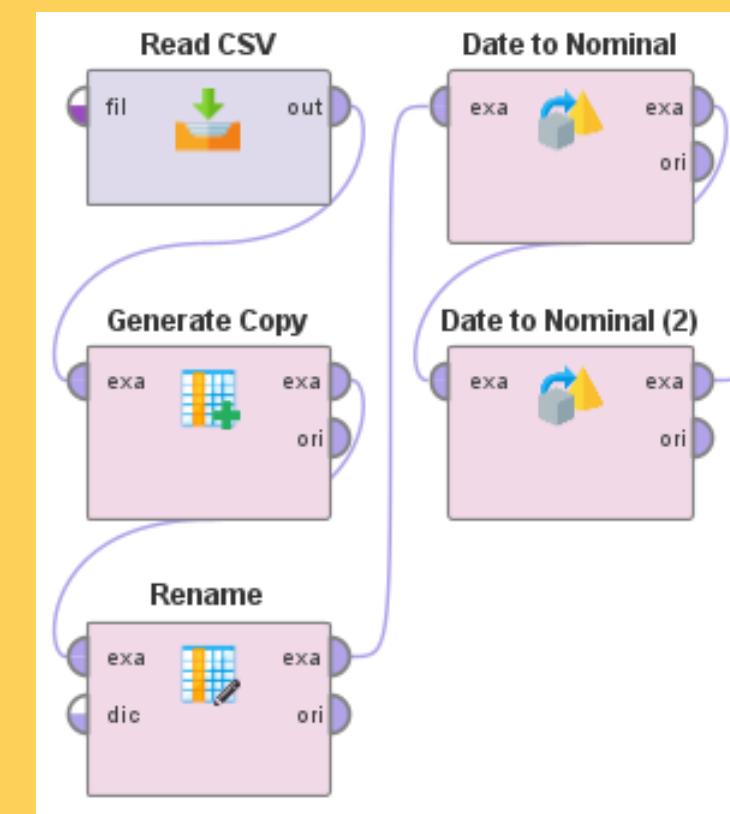
Presented by Malila

Data Preparation

Fix problems in the data; create derived variables

ทำการ Extract Column DateTime ออกจากกันเป็น Date และ Time ดังนี้

Row No.	Junction	Vehicles	ID	Date	Time
1	1	15	20151101001	01.11.2015	00:00:00
2	1	13	20151101011	01.11.2015	01:00:00
3	1	10	20151101021	01.11.2015	02:00:00
4	1	7	20151101031	01.11.2015	03:00:00
5	1	9	20151101041	01.11.2015	04:00:00



ตั้งค่า Date มีแค่วันที่ และ Time มีแค่เวลา

attribute name	DateTime
new name	Date
attribute name	Date
date format	dd.MM.yyyy
attribute name	Time
date format	HH:mm:ss

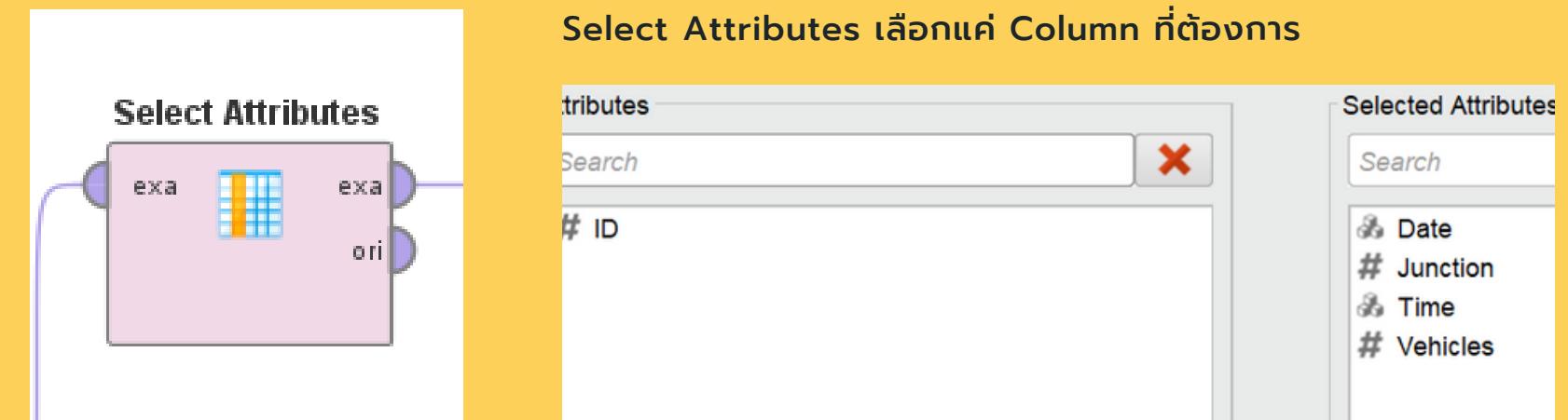


Data Preparation

Fix problems in the data; create derived variables

Data Problem : Data inconsistency
เป็นข้อมูลที่มีการเก็บมาแต่ไม่ได้สอดคล้องกับปัจจุหา
จึงต้องทำการลบ Column ID ออกดังนี้

Row No.	Junction	Vehicles	Date	Time
1	1	15	01.11.2015	00:00:00
2	1	13	01.11.2015	01:00:00
3	1	10	01.11.2015	02:00:00
4	1	7	01.11.2015	03:00:00
5	1	9	01.11.2015	04:00:00
6	1	6	01.11.2015	05:00:00



Data Preparation

Fix problems in the data; create derived variables

เมื่อเตรียมข้อมูลเสร็จสิ้นแล้วทำการจัดเรียงข้อมูล
จะได้ Datasets เป็นดังนี้

- Date วันที่
- Time เวลา
- Junction แยก
- Vehicles จำนวนรถ

Row No.	Date	Time	Junction	Vehicles
1	Nov 1, 2015	12:00:00 AM ICT	1	15
2	Nov 1, 2015	1:00:00 AM ICT	1	13
3	Nov 1, 2015	2:00:00 AM ICT	1	10

เรียง Attributes โดยใช้ Reorder Attributes

The screenshot shows the KNIME interface with a 'Reorder Attributes' node in the center. The node has four ports: 'ref' (bottom-left), 'ori' (bottom-right), and two unlabeled purple ports on top. A green checkmark is visible near the bottom-left port. To the right of the node is a 'Reorder Attributes' dialog box with the title 'Attribute Ordering'. Inside the dialog, there is a list of attributes: Date, Time, Junction, and Vehicles. A green '+' button is located at the top right of the list area.



Modeling

Build predictive or descriptive models



Presented by Malila



Modeling

Building some models

- Decision Tree
- Random Forest
- Naive Bayes

สร้างโมเดลโดยการใช้ 3 เทคนิค
โดยโมเดลจะแบ่งแต่ละแยก กั้งหมวด 4 แยก
เพื่อกำการพยากรณ์แต่ละแยกในช่วงวันเวลาหนึ่น ๆ
มีการรดติดหรือไม่

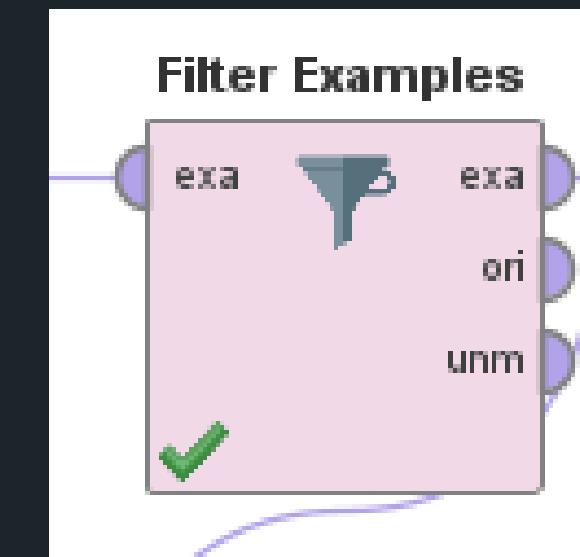


Presented by Malila

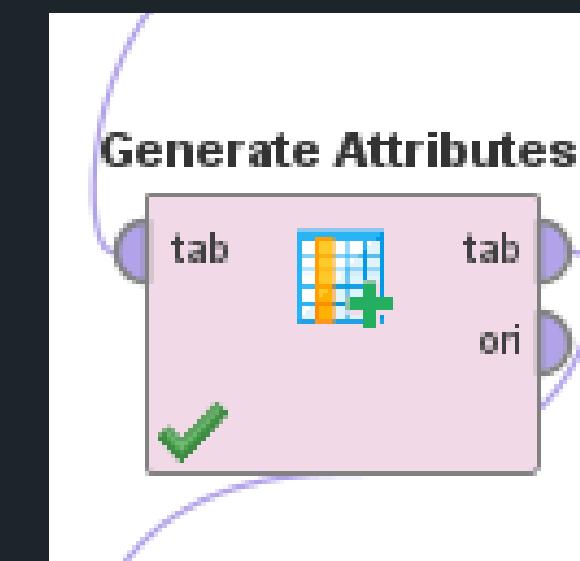
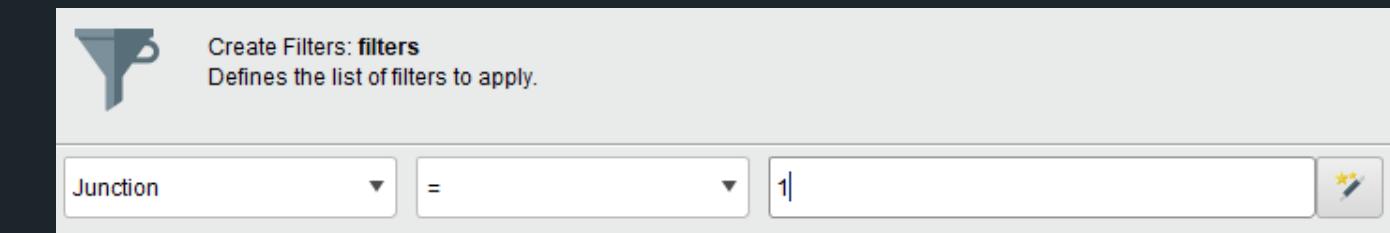
Modeling



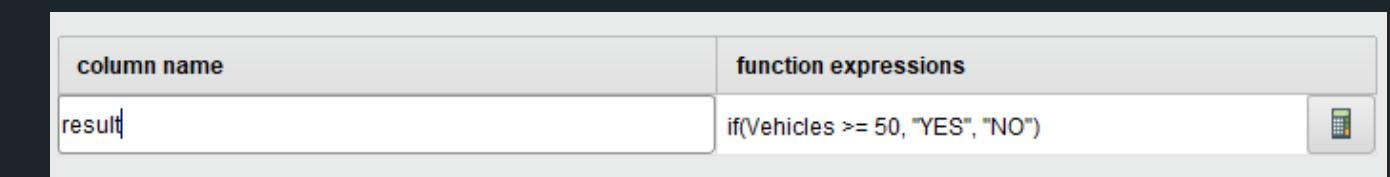
Presented by Malila



แบ่งแต่ละแยกโดยการใช้
Filter Examples : เลือก Column Junction
ให้เก่ากับ 1, 2, 3 หรือ 4 ตามลำดับ



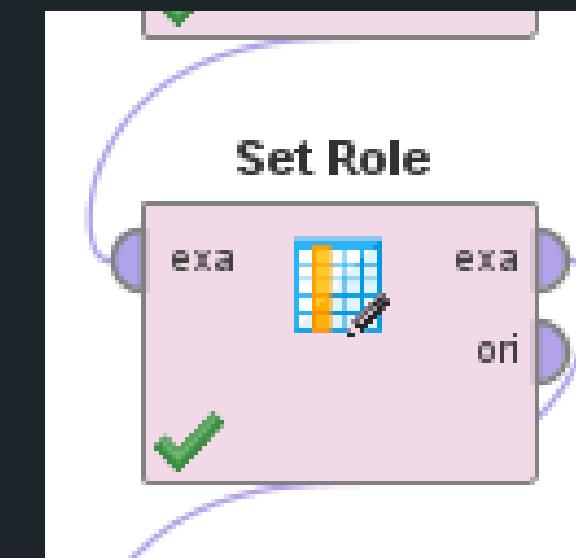
สร้าง Column ขึ้นมาใหม่ เพื่อแสดงผลลัพธ์การรถติดโดยการใช้
Generate Attributes : กำหนดชื่อ Column เป็น result
แล้วกำหนดเงื่อนไขดังนี้ ถ้าจำนวนรถมากกว่า 25 คัน
ให้แสดง "YES" ถ้าไม่ ให้แสดง "NO"



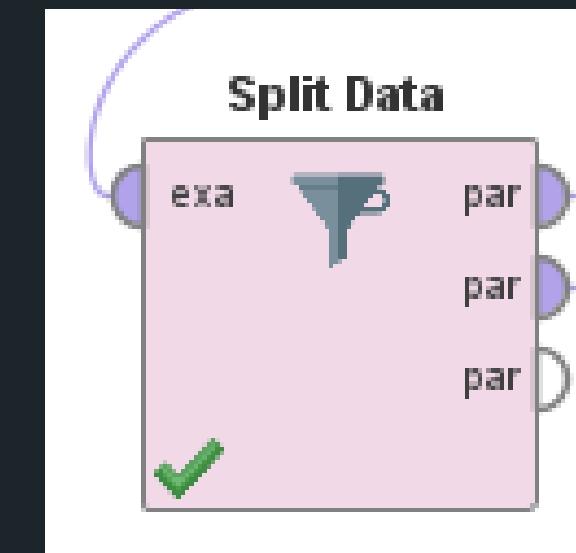
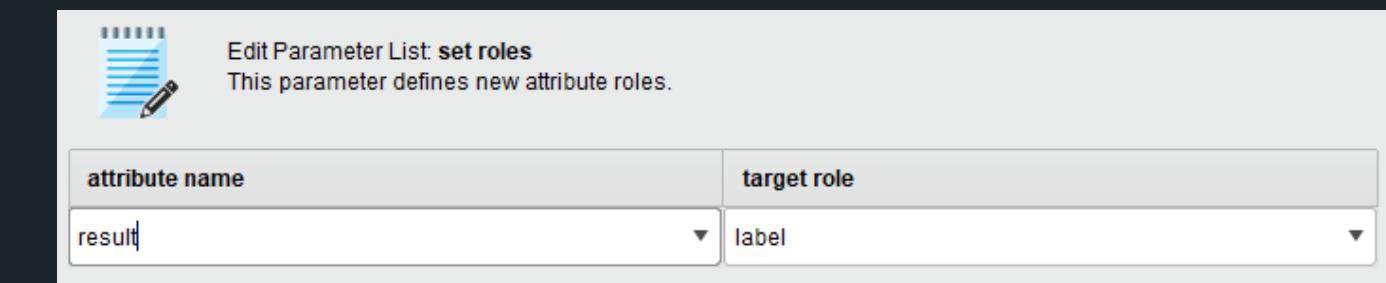
Modeling



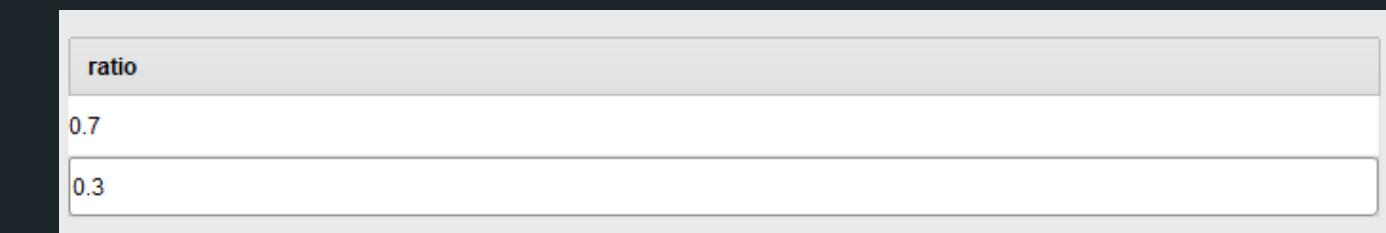
Presented by Malila



กำหนด result ให้เป็น label โดยการใช้
Set Role : เลือก result เป็น label

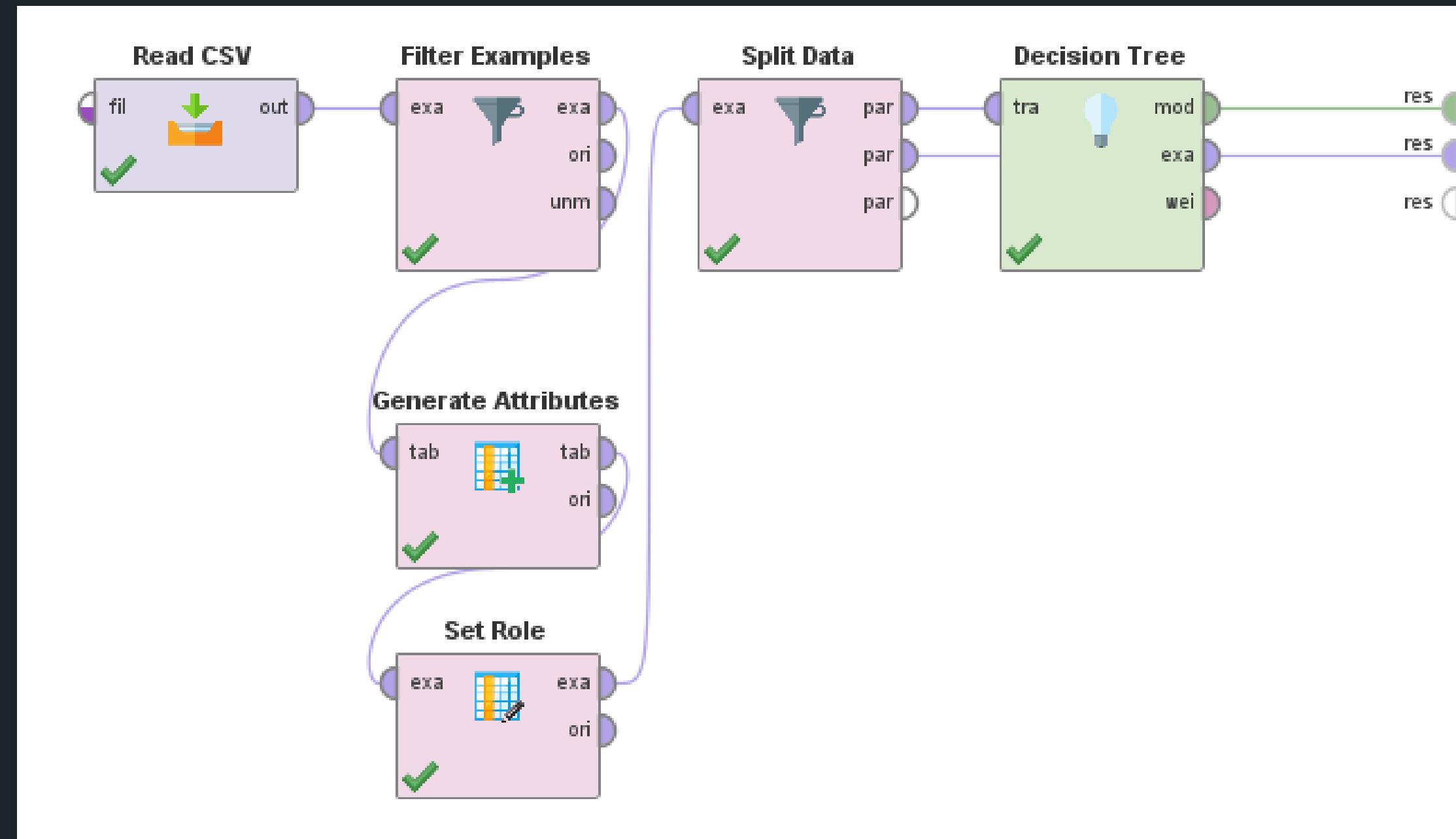


แบ่งข้อมูลออกเป็น 2 ส่วน โดยการใช้
Split Data : แบ่ง 70% ของข้อมูลกั้งหมดใช้ในการสร้างโมเดล
และแบ่ง 30% ของข้อมูลกั้งหมดใช้ในการทดสอบประสิทธิภาพ



Decision Tree

1 ใน 4 แยก ข้องการพยายามลดติด

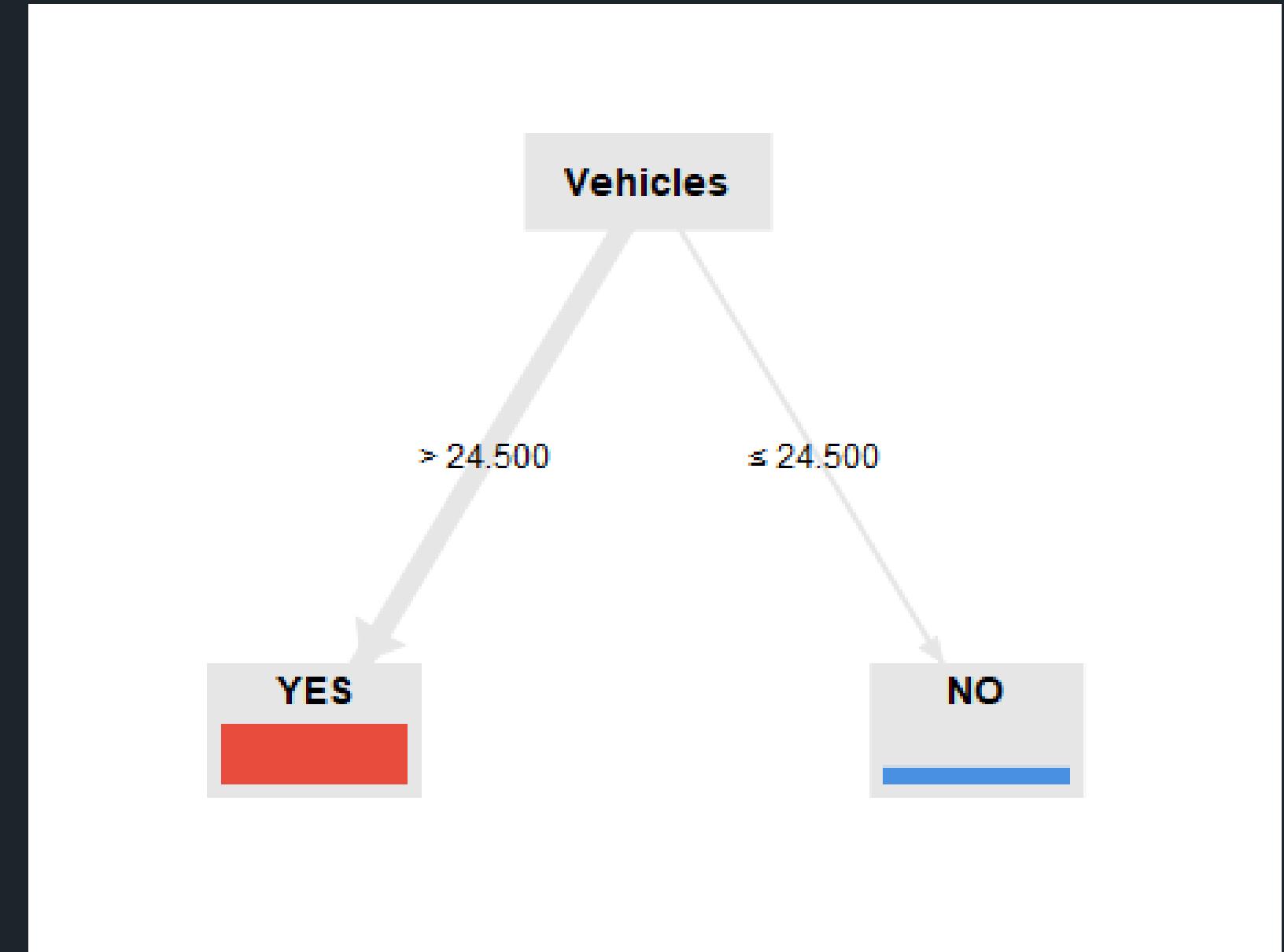


Presented by Malila

Decision Tree

ผลลัพธ์ของจำนวนรถว่ามีการรถติดหรือไม่

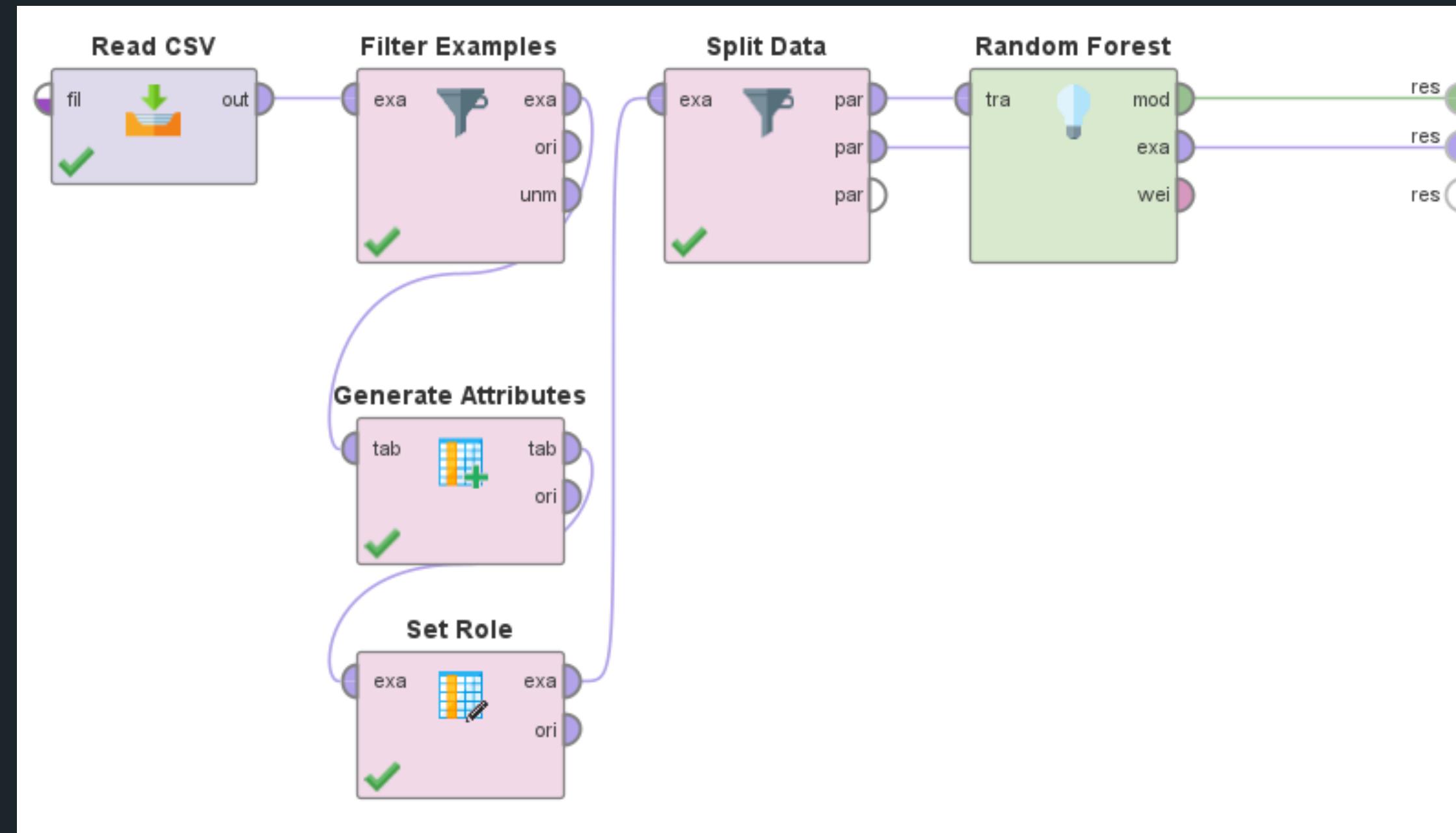
Row No.	Date	Time	Junction	Vehicles	result
1	Nov 1, 2015	3:00:00 AM	1	7	NO
2	Nov 1, 2015	5:00:00 AM	1	6	NO
3	Nov 1, 2015	11:00:00 AM	1	17	NO
4	Nov 1, 2015	6:00:00 PM	1	17	NO
5	Nov 1, 2015	7:00:00 PM	1	20	NO
6	Nov 1, 2015	9:00:00 PM	1	19	NO
7	Nov 2, 2015	12:00:00 AM	1	14	NO
8	Nov 2, 2015	1:00:00 AM	1	12	NO
9	Nov 2, 2015	5:00:00 AM	1	11	NO
10	Nov 2, 2015	6:00:00 AM	1	13	NO
11	Nov 2, 2015	7:00:00 AM	1	14	NO
12	Nov 2, 2015	8:00:00 AM	1	12	NO
13	Nov 2, 2015	9:00:00 AM	1	22	NO
14	Nov 2, 2015	10:00:00 AM	1	32	YES
15	Nov 2, 2015	11:00:00 AM	1	31	YES



Presented by Malila

Random Forest

1 ใน 4 แยก ข้องการพยายามลดติด

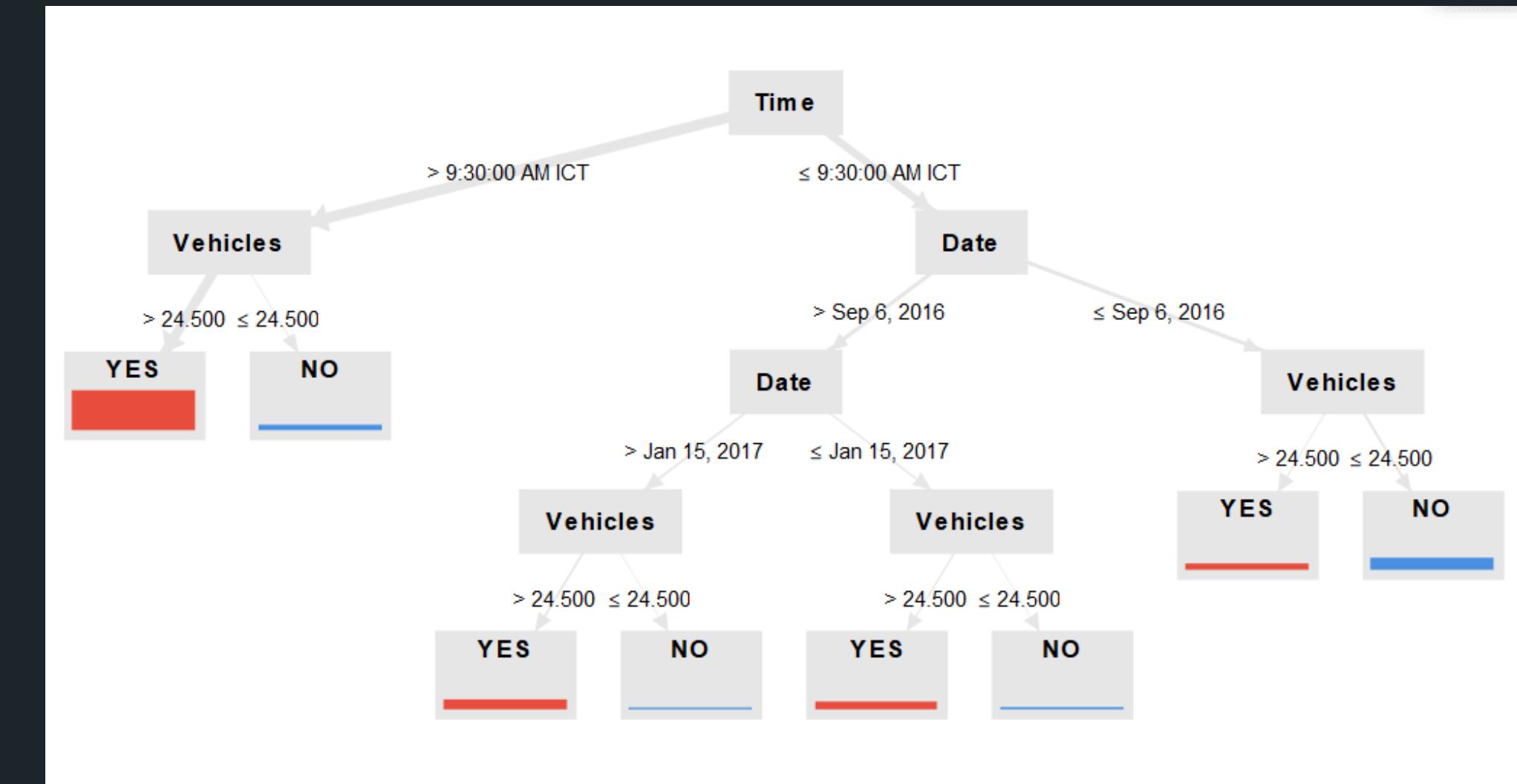


Presented by Malila

Random Forest

ผลลัพธ์ของจำนวนรถว่ามีการลดติดหรือไม่

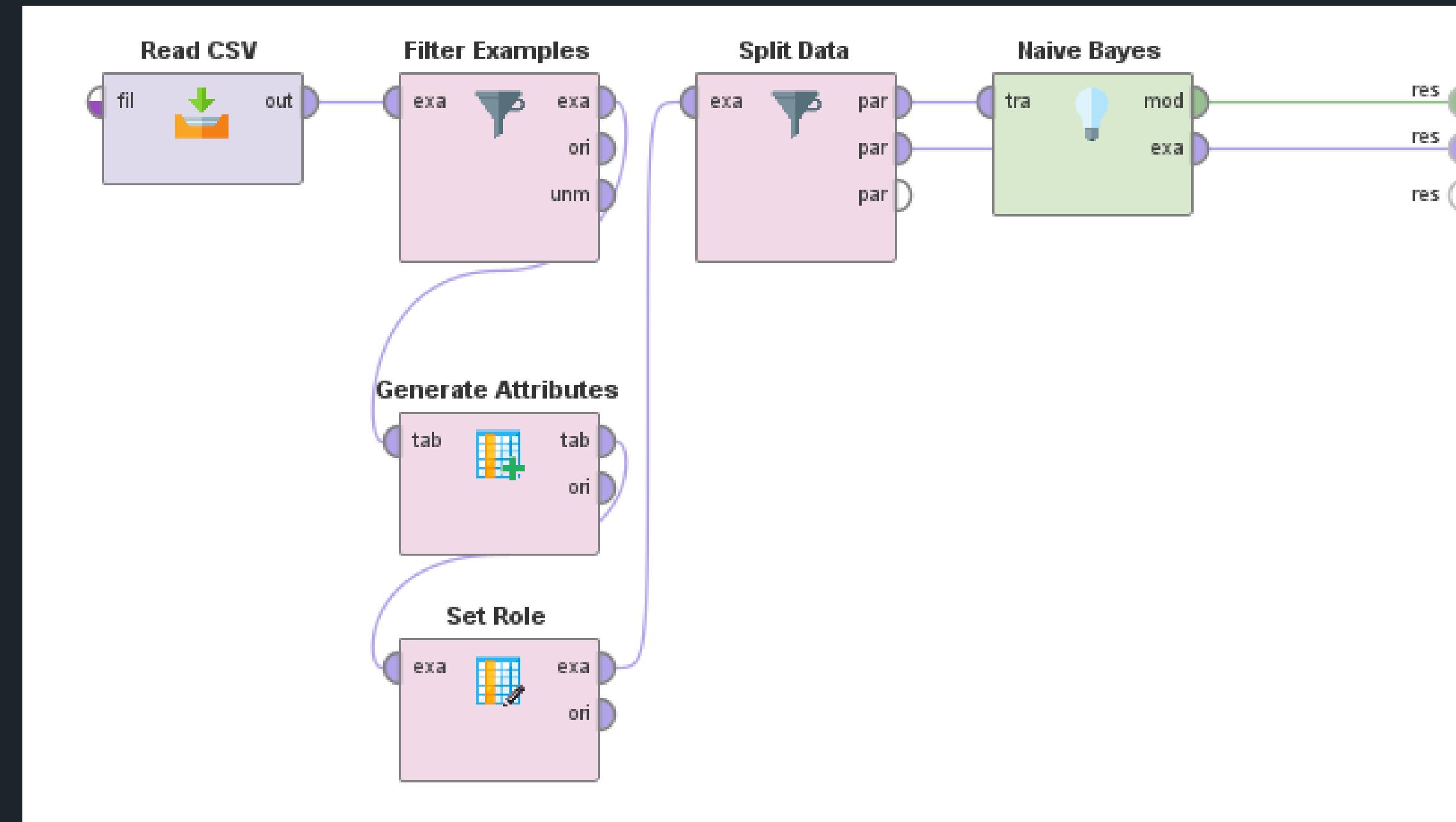
Row No.	Date	Time	Junction	Vehicles	result
1	Nov 1, 2015	3:00:00 AM	1	7	NO
2	Nov 1, 2015	5:00:00 AM	1	6	NO
3	Nov 1, 2015	11:00:00 AM	1	17	NO
4	Nov 1, 2015	6:00:00 PM	1	17	NO
5	Nov 1, 2015	7:00:00 PM	1	20	NO
6	Nov 1, 2015	9:00:00 PM	1	19	NO
7	Nov 2, 2015	12:00:00 AM	1	14	NO
8	Nov 2, 2015	1:00:00 AM	1	12	NO
9	Nov 2, 2015	5:00:00 AM	1	11	NO
10	Nov 2, 2015	6:00:00 AM	1	13	NO
11	Nov 2, 2015	7:00:00 AM	1	14	NO
12	Nov 2, 2015	8:00:00 AM	1	12	NO
13	Nov 2, 2015	9:00:00 AM	1	22	NO
14	Nov 2, 2015	10:00:00 AM	1	32	YES
15	Nov 2, 2015	11:00:00 AM	1	31	YES



Presented by Malila

Naive Bayes

1 ใน 4 แยก ข้องการพยายามลดติด

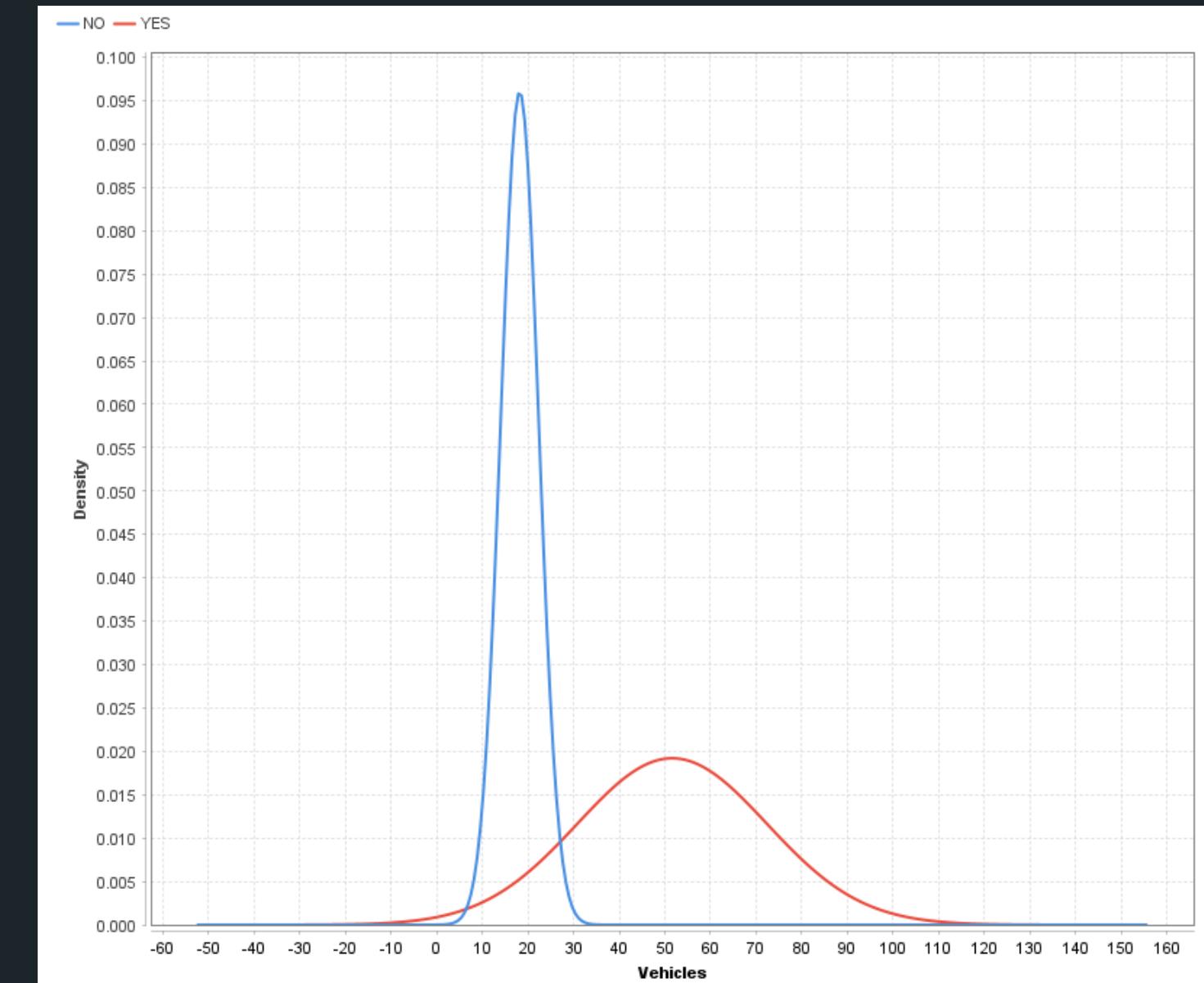


Presented by Malila

Naive Bayes

ผลลัพธ์ของจำนวนรถว่ามีการรถติดหรือไม่

Row No.	Date	Time	Junction	Vehicles	result
1	Nov 1, 2015	3:00:00 AM	1	7	NO
2	Nov 1, 2015	5:00:00 AM	1	6	NO
3	Nov 1, 2015	11:00:00 AM	1	17	NO
4	Nov 1, 2015	6:00:00 PM	1	17	NO
5	Nov 1, 2015	7:00:00 PM	1	20	NO
6	Nov 1, 2015	9:00:00 PM	1	19	NO
7	Nov 2, 2015	12:00:00 AM	1	14	NO
8	Nov 2, 2015	1:00:00 AM	1	12	NO
9	Nov 2, 2015	5:00:00 AM	1	11	NO
10	Nov 2, 2015	6:00:00 AM	1	13	NO
11	Nov 2, 2015	7:00:00 AM	1	14	NO
12	Nov 2, 2015	8:00:00 AM	1	12	NO
13	Nov 2, 2015	9:00:00 AM	1	22	NO
14	Nov 2, 2015	10:00:00 AM	1	32	YES
15	Nov 2, 2015	11:00:00 AM	1	31	YES



Presented by Malila

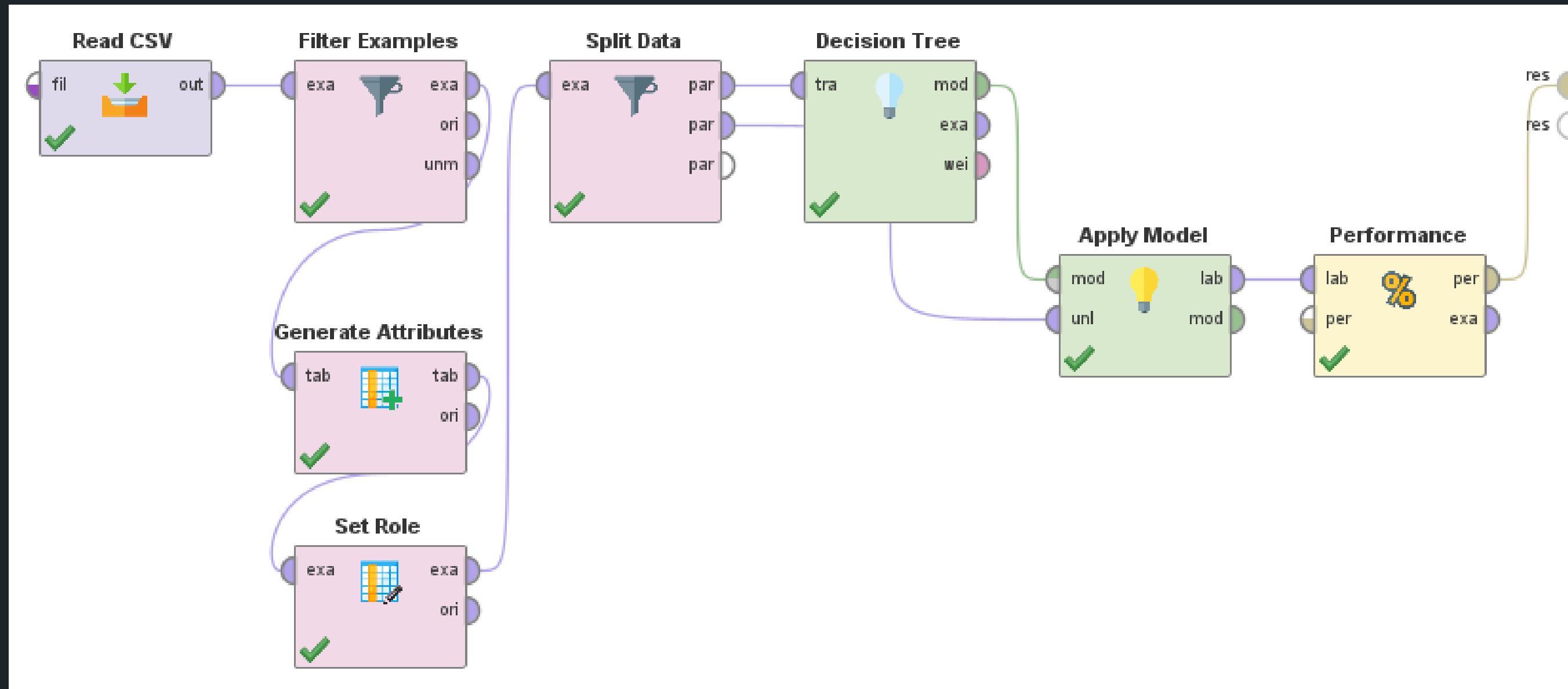
Evaluation

Assess models; report on the expected effects of models



Presented by Malila

Decision Tree



Presented by Malila

Decision Tree

	Accuracy	Squared Correlation
แยกที่ 1	100.00%	1.000
แยกที่ 2	100.00%	1.000
แยกที่ 3	100.00%	1.000
แยกที่ 4	100.00%	1.000

แยกที่ 1

accuracy: 100.00%			
	true NO	true YES	class precision
pred. NO	862	0	100.00%
pred. YES	0	3515	100.00%
class recall	100.00%	100.00%	

squared_correlation

squared_correlation: 1.000

แยกที่ 2

accuracy: 100.00%			
	true NO	true YES	class precision
pred. NO	3953	0	100.00%
pred. YES	0	425	100.00%
class recall	100.00%	100.00%	

squared_correlation

squared_correlation: 1.000

แยกที่ 3

accuracy: 100.00%			
	true NO	true YES	class precision
pred. NO	3913	0	100.00%
pred. YES	0	464	100.00%
class recall	100.00%	100.00%	

squared_correlation

squared_correlation: 1.000

แยกที่ 4

accuracy: 100.00%			
	true NO	true YES	class precision
pred. NO	1300	0	100.00%
pred. YES	0	3	100.00%
class recall	100.00%	100.00%	

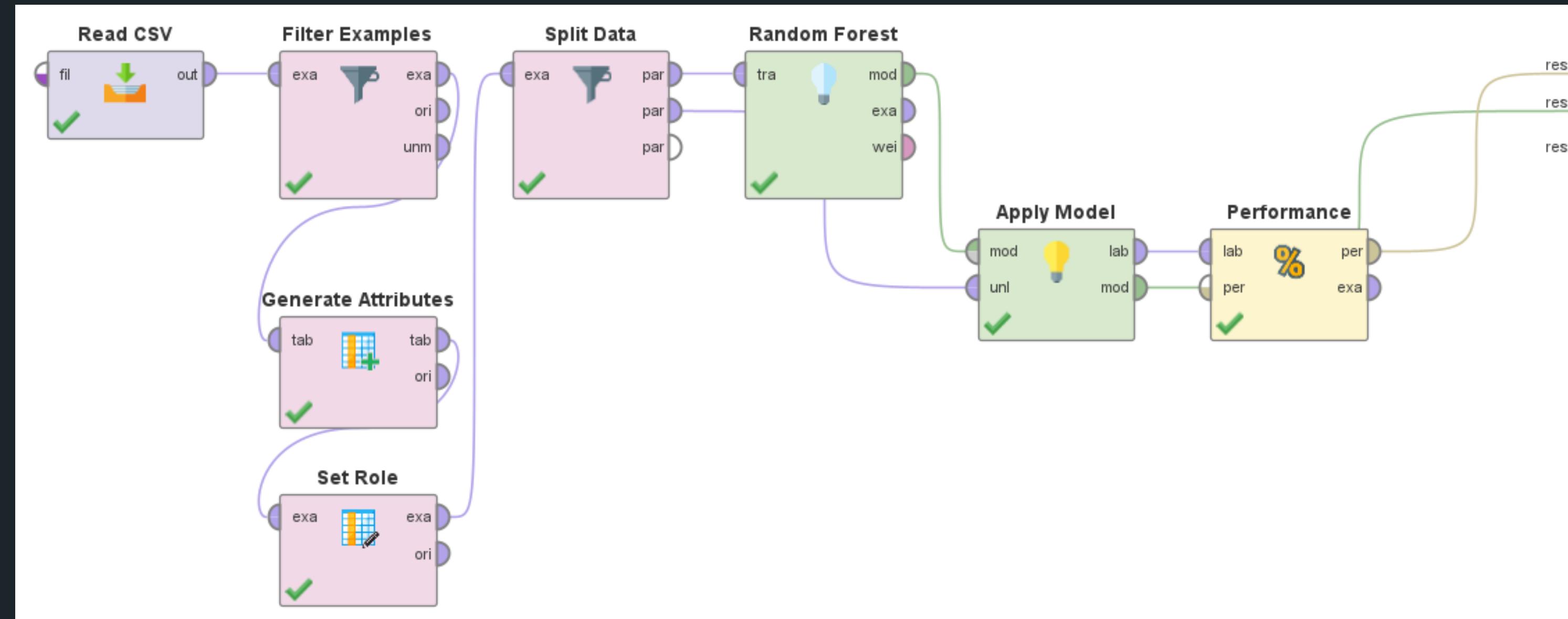
squared_correlation

squared_correlation: 1.000



Presented by Malila

Random Forest



Presented by Malila



Random Forest

	Accuracy	Squared Correlation
แยกที่ 1	100.00%	1.000
แยกที่ 2	100.00%	1.000
แยกที่ 3	100.00%	1.000
แยกที่ 4	100.00%	1.000

แยกที่ 1

accuracy: 100.00%			
	true NO	true YES	class precision
pred. NO	862	0	100.00%
pred. YES	0	3515	100.00%
class recall	100.00%	100.00%	

squared_correlation

squared_correlation: 1.000

แยกที่ 2

accuracy: 100.00%			
	true NO	true YES	class precision
pred. NO	3953	0	100.00%
pred. YES	0	425	100.00%
class recall	100.00%	100.00%	

squared_correlation

squared_correlation: 1.000

แยกที่ 3

accuracy: 100.00%			
	true NO	true YES	class precision
pred. NO	3913	0	100.00%
pred. YES	0	464	100.00%
class recall	100.00%	100.00%	

squared_correlation

squared_correlation: 1.000

แยกที่ 4

accuracy: 100.00%			
	true NO	true YES	class precision
pred. NO	1300	0	100.00%
pred. YES	0	3	100.00%
class recall	100.00%	100.00%	

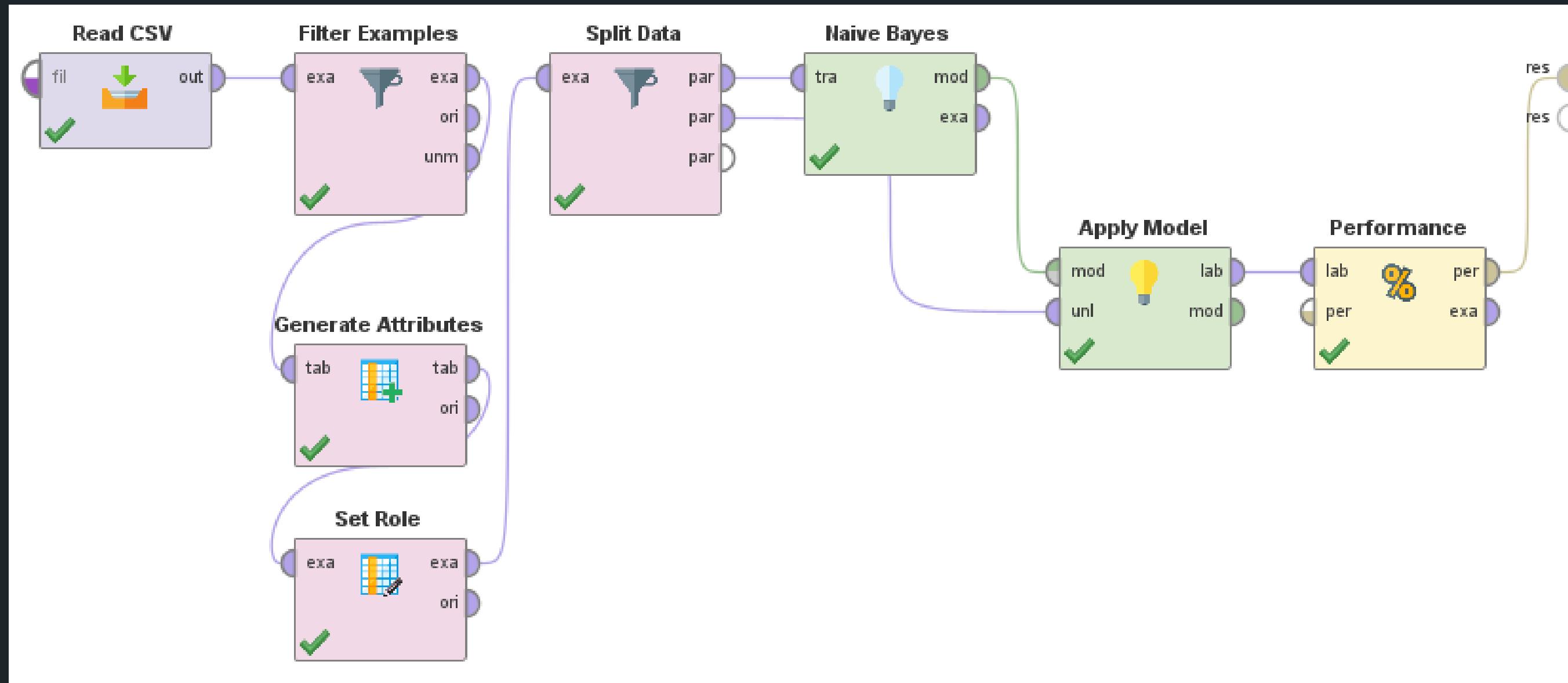
squared_correlation

squared_correlation: 1.000



Presented by Malila

Naive Bayes



Presented by Malila

Naive Bayes

	Accuracy	Squared Correlation
แยกที่ 1	96.55%	0.804
แยกที่ 2	98.08%	0.816
แยกที่ 3	98.54%	0.851
แยกที่ 4	99.54%	0.332

แยกที่ 1

accuracy: 96.55%

	true NO	true YES	class precision
pred. NO	831	120	87.38%
pred. YES	31	3395	99.10%
class recall	96.40%	96.59%	

squared_correlation

squared_correlation: 0.804

แยกที่ 2

accuracy: 98.08%

	true NO	true YES	class precision
pred. NO	3871	2	99.95%
pred. YES	82	423	83.76%
class recall	97.93%	99.53%	

squared_correlation

squared_correlation: 0.816

แยกที่ 3

accuracy: 98.54%

	true NO	true YES	class precision
pred. NO	3885	36	99.08%
pred. YES	28	428	93.86%
class recall	99.28%	92.24%	

squared_correlation

squared_correlation: 0.851

แยกที่ 4

accuracy: 99.54%

	true NO	true YES	class precision
pred. NO	1294	0	100.00%
pred. YES	6	3	33.33%
class recall	99.54%	100.00%	

squared_correlation

squared_correlation: 0.332



Presented by Malila

เปรียบเทียบประสิทธิภาพของแต่ละโมเดล

	Decision Tree		Random Forest		Naive Bayes	
	Accuracy	Squared Correlation	Accuracy	Squared Correlation	Accuracy	Squared Correlation
แยกที่ 1	100.00%	1.000	100.00%	1.000	96.55%	0.804
แยกที่ 2	100.00%	1.000	100.00%	1.000	98.08%	0.816
แยกที่ 3	100.00%	1.000	100.00%	1.000	98.54%	0.851
แยกที่ 4	100.00%	1.000	100.00%	1.000	99.54%	0.332



Presented by Malila

CPE408 - BIG DATA AND AI

THANK YOU



Presented by Malila