



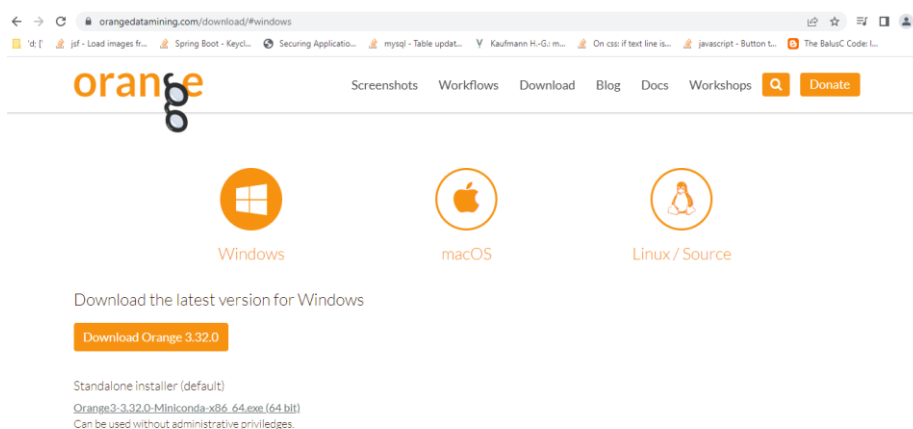
Pada modul kali ini kita akan mencoba Instalasi Tools untuk Data Mining yaitu Orange Data Mining. Data Mining salah satu teknik implementasi dari pembelajaran mesin (*machine learning*) yang saat ini banyak digunakan di berbagai bidang. Data mining menjadi salah satu primadona teknik pemodelan dan pengambilan keputusan berdasarkan data yang dapat menghasilkan wawasan (*insight*) dan perkiraan/prediksi (*forecast*) yang dapat membantu kita di kehidupan sehari-hari. Dengan aplikasi Orange Data Mining kita bisa melakukan proses data mining tanpa perlu menulis kode program sama sekali.

### A. Orange Data Mining

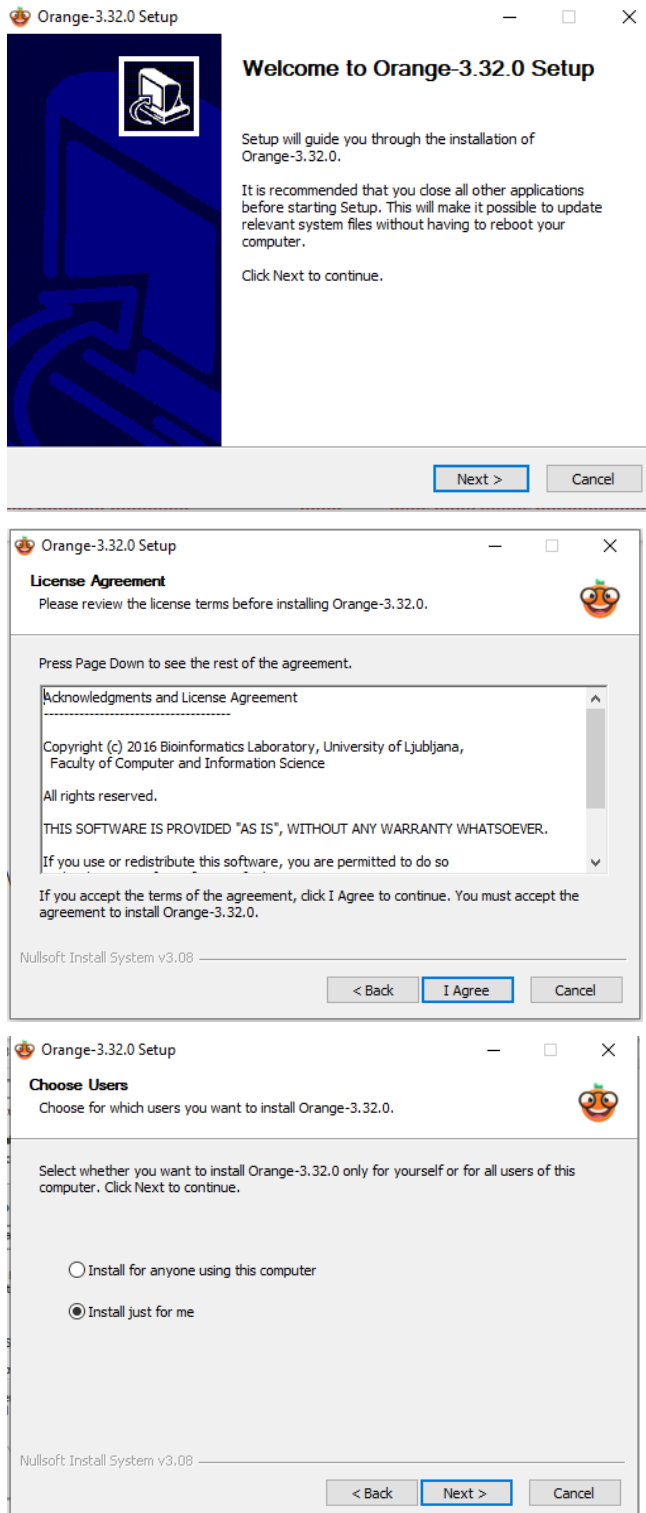
Orange merupakan aplikasi yang dikembangkan oleh University of Ljubljana, dengan target menghadirkan sebuah aplikasi untuk mengolah data secara visual tanpa perlu pengalaman membuat program sebelumnya. Pada Orange, untuk melakukan data mining kita akan menggunakan sistem widget. Setiap widget punya fungsinya masing-masing dan dapat menerima input atau mengeluarkan output. Misalnya kita ingin membaca data dari sebuah file, kita bisa menggunakan widget File untuk membaca data dari file tersebut dan menggunakan widget Data Table untuk menampilkan isi data yang sudah dibaca. Sama halnya jika kita ingin membuat sebuah model untuk melakukan data mining seperti klasifikasi atau regresi. Kita bisa menambahkan widget seperti K-nearest neighbors (KNN) dan Naive Bayes untuk membuat model dan memberikan data ke dalam model tersebut dengan cara menghubungkan sumber data dengan model dengan cara menarik garis penghubung.

### B. Instalasi Orange Data Mining

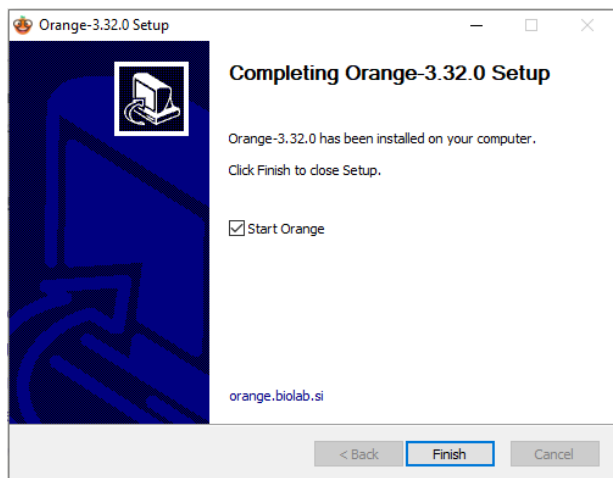
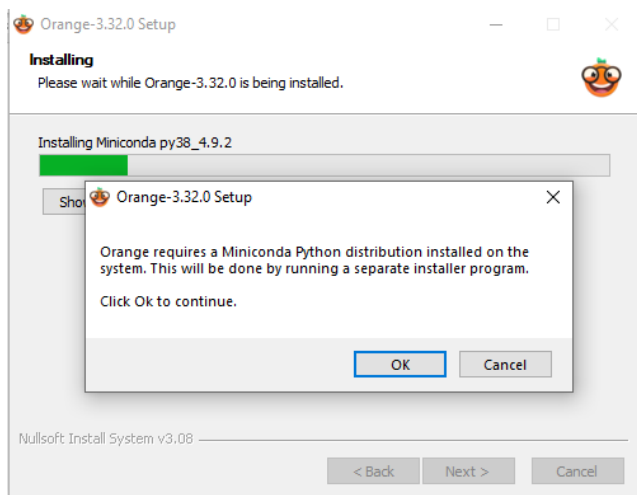
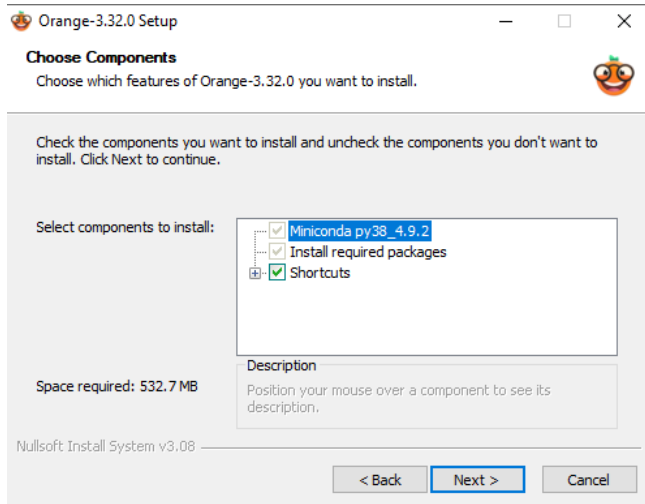
Untuk unduh Orange Data Mining bisa melalui alamat URL berikut ini <https://orangedatamining.com/download/#windows>. Pada modul ini menggunakan sistem operasi windows, selain operasi windows silakan disesuaikan dengan sistem operasi masing-masing.



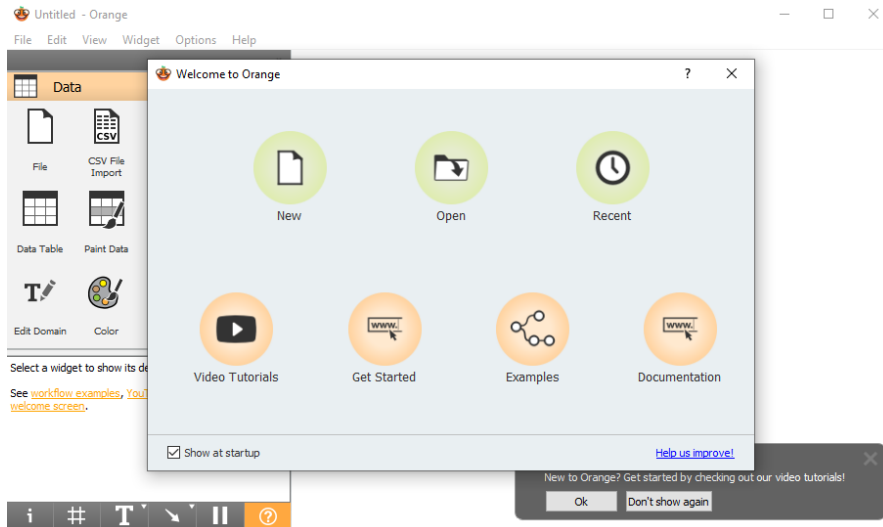
Setelah selesai unduh silakan di double klik file nya lalu ikuti tahap-tahap selanjutnya sampai aplikasi sukses terinstall.



Modul Praktikum 8 : Instalasi Orange Data Mining

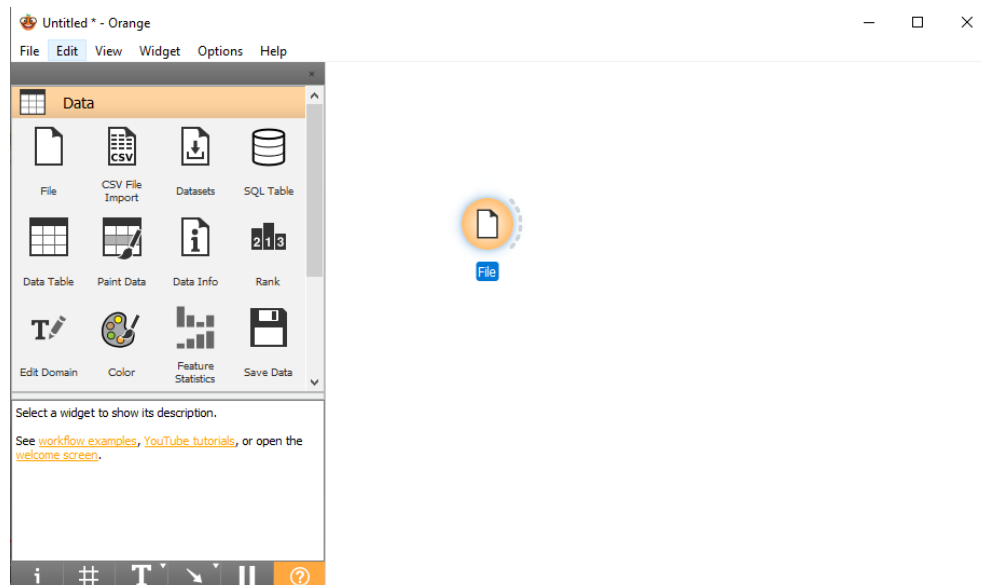


Berikut tampilan awal orange

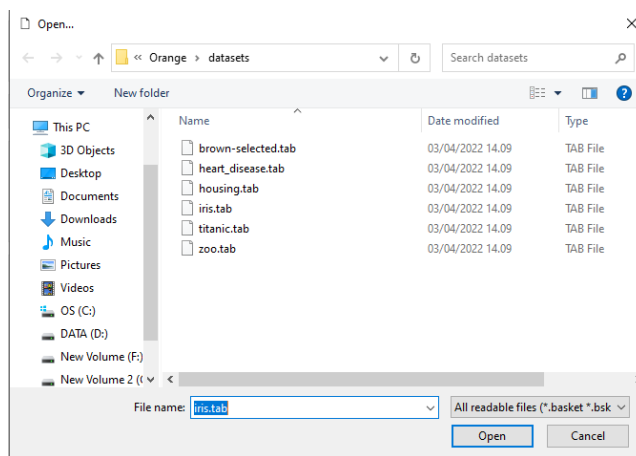
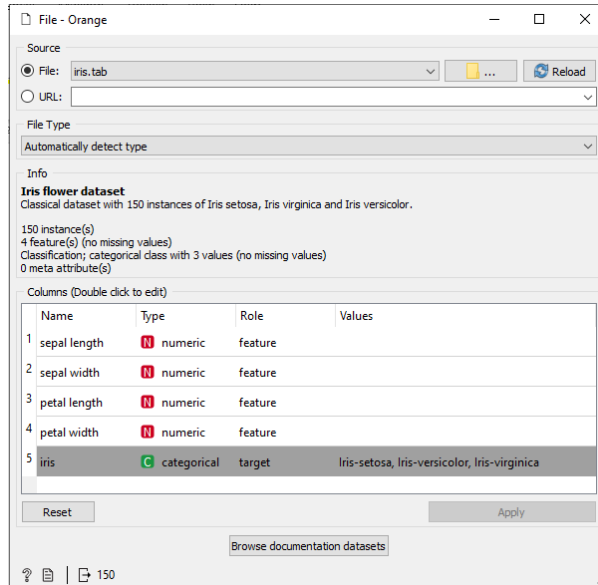


### C. Penggunaan Orange Data Mining

- Pilih widget data kemudian klik file.

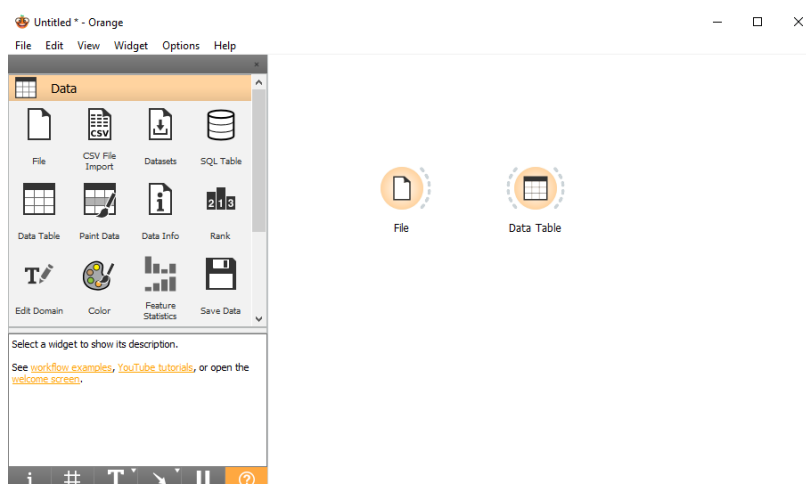


- Double klik file tersebut lalu pilih file dan sumber data yang diinginkan, disini kita menggunakan data set yang sudah disediakan oleh orange.



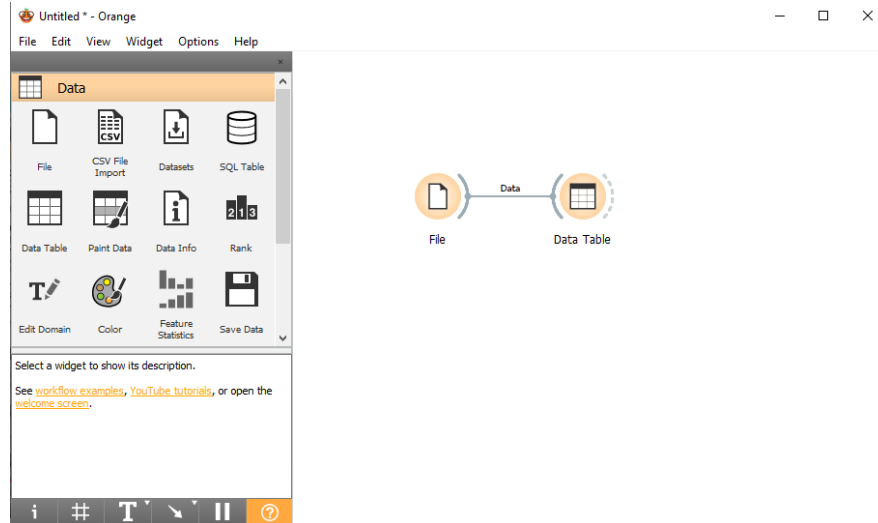
Pilih iris.tab kemudian setelah dipilih klik tanda silang (close) dialog file tersebut

- Klik widget Data Table



- Lalu tarik garis dari kanan (output) widget file ke kiri (input) widget Data Table

## Modul Praktikum 8 : Instalasi Orange Data Mining



- Kemudian klik Data Table, akan tampil data berupa tabel

Data Table - Orange

Info  
150 instances (no missing data)  
4 features  
Target with 3 values  
No meta attributes

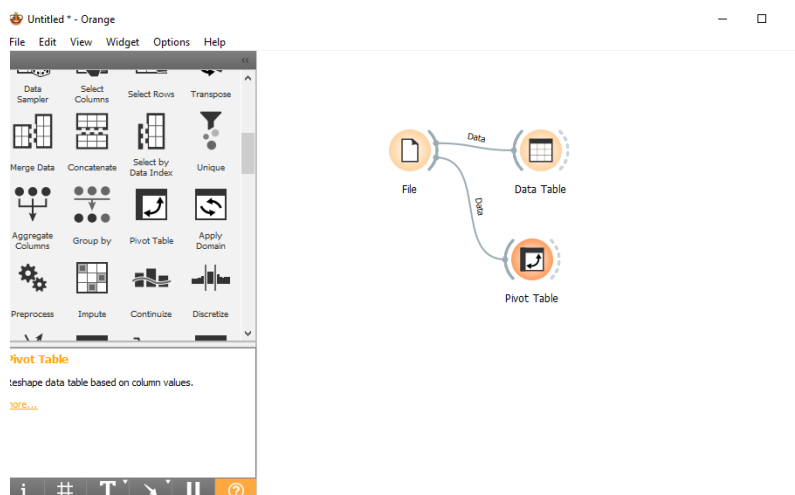
Variables  
☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

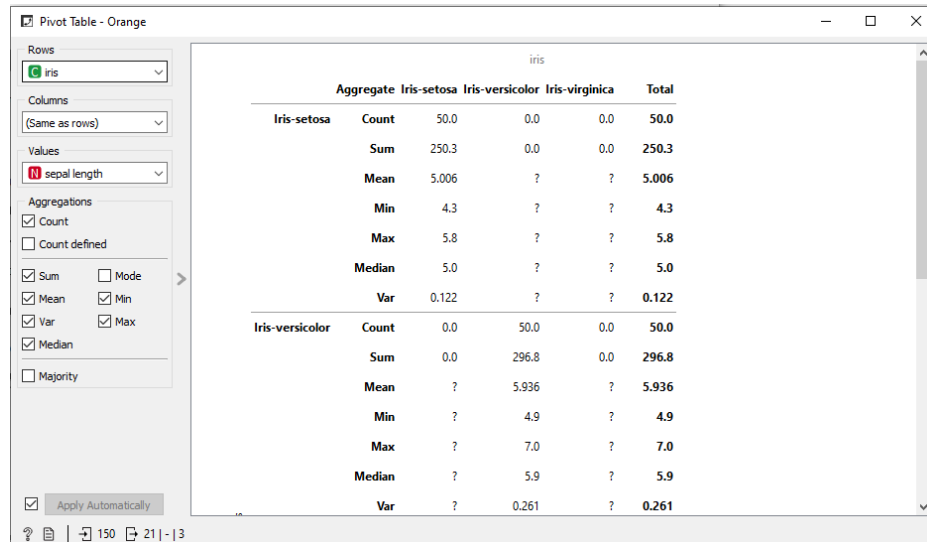
Selection  
☒ Select full rows

Restore Original Order  
☒ Send Automatically

	iris	sepal length	sepal width	petal length	petal width
1	Iris-setosa	5.1	3.5	1.4	0.2
2	Iris-setosa	4.9	3.0	1.4	0.2
3	Iris-setosa	4.7	3.2	1.3	0.2
4	Iris-setosa	4.6	3.1	1.5	0.2
5	Iris-setosa	5.0	3.6	1.4	0.2
6	Iris-setosa	5.4	3.9	1.7	0.4
7	Iris-setosa	4.6	3.4	1.4	0.3
8	Iris-setosa	5.0	3.4	1.5	0.2
9	Iris-setosa	4.4	2.9	1.4	0.2
10	Iris-setosa	4.9	3.1	1.5	0.1
11	Iris-setosa	5.4	3.7	1.5	0.2
12	Iris-setosa	4.8	3.4	1.6	0.2
13	Iris-setosa	4.8	3.0	1.4	0.1
14	Iris-setosa	4.3	3.0	1.1	0.1
15	Iris-setosa	5.8	4.0	1.2	0.2
16	Iris-setosa	5.7	4.4	1.5	0.4
17	Iris-setosa	5.4	3.9	1.3	0.4
18	Iris-setosa	5.1	3.5	1.4	0.3
19	Iris-setosa	5.7	3.8	1.7	0.3
20	Iris-setosa	5.1	3.8	1.5	0.3
21	Iris-setosa	5.4	3.4	1.7	0.2

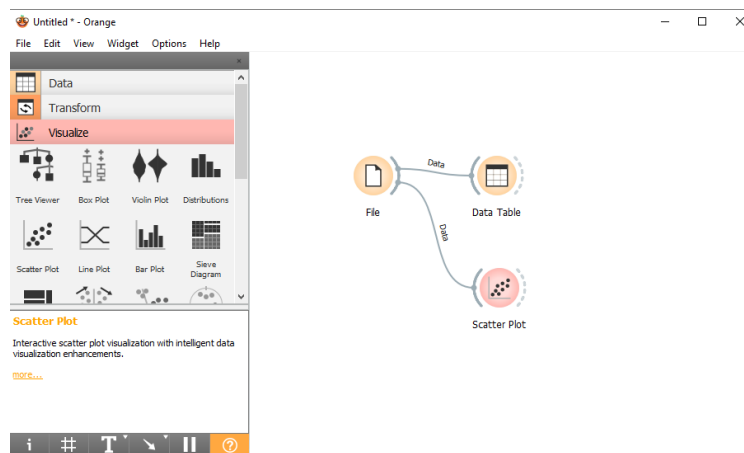
- Tambahkan widget pivot table, lalu hubungkan garis dari kanan (output) widget file ke kiri (input) widget pivot table dan double klik widget pivot table.

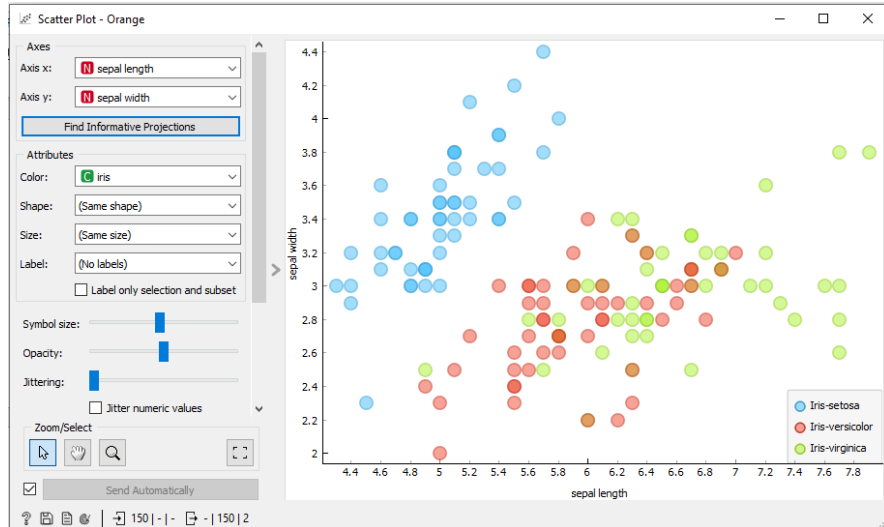




Silakan ceklist count, sum, mean, median, min, max sehingga tampil seperti gambar diatas.

- Tambahkan widget Visualize Scatter Plot, lalu hubungkan garis dari kanan (output) widget file ke kiri (input) widget Scatter Plot dan double klik widget Scatter Plot.





Akan terlihat distribusi data yang ditampilkan dalam diagram.

- Tambahkan widget Visualize Distribution, lalu hubungkan garis dari kanan (output) widget file ke kiri (input) widget Distribution dan double klik widget Distribution.

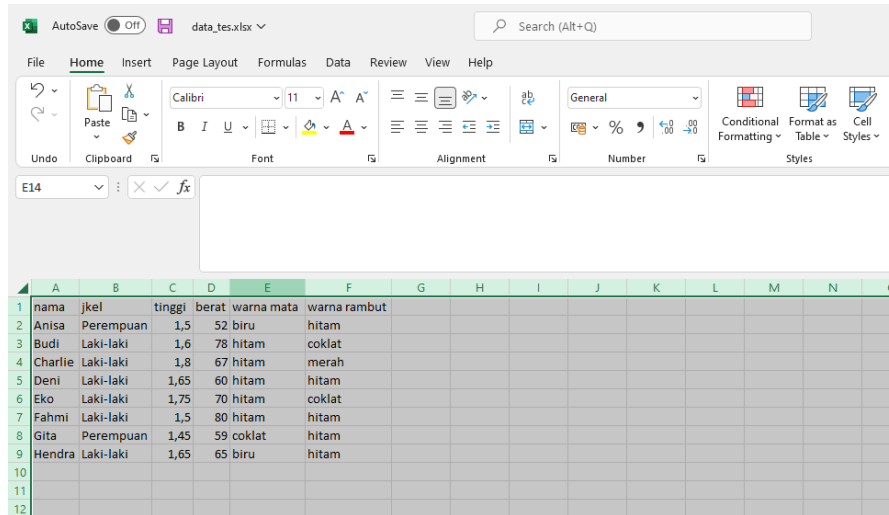


- Tambahkan widget Visualize Distribution, lalu hubungkan garis dari kanan (output) widget file ke kiri (input) widget Distribution dan double klik widget Distribution.



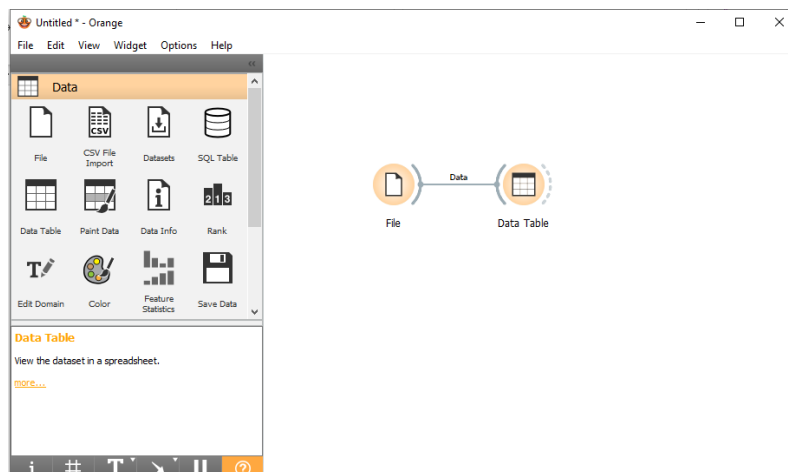
### D. Loading Data Orange Data Mining

- Buat file excel seperti dibawah ini

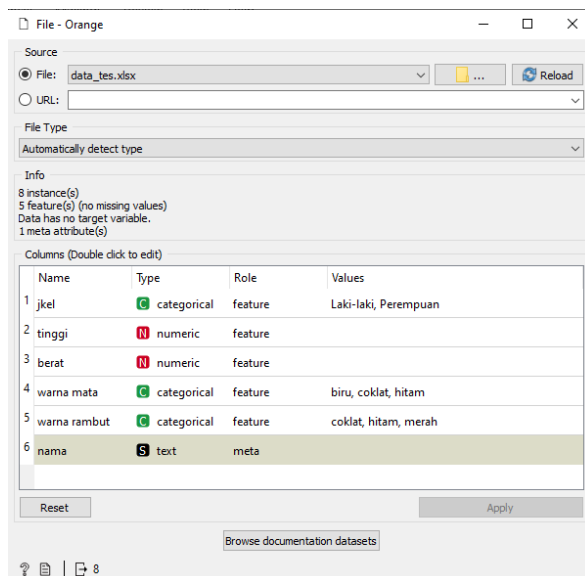


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	nama	jkel	tinggi	berat	warna mata	warna rambut									
2	Anisa	Perempuan	1,5	52	biru	hitam									
3	Budi	Laki-laki	1,6	78	hitam	coklat									
4	Charlie	Laki-laki	1,8	67	hitam	merah									
5	Deni	Laki-laki	1,65	60	hitam	hitam									
6	Eko	Laki-laki	1,75	70	hitam	coklat									
7	Fahmi	Laki-laki	1,5	80	hitam	hitam									
8	Gita	Perempuan	1,45	59	coklat	hitam									
9	Hendra	Laki-laki	1,65	65	biru	hitam									
10															
11															
12															

- Lalu buat kan pada orange widget file dan data table.



- Double klik pada widget file, kemudian pilih file excel yang tadi sudah dibuat



File - Orange

Source

☒ File: data\_tes.xlsx ☐ URL:

File Type

Automatically detect type

Info

8 instance(s)  
5 feature(s) (no missing values)  
Data has no target variable.  
1 meta attribute(s)

Columns (Double click to edit)

Name	Type	Role	Values
1 jkel	categorical	feature	Laki-laki, Perempuan
2 tinggi	numeric	feature	
3 berat	numeric	feature	
4 warna mata	categorical	feature	biru, coklat, hitam
5 warna rambut	categorical	feature	coklat, hitam, merah
6 nama	text	meta	

Reset Apply

Browse documentation datasets

## Modul Praktikum 8 : Instalasi Orange Data Mining

- Kemudian close lalu klik data table maka akan terlihat data yang ada di excel.

Data Table - Orange

Info  
8 instances (no missing data)  
5 features  
No target variable.  
1 meta attribute

Variables  
☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

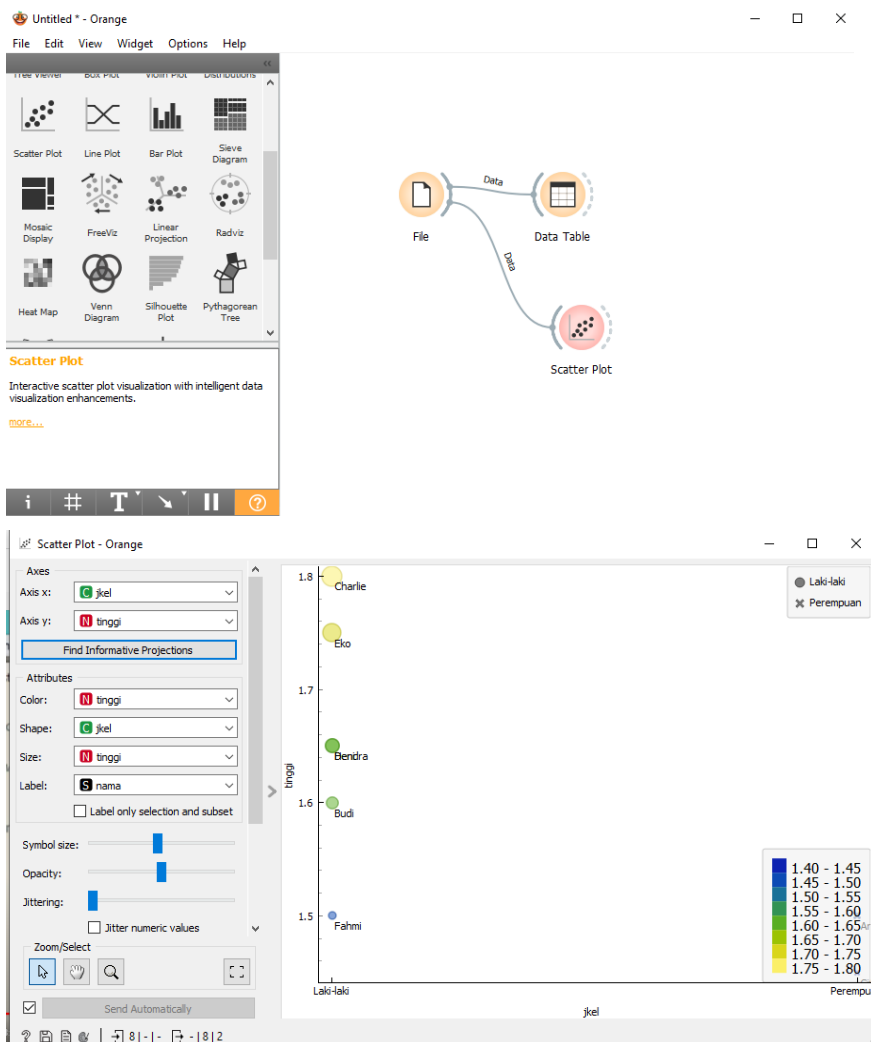
Selection  
☒ Select full rows

Restore Original Order

☒ Send Automatically

	nama	jenis	tinggi	berat	warna mata	warna rambut
1	Anisa	Perempuan	1.50	52	biru	hitam
2	Budi	Laki-laki	1.60	78	hitam	coklat
3	Charlie	Laki-laki	1.80	67	hitam	merah
4	Deni	Laki-laki	1.65	60	hitam	hitam
5	Eko	Laki-laki	1.75	70	hitam	coklat
6	Fahmi	Laki-laki	1.50	80	hitam	hitam
7	Gita	Perempuan	1.45	59	coklat	hitam
8	Hendra	Laki-laki	1.65	65	biru	hitam

- Tambahkan widget scatter plot lalu bisa dilihat data ditampilkan dalam bentuk diagram.



## E. Making Predictions

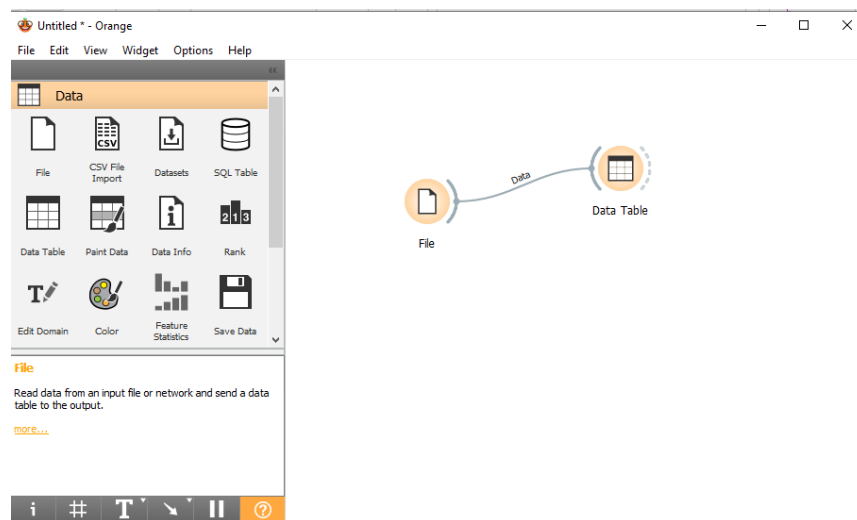
- Download data set berikut ini untuk melakukan prediksi dengan *classification tree* dan *logistic regression*

<https://raw.githubusercontent.com/ajdapretnar/datasets/master/data/fruits-and-vegetables-train.tab>

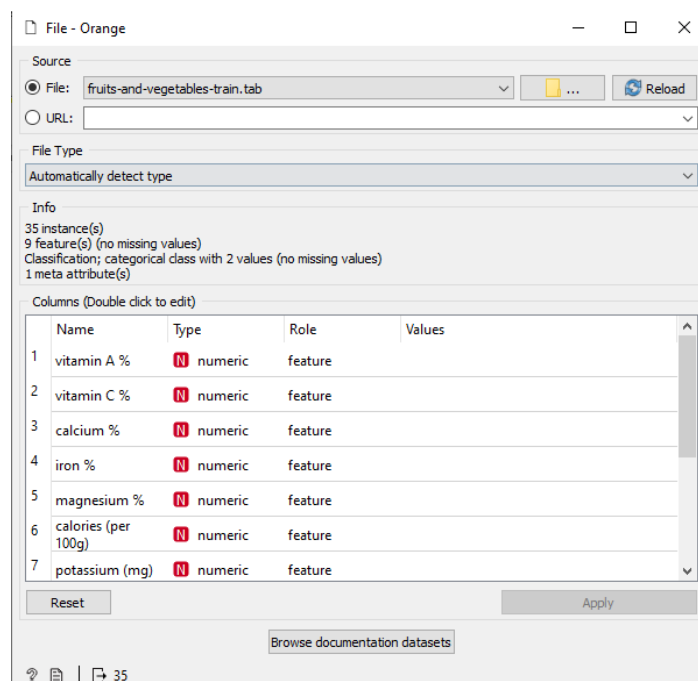
<https://raw.githubusercontent.com/ajdapretnar/datasets/master/data/fruits-and-vegetables-test.tab>

simpan dengan ekstensi .tab

- Buat file new baru untuk membuat widget seperti berikut



- Sama seperti sebelumnya ambil data set pada widget file yaitu fruits-and-vegetables-train.tab



## Modul Praktikum 8 : Instalasi Orange Data Mining

- Double klik widget Data Table maka isi data akan tampil seperti ini.

Data Table - Orange

Info  
35 instances (no missing data)  
9 features  
Target with 2 values  
1 meta attribute

Variables  
☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

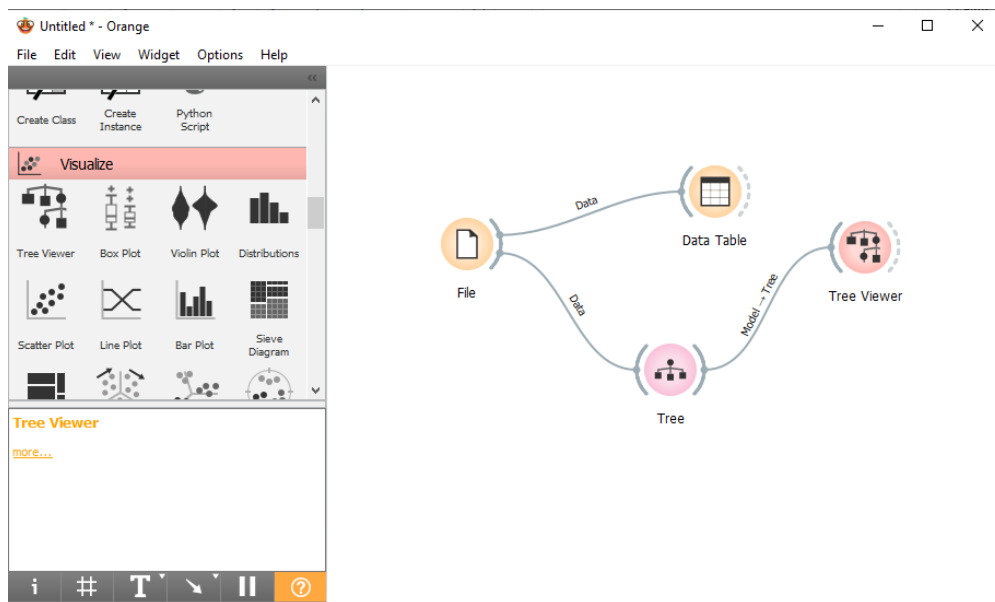
Selection  
☒ Select full rows

Restore Original Order

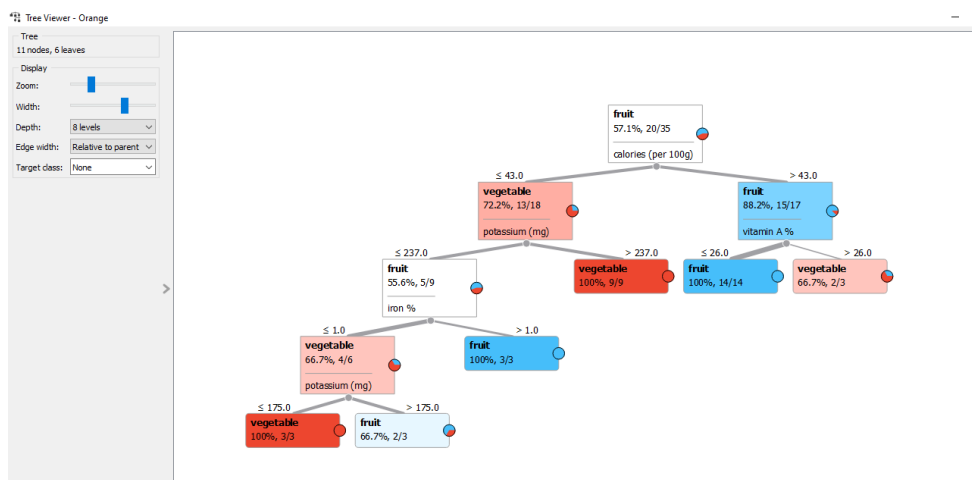
☒ Send Automatically

	classification	name	vitamin A %	vitamin C %	calcium %
1	fruit	apple	1.0	7.0	0.0
2	fruit	apricot	38.0	16.0	1.0
3	fruit	avocado	2.0	16.0	1.0
4	fruit	banana	1.0	14.0	0.0
5	vegetable	beetroot	0.0	8.0	1.0
6	fruit	blackberry	4.0	35.0	2.0
7	fruit	blueberry	1.0	16.0	0.0
8	vegetable	broccoli	12.0	148.0	4.0
9	vegetable	brussels sprouts	15.0	141.0	4.0
10	vegetable	carrot	334.0	9.0	3.0
11	vegetable	cauliflower	0.0	80.0	2.0
12	vegetable	celery	8.0	5.0	4.0
13	fruit	cherry	26.0	16.0	1.0
14	vegetable	cucumber	2.0	4.0	1.0
15	vegetable	eggplant	0.0	3.0	0.0
16	fruit	elderberry	12.0	60.0	3.0
17	fruit	grape	2.0	6.0	1.0
18	vegetable	leek	33.0	20.0	5.0
19	fruit	lemon	0.0	88.0	2.0
20	fruit	mango	21.0	60.0	1.0
21	fruit	orange	4.0	88.0	4.0
22	fruit	pineapple	1.0	79.0	1.0

- Tambahkan widget tree dan tree viewer sehingga menjadi seperti ini

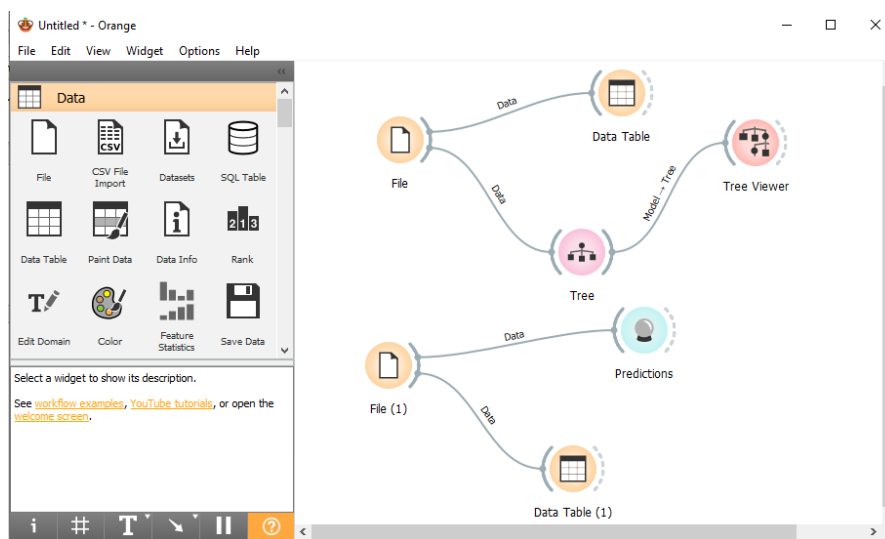


- Double klik tree viewer nanti akan terlihat klasifikasi data sesuai data set

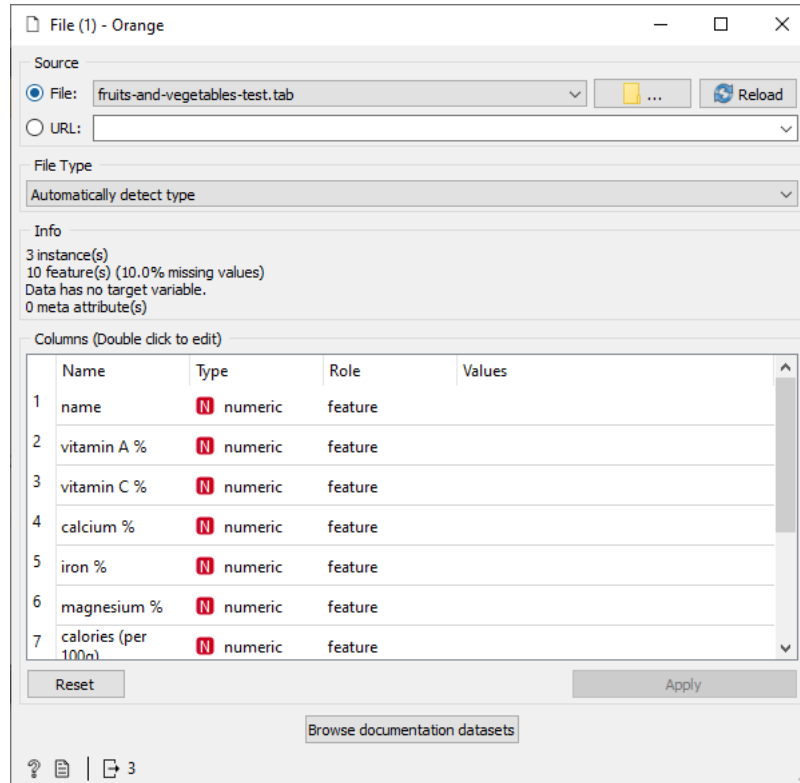


Pada gambar diatas terlihat klasifikasi fruit dan vegetables sesuai dengan nilai masing-masing parameter. Pertama adalah kalori, jika kalori lebih kecil sama dengan 43 maka masuk kategori vegetables lalu jika lebih besar 43 masuk kategori fruit. Nilai 43 adalah nilai yang tertera pada dataset untuk kalori pada fruit dan vegetables. Kemudian selanjutnya adalah klasifikasi sesuai parameter yang lain.

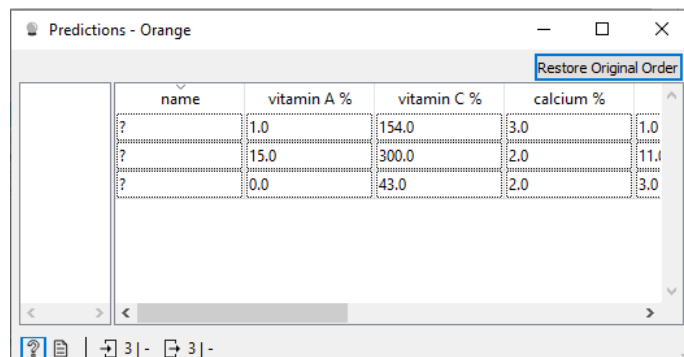
- Selanjutnya tambahkan dibagian bawah seperti bagan berikut, buat widget file, data table dan predictions untuk melakukan *logistic regression*.



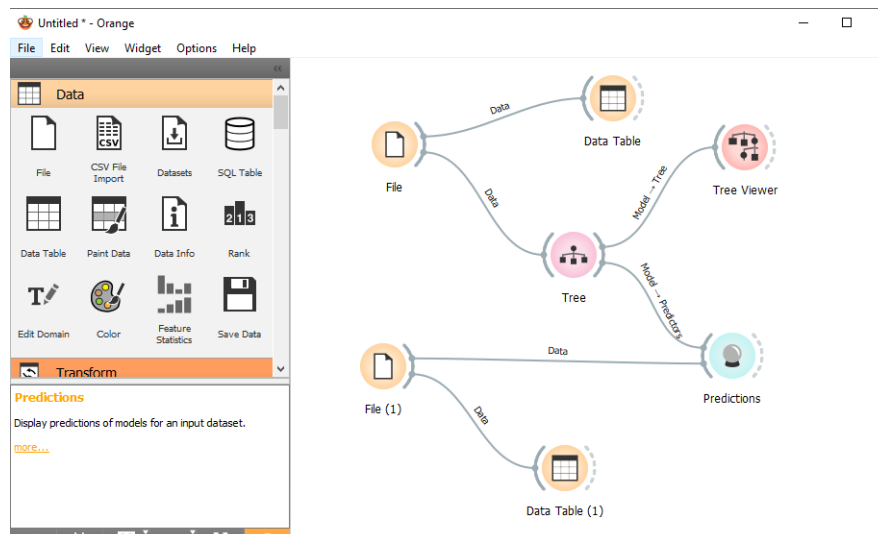
- Masukkan data set pada file (1) fruits-and-vegetables-test.tab



- Double klik predictions



- Sambungkan output tree ke input dari predictions seperti gambar berikut



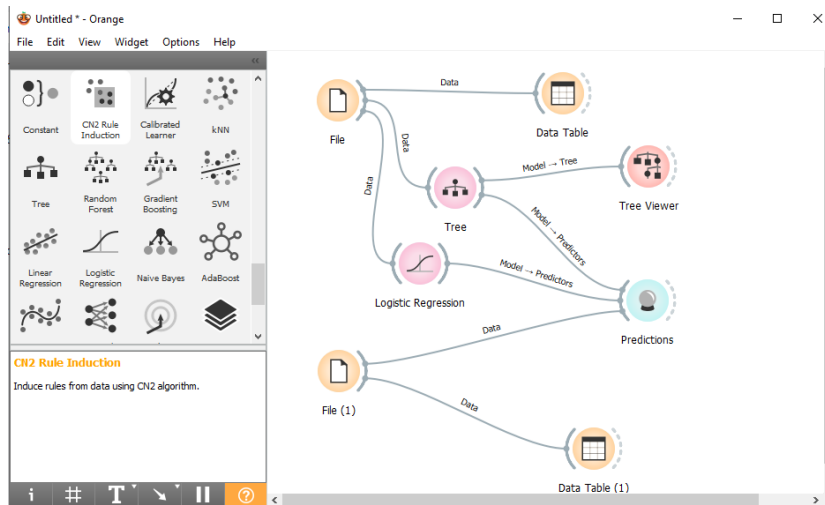
## Modul Praktikum 8 : Instalasi Orange Data Mining

- Lakukan double klik kembali pada predictions

Predictions - Orange

Tree	name	vitamin A %	vitamin C %	calcium %	iron %	magnesium %	calories (per 100g)	potassium (mg)	protein (g)	fiber (g)
1 fruit	?	1.0	154.0	3.0	1.0	4.0	61.0	213.0	1.1	3.0
2 fruit	?	15.0	300.0	2.0	11.0	3.0	20.0	202.0	2.2	2.1
3 fruit	?	0.0	43.0	2.0	3.0	5.0	53.0	151.0	1.1	7.0

- Tambahkan logistic regression sesuai gambar lalu lihat hasilnya pada predictions



Predictions - Orange

Tree	Logistic Regression	name	vitamin A %	vitamin C %	calcium %	iron %	magnesium %	calories (per 100g)	potassium (mg)	protein (g)	fiber (g)
1 fruit	fruit	?	1.0	154.0	3.0	1.0	4.0	61.0	213.0	1.1	3.0
2 fruit	vegetable	?	15.0	300.0	2.0	11.0	3.0	20.0	202.0	2.2	2.1
3 fruit	fruit	?	0.0	43.0	2.0	3.0	5.0	53.0	151.0	1.1	7.0

Terlihat prediksi untuk data menggunakan tree dan juga logistic regression. Dari sini bisa diambil kesimpulan dengan tools orange ini user bisa melakukan beberapa proses terkait data mining selama data set tersedia.

----- Sekian Terima Kasih -----

- <https://orangedatamining.com/>