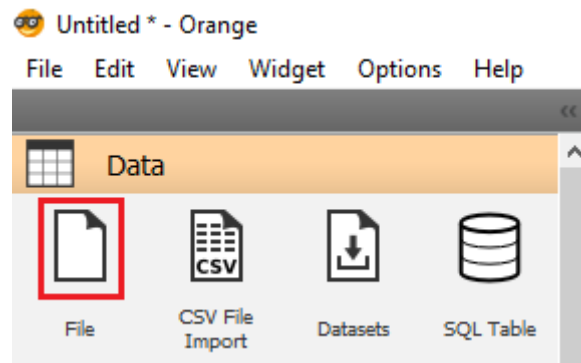


Langkah pertama yang perlu dilakukan adalah dengan melakukan *input data* ke Orange terlebih dahulu melalui fitur File yang dapat dilihat sebagai berikut.



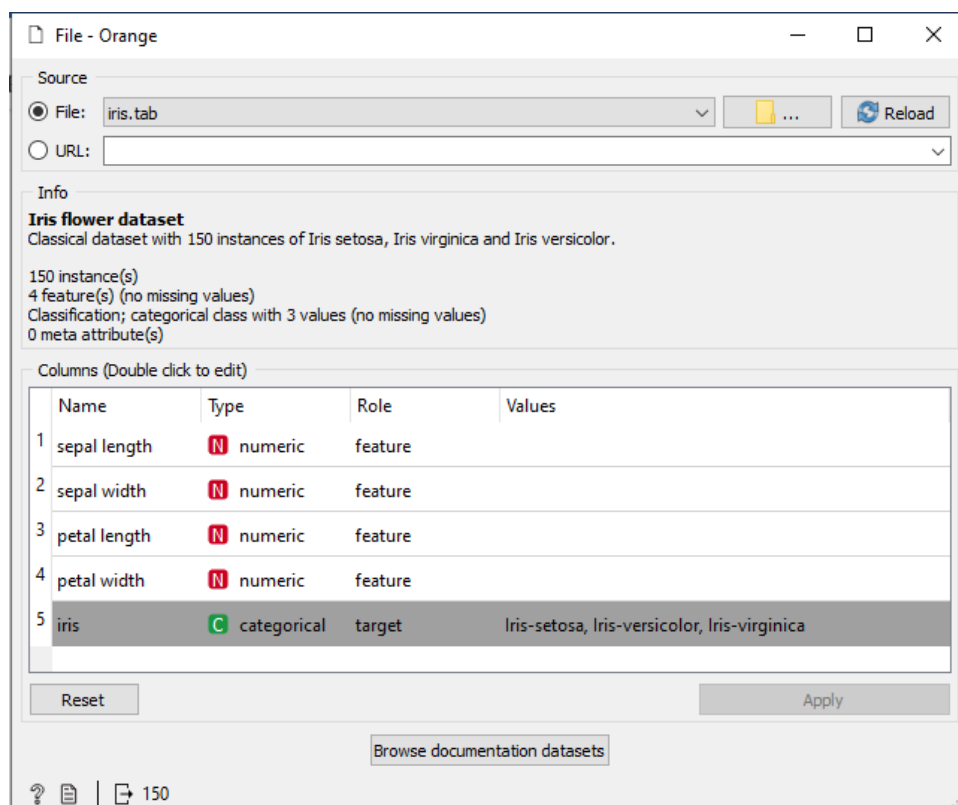
Gambar 2. 1 Fitur *File*

Langkah selanjutnya, lakukan klik pada fitur *File*, sehingga pada lembar kerja Orange akan muncul secara otomatis ikon File sebagai berikut.



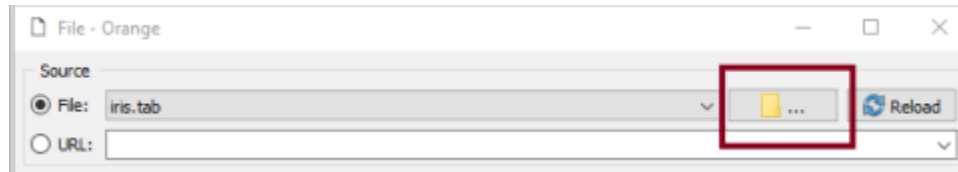
Gambar 2. 2 ikon Fitur *File*

Selanjutnya, lakukan klik ganda pada ikon *File* sehingga muncul tampilan sebagai berikut.



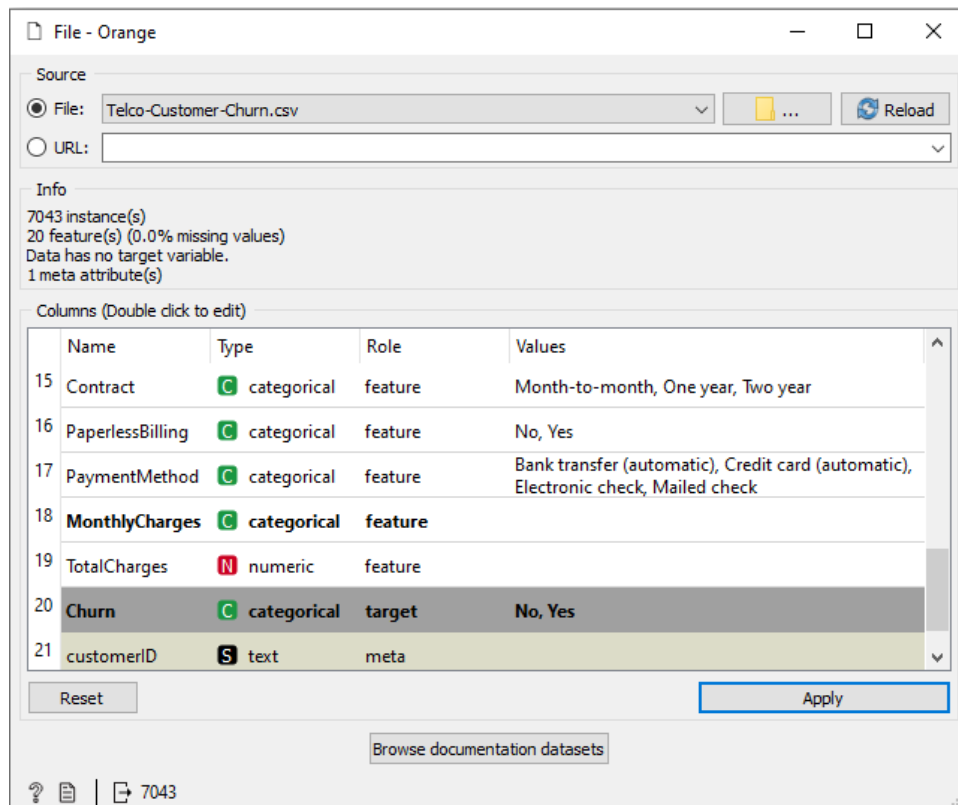
Gambar 2. 3 Tampilan setelah melakukan klik ganda pada ikon fitur *File*

Lalu, pilih *file* data Telco-Customer-Churn yang tersimpan pada folder komputer dengan mengklik logo folder di sebelah kanan atas sebagai berikut.



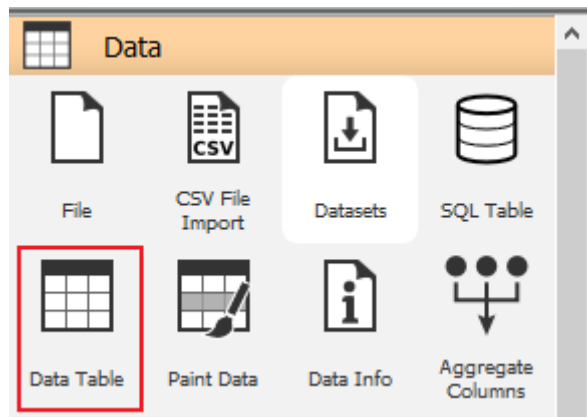
Gambar 2. 4 Klik logo *folder*

Sehingga nama Source File yang tampil sudah sesuai dengan data yang akan dilakukan analisis seperti pada gambar berikut. Lalu, ganti *role* untuk kolom Churn dari *feature* menjadi *target*, tahapan ini dilakukan untuk memastikan bahwa target pemetaan adalah benar.



Gambar 2. 5 Tampilan akhir pada ikon fitur *File*

Langkah selanjutnya, lakukan klik pada fitur Data Table sebagai berikut.



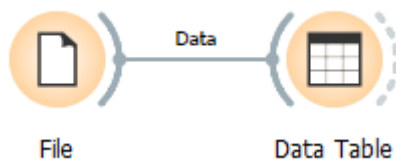
Gambar 2. 6 Fitur Data Table

Sehingga pada lembar kerja Orange akan terlihat sebagai berikut.



Gambar 2. 7 Menambahkan ikon Data Table

Lalu, membuat garis hubung dari ikon File ke ikon Data Tabel dengan cara mengklik lengkungan kanan pada ikon File, lalu arahkan ke lengkungan kiri pada ikon Data Table sebagai berikut.



Gambar 2. 8 Menyambungkan ikon File dengan ikon Data Table

Lalu, untuk melihat data yang sudah di input dapat dilakukan klik ganda pada ikon *Data Table* sehingga data dapat terlihat dalam bentuk tabel sebagai berikut.

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup
1	7580-WH9EG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes
2	3375-W9WEG	Male	0	No	No	34	Yes	No	DSL	Yes	No
3	3668-QP9BK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes
4	7795-CF0CW	Male	0	No	No	45	No	No phone service	DSL	Yes	No
5	8237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No
6	9335-C25AC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No
7	1432-K0VK	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes
8	6713-OKDMC	Female	0	No	No	10	No	No phone service	DSL	Yes	No
9	7892-P00KP	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No
10	6388-S0M0U	Male	0	No	Yes	42	Yes	No	DSL	Yes	Yes
11	8763-GR5GD	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No
12	7489-LK8IC	Male	0	No	No	16	Yes	No	No	No internet ser...	No internet ser...
13	8091-TNAX	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No
14	5230-KJEX	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes
15	5126-LUPS	Male	0	No	No	25	Yes	No	Fiber optic	Yes	No
16	3655-DNQYZ	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	Yes
17	8191-WXZG	Female	0	No	No	52	Yes	No	No	No internet ser...	No internet ser...
18	9959-W0RKT	Male	0	No	Yes	71	Yes	Yes	Fiber optic	Yes	No
19	4190-MFL0W	Female	0	Yes	Yes	10	Yes	No	DSL	No	No
20	4183-MYFRB	Female	0	No	No	21	Yes	No	Fiber optic	No	Yes
21	8779-Q8DMV	Male	1	No	No	1	No	No phone service	DSL	No	No
22	1680-V0CWV	Male	0	Yes	No	12	Yes	No	No	No internet ser...	No internet ser...
23	5966-W5XG	Male	0	No	No	1	Yes	No	No	No internet ser...	No internet ser...
24	3638-WEABW	Female	0	Yes	No	58	Yes	Yes	DSL	No	Yes
25	8322-H9PFA	Male	0	Yes	Yes	49	Yes	No	DSL	Yes	Yes
26	8885-D2W0	Female	0	No	No	30	Yes	No	DSL	Yes	Yes
27	6887-CV2W	Male	0	Yes	No	47	Yes	Yes	Fiber optic	No	Yes
28	8665-UTD4Z	Male	0	Yes	Yes	1	No	No phone service	DSL	No	Yes
29	5248-YGRN	Male	0	Yes	No	72	Yes	Yes	DSL	Yes	Yes
30	8773-H4K0Z	Female	0	No	Yes	17	Yes	No	DSL	No	No

Gambar 2. 9 Tampilan data dalam bentuk tabel

Setelah data ditampilkan, selanjutnya dilakukan deskripsi terhadap setiap atribut yang ada pada *dataset*, penjelasannya adalah sebagai berikut.

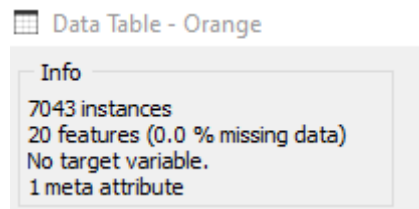
Tabel 2. 1 Deskripsi Atribut

Atribut	Deskripsi
customerID	Kode unik pelanggan
Gender	Jenis kelamin pelanggan
SeniorCitizen	Menunjukkan pelanggan berusia 65 tahun ke atas atau tidak
Partner	Menunjukkan pelanggan memiliki pasangan atau tidak
Dependents	Menunjukkan pelanggan tinggal dengan tanggungan atau tidak
tenure	Periode pelanggan berlangganan
PhoneService	Menunjukkan pelanggan berlangganan layanan telepon atau tidak
MultipleLines	Menunjukkan pelanggan berlangganan beberapa saluran telepon atau tidak
InternetService	Menunjukkan pelanggan berlangganan layanan internet dengan tipe tertentu atau tidak
OnlineSecurity	Menunjukkan pelanggan berlangganan layanan keamanan tambahan atau tidak
OnlineBackup	Menunjukkan pelanggan berlangganan layanan <i>online backup</i> tambahan atau tidak
DeviceProtection	Menunjukkan pelanggan berlangganan layanan proteksi perangkat tambahan atau tidak
TechSupport	Menunjukkan pelanggan berlangganan layanan <i>technical support</i> tambahan atau tidak
StreamingTV	Menunjukkan pelanggan menggunakan layanan internet untuk <i>streaming</i> TV
StreamingMovies	Menunjukkan pelanggan menggunakan layanan internet untuk <i>streaming</i> film
Contract	Tipe kontrak pelanggan
PaperlessBilling	Menunjukkan pelanggan memilih tagihan non-cetak

PaymentMethod	Tipe pembayaran yang digunakan pelanggan
MonthlyCharges	Tagihan yang diterima pelanggan perbulan
TotalCharges	Total tagihan yang diterima pelanggan dari awal berlangganan
Churn	Menunjukkan pelanggan berhenti berlangganan atau tidak

2.1.3 Data Preprocessing

Dari informasi yang didapatkan pada info Data Table di sebelah kiri atas, dapat terlihat bahwa tidak ada *missing data* sebagai berikut.



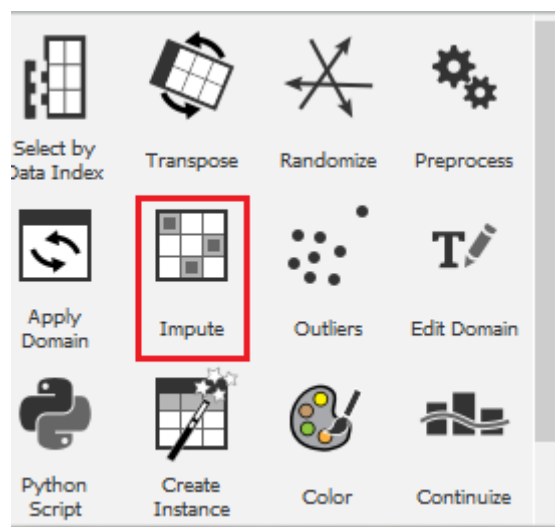
Gambar 2. 10 Info data

Tetapi setelah dilakukan *sort* terhadap masing-masing kolom yang ada, ditemukan untuk kolom TotalCharges terdapat *missing value* sebanyak 11 baris sebagai berikut.

	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
4381	No internet serv...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Two year	No	Mailed check	20.00	?	No
937	Yes	Yes	No	Yes	Yes	Two year	No	Mailed check	80.85	?	No
6671	Yes	Yes	Yes	Yes	No	Two year	No	Mailed check	73.35	?	No
1341	Yes	Yes	Yes	Yes	No	Two year	No	Credit card (aut...	56.05	?	No
489	No	Yes	Yes	Yes	No	Two year	Yes	Bank transfer (a...	52.55	?	No
754	No internet serv...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Two year	No	Mailed check	20.25	?	No
6755	Yes	No	Yes	No	No	Two year	Yes	Bank transfer (a...	61.90	?	No
3332	No internet serv...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Two year	No	Mailed check	19.85	?	No
5219	No internet serv...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	One year	Yes	Mailed check	19.70	?	No
1083	No internet serv...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Two year	No	Mailed check	25.75	?	No
3827	No internet serv...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Two year	No	Mailed check	25.35	?	No

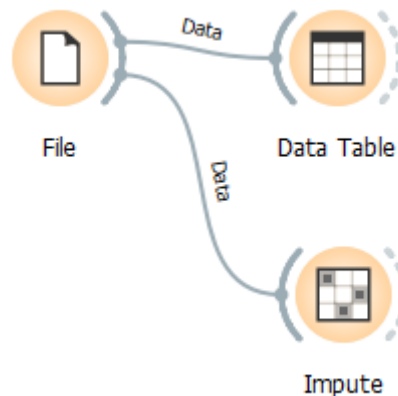
Gambar 2. 11 Tampilan data dengan *missing data*

Oleh karena itu, perlu dilakukan penghapusan terhadap 11 baris yang memiliki *missing value* pada kolom TotalCharges menggunakan fitur *Impute* sebagai berikut, hal ini dilakukan agar data yang akan dianalisis memiliki kualitas yang lebih baik.



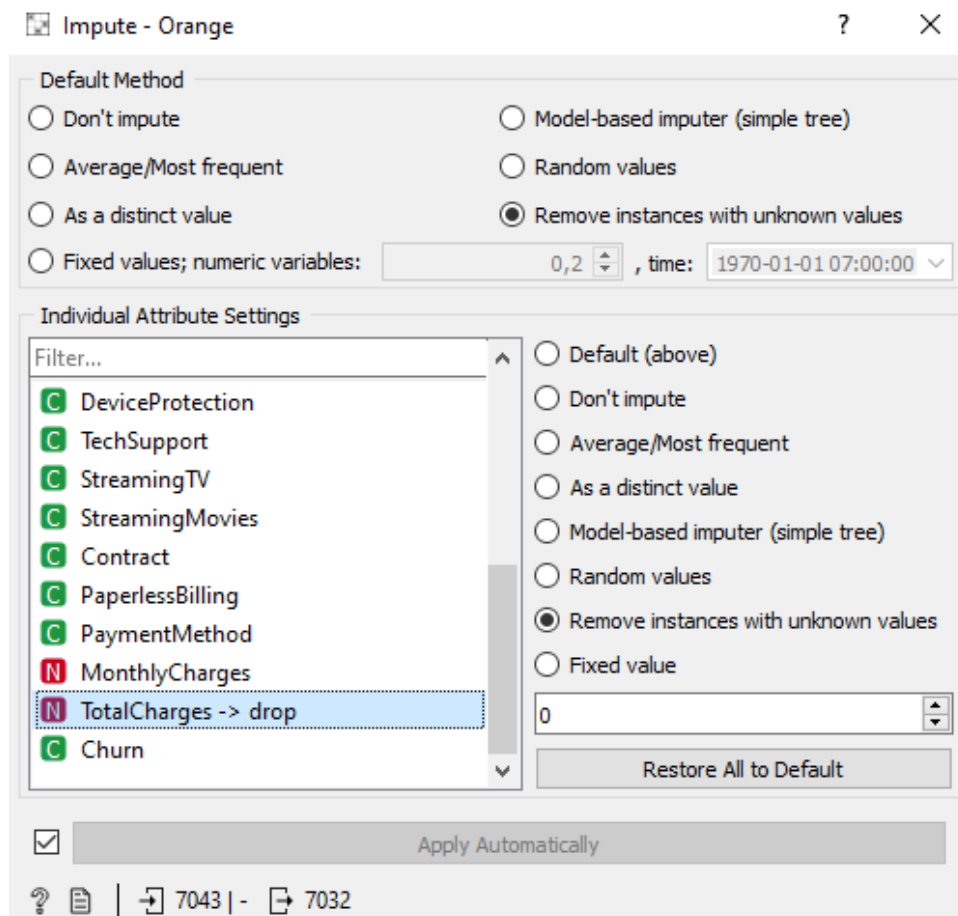
Gambar 2. 12 fitur *Impute*

Lalu, klik fitur *Impute* dan hubungkan ikon File dengan ikon Impute sebagai berikut.



Gambar 2. 13 Hubungkan fitur *impute*

Kemudian ditampilkan halaman *Impute* dan pilih atribut TotalCharges. Untuk menghilangkan data yang tidak memiliki nilai, dapat memilih “*Remove instances with unknown values*” sebagai berikut.

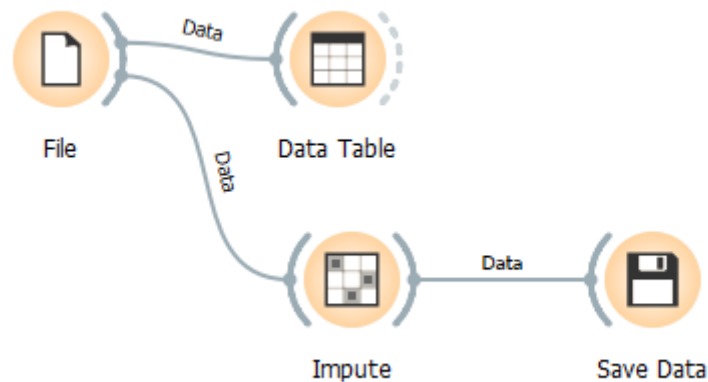


Gambar 2. 14 Menghilangkan data tak bernilai

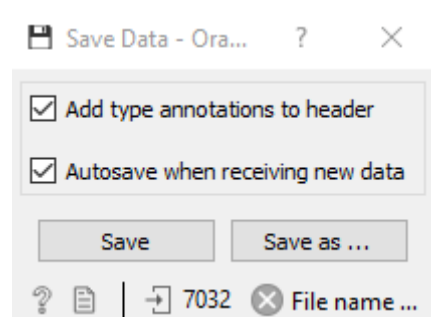
Setelah data yang tidak memiliki nilai dihilangkan, *dataset* yang baru harus disimpan dengan memilih fitur *Save Data* dan dihubungkan ke fitur *Impute* sebagai berikut.



Gambar 2. 15 *Save data*

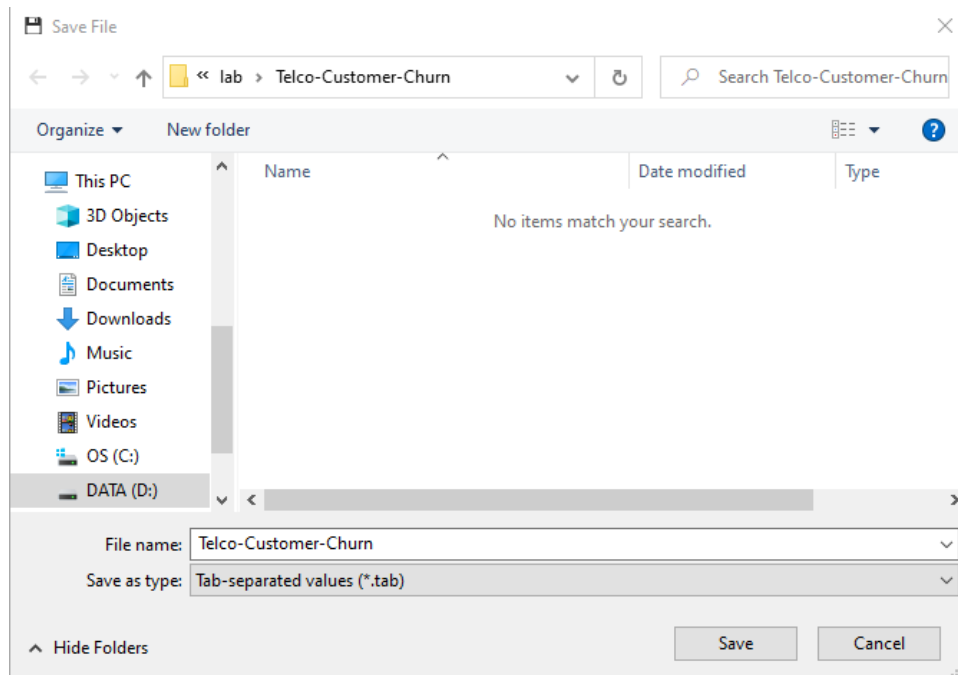


Gambar 2. 16 Hubungkan fitur *save data*

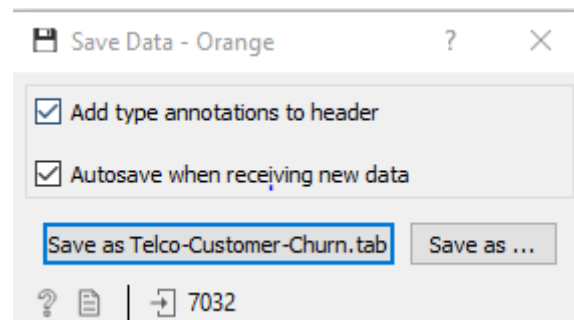


Gambar 2. 17 Pengaturan *save data*

Fitur *Save Data* dapat diklik ganda dan ditampilkan halaman untuk menyimpan *dataset* yang baru dengan nama *file* Telco-Customer-Churn.tab sebagai berikut.

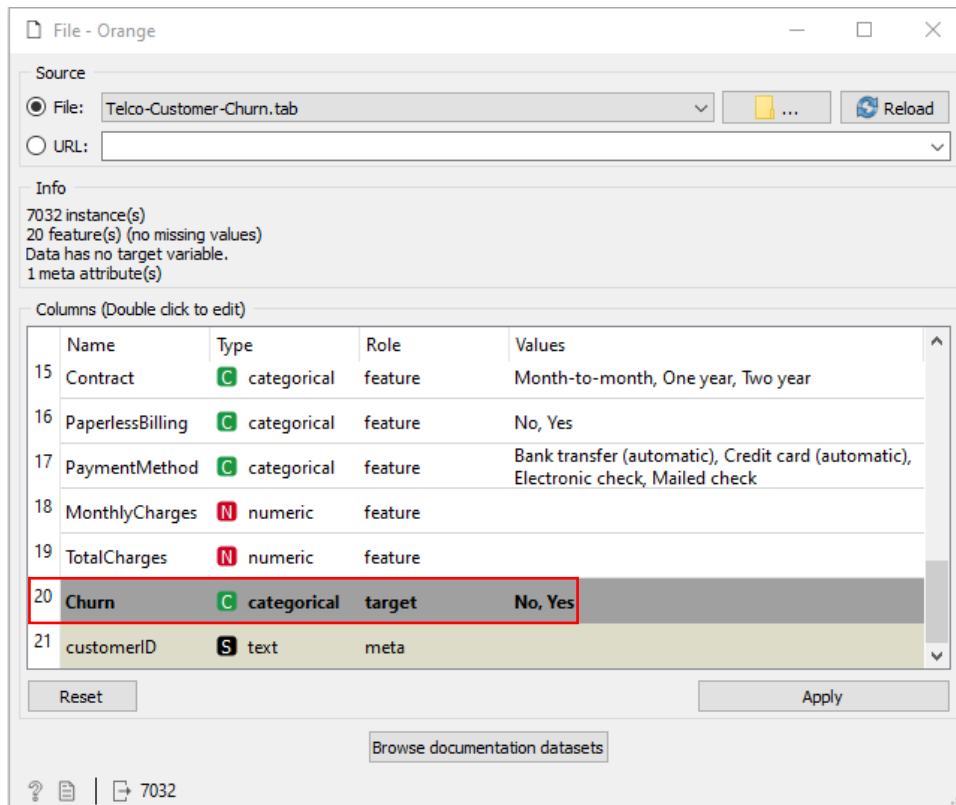


Gambar 2. 18 menyimpan *dataset* baru



Gambar 2. 19 menyimpan *dataset* baru

Setelah dataset tersimpan, klik ganda ikon File dan pilih dataset baru dengan nama *file* Telco-Customer-Churn.tab, sehingga dataset yang baru pada gambar berikut berhasil digunakan dengan data TotalCharges yang sudah bersih. *Role* pada atribut Churn dijadikan sebagai *target*.



Gambar 2. 20 Informasi *file data*

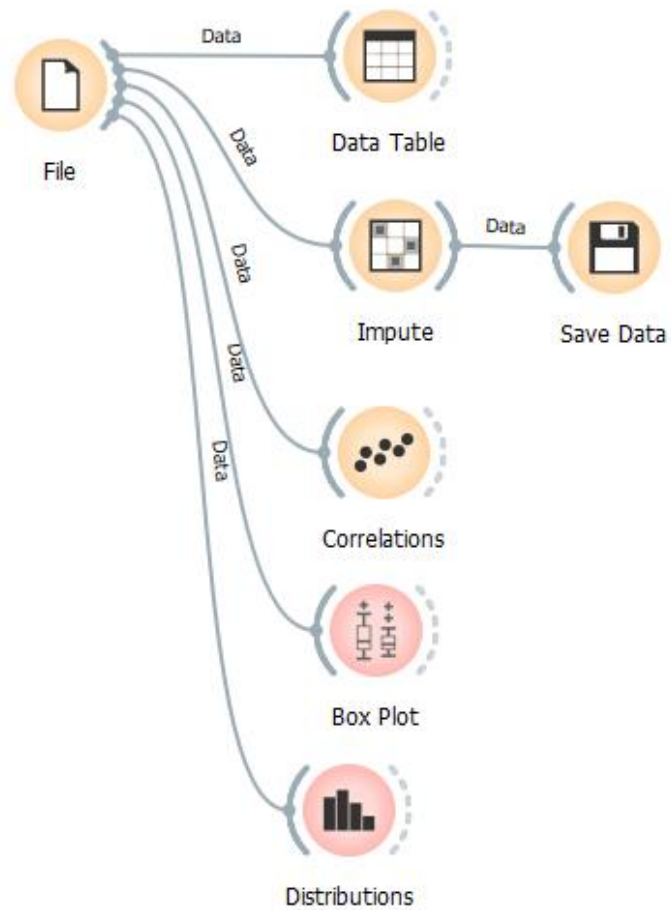
Dataset yang baru sudah bersih dan dapat digunakan untuk analisis. Dari *Data Table* pada di atas dapat diketahui bahwa sudah tidak terdapat *missing data* dan *role* target juga sudah ditentukan. Pada Gambar 2. 10 diketahui bahwa terdapat 7.043 *instances* yang kemudian setelah dilakukan *pre-processing*, dataset yang baru menunjukkan pada gambar berikut memiliki 7.032 *instances*.

	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges
1630	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	One year	No	Mailed check	18.80	18.80
6461	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	Yes	Mailed check	18.85	18.85
1148	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Electronic check	18.85	18.85
4932	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Mailed check	18.90	18.90
583	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	Yes	Mailed check	19.00	19.00
1729	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Mailed check	19.05	19.05
5328	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Mailed check	19.10	19.10
4342	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	Yes	Mailed check	19.10	19.10
3106	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Mailed check	19.10	19.10
368	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Mailed check	19.15	19.15
4385	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Mailed check	19.20	19.20
3309	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Electronic check	19.20	19.20
2985	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	Yes	Mailed check	19.20	19.20
1812	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Mailed check	19.20	19.20
6243	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Electronic check	19.25	19.25
5638	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	No	Mailed check	19.25	19.25
311	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	Yes	Bank transfer (a...	19.25	19.25
6830	o internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	No internet ser...	Month-to-month	Yes	Mailed check	19.30	19.30

Gambar 2. 21 *Dataset* bersih

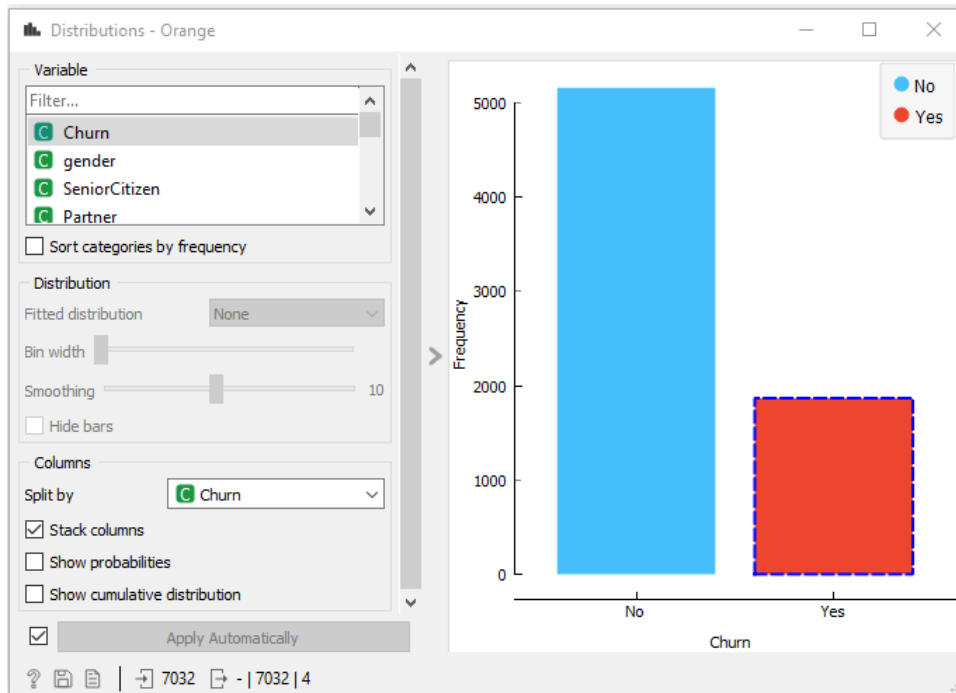
2.1.4 Data Exploratory & Distribution

Dataset yang sudah bersih setelah dilakukan *pre-processing* dapat digunakan untuk melakukan analisis. Untuk dapat melakukan analisis, dilakukan eksplorasi data dengan data *categorical* dan *numerical*. Untuk data *categorical*, dapat menggunakan fitur *Distribution* yang dihubungkan ke ikon File untuk melihat persebaran data sebagai berikut.



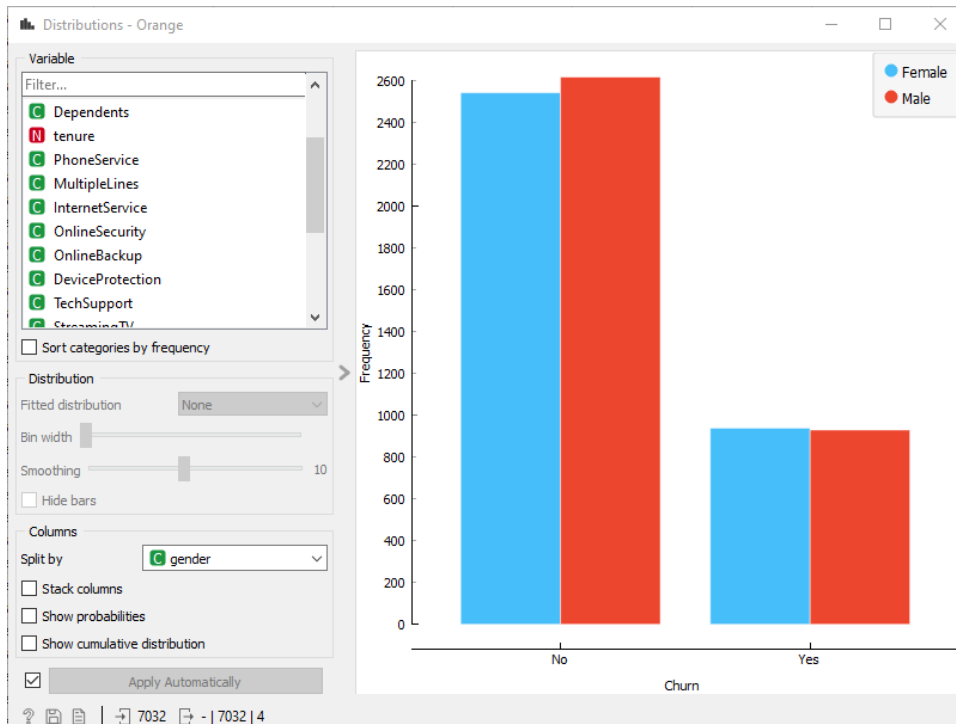
Gambar 2. 22 menghubungkan fitur *distribution*

Dengan fitur *Distribution*, dapat dilihat persebaran data untuk masing-masing nilai pada setiap atribut. Pada gambar berikut menunjukkan persebaran data *Churn*.



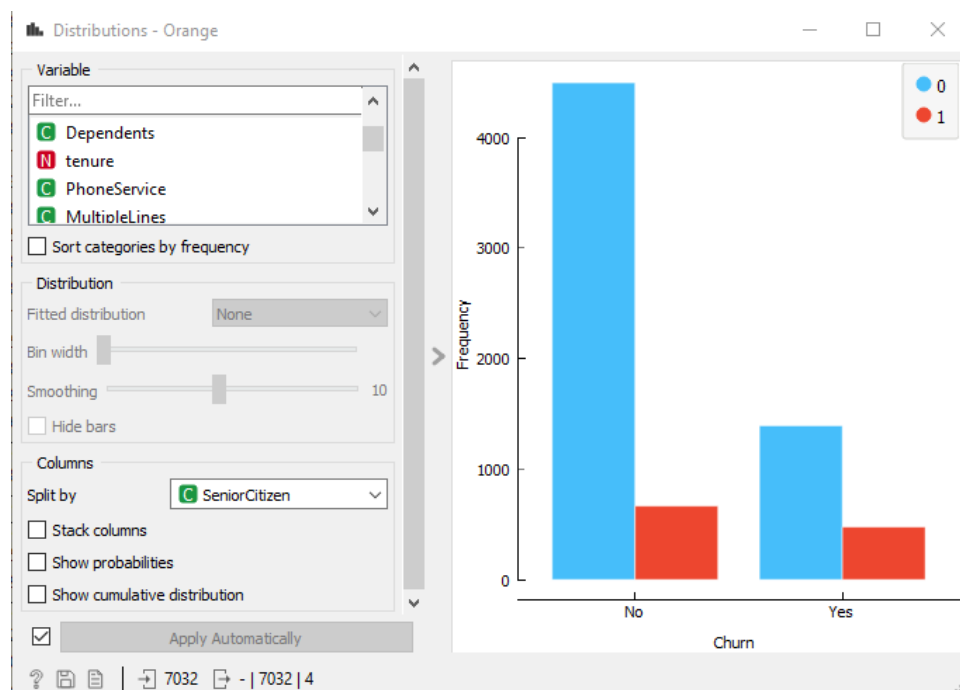
Gambar 2. 23 Data Churn

Dari keseluruhan pelanggan, ditunjukkan bahwa sebagian besar pelanggan tetap berlangganan dan tidak meninggalkan perusahaan, yaitu sebesar 73%. Sementara itu, 26% dari keseluruhan pelanggan berhenti berlangganan atau terindikasi *churn*. Setelah diketahui hal ini, dilakukan analisis lebih lanjut terhadap metrik data pelanggan lainnya sehingga didapatkan langkah yang tepat untuk perusahaan mempertahankan pelanggan. Pada Gambar berikut menunjukkan persebaran data *gender*.



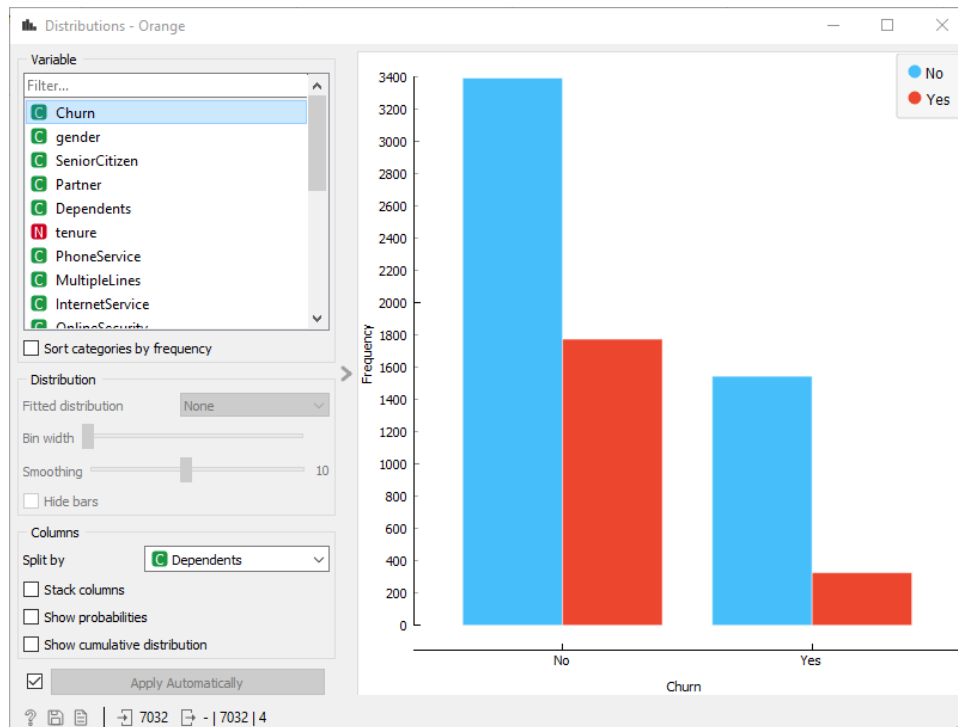
Gambar 2. 24 Data Gender

Dari data *gender*, diketahui bahwa pria dan wanita memiliki peluang yang sama untuk berpindah dari perusahaan, yaitu sebesar 50% untuk masing-masing jenis kelamin. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *SeniorCitizen*.



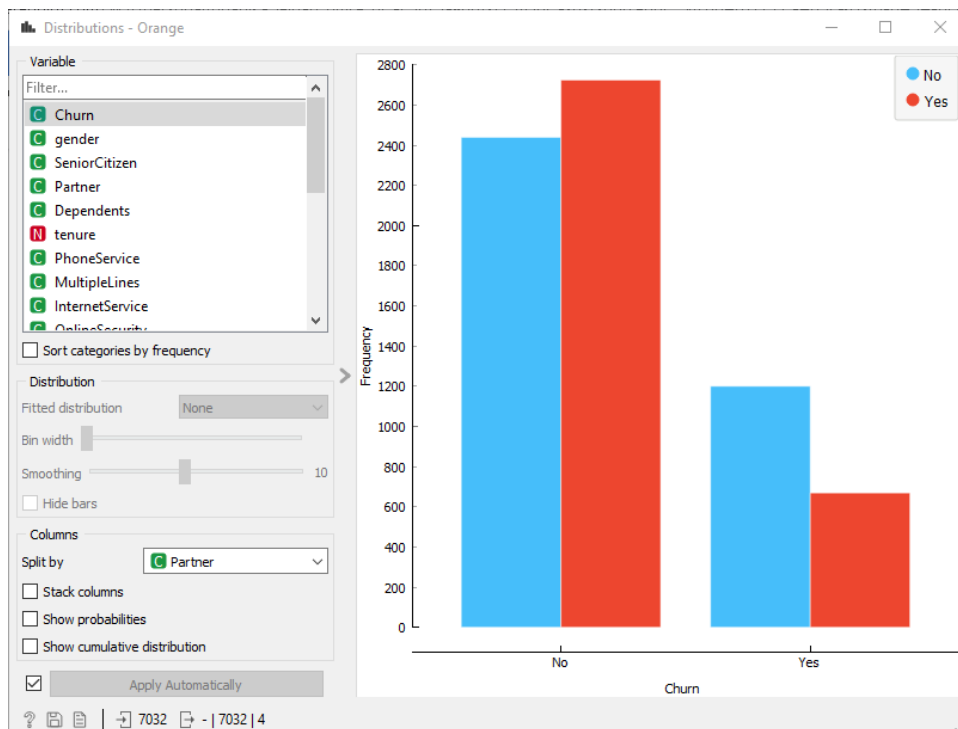
Gambar 2. 25 Data *Senior Citizen*

Dari data *SeniorCitizen*, diketahui bahwa pelanggan yang berusia 65 tahun ke bawah, yaitu dikategorikan sebagai non-*SeniorCitizen*, memiliki kecenderungan yang lebih tinggi untuk berpindah dari perusahaan (sebesar 74%) dibandingkan dengan pelanggan *Senior*. Salah satu faktor didapatkannya hasil ini yaitu bahwa pelanggan non-*Senior* memiliki mobilitas yang lebih tinggi dibandingkan dengan pelanggan *Senior* dalam penggunaan layanan telekomunikasi, sehingga pelanggan dalam kategori non-*Senior* cenderung lebih eksploratif dalam penggunaan layanan ini. Untuk mengatasi hal ini, pendekatan atau promosi khusus untuk kategori pelanggan non-*Senior* dapat diberikan agar mengurangi peluang pelanggan berpindah dari perusahaan. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *Dependents*.



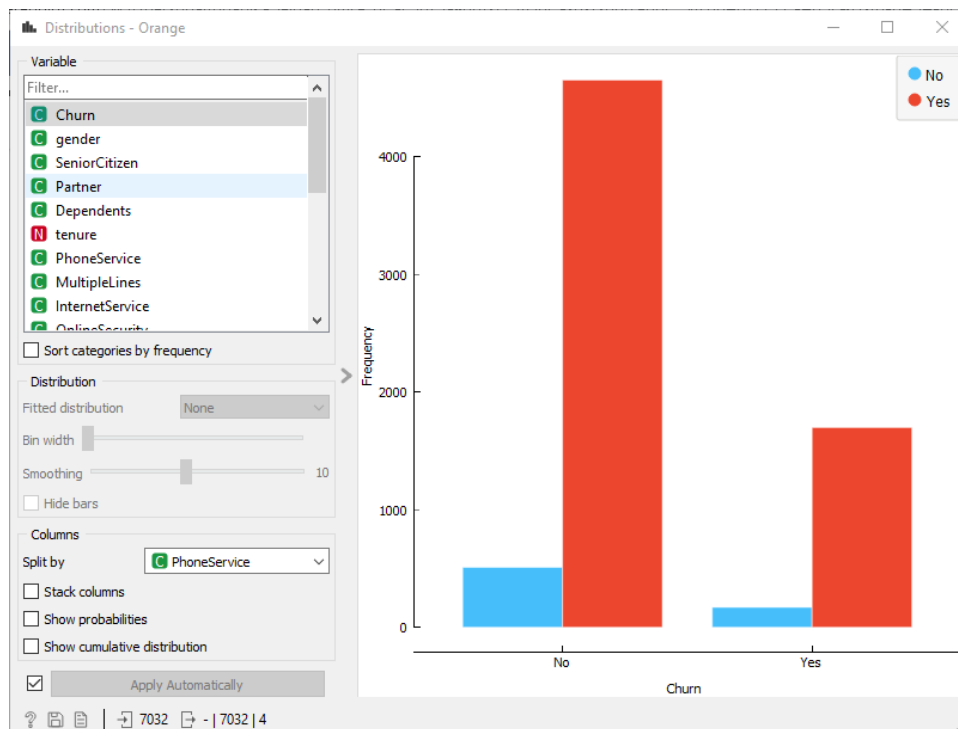
Gambar 2. 26 Data *Dependents*

Dari data *Dependents*, diketahui bahwa pelanggan yang tidak tinggal dengan tanggungan memiliki kecenderungan yang lebih tinggi untuk berpindah dari perusahaan, yaitu sebesar 82%. Dari hasil tersebut, dapat diindikasikan bahwa pelanggan yang tidak memiliki tanggungan merupakan pelanggan yang hidup sendiri di tempat tinggalnya, sehingga untuk mengatasi hal ini, dapat dibuatkan program atau paket khusus untuk pelanggan yang hanya menggunakan layanan telekomunikasi perseorangan agar pelanggan dalam kategori tersebut bisa menjadi lebih hemat dan tidak berpindah dari perusahaan. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *Partner*.



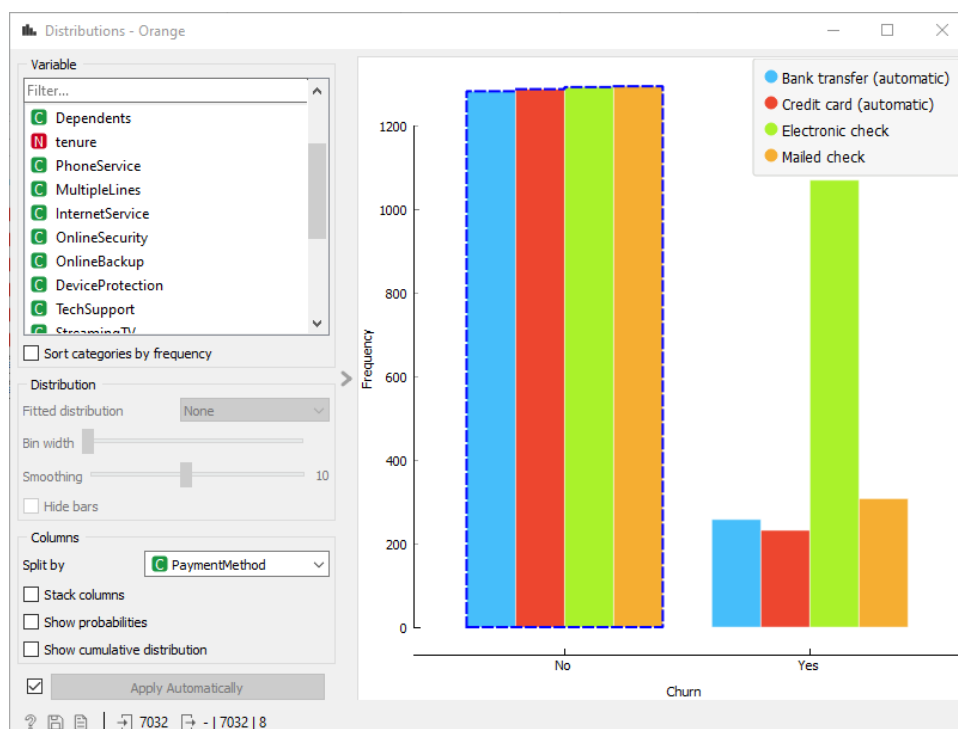
Gambar 2. 27 Data *partner*

Dari data *Partner*, diketahui bahwa pelanggan yang tidak memiliki pasangan cenderung untuk berpindah dari perusahaan, yaitu sebesar 64%. Diketahui bahwa hasil tersebut memiliki kemiripan dengan data *Dependents*, sehingga untuk mengatasi hal ini, dapat dibuatkan program atau paket khusus untuk pelanggan yang hanya menggunakan layanan telekomunikasi perseorangan agar pelanggan dalam kategori tersebut bisa menjadi lebih hemat dan tetap setia dengan perusahaan. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *PhoneService*.



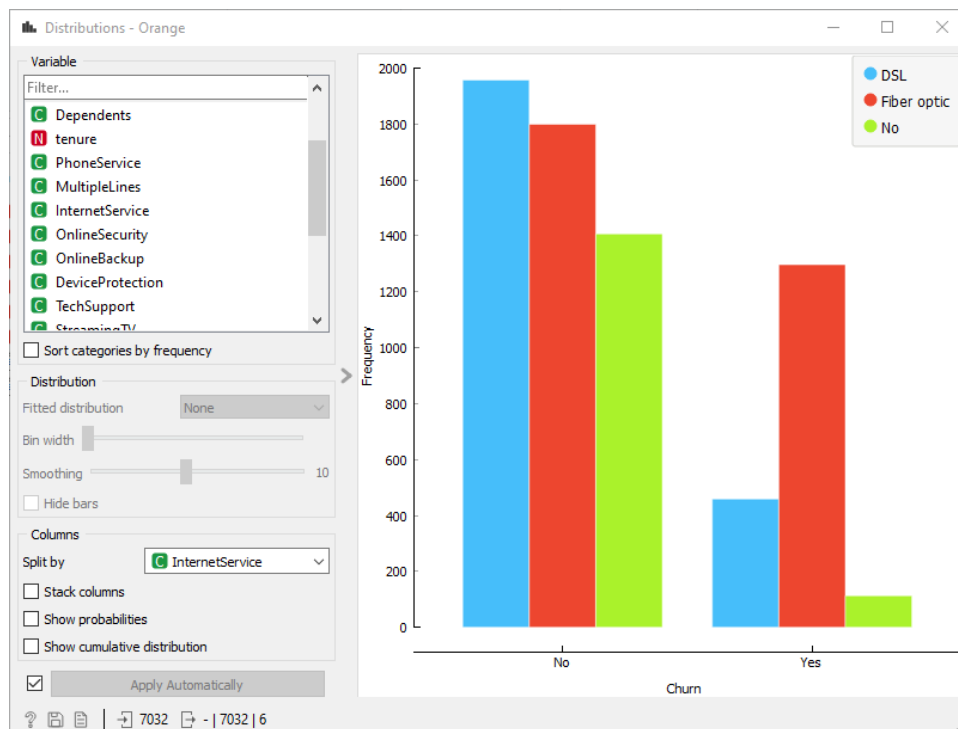
Gambar 2. 28 Data *Phone Service*

Dari data *PhoneService*, diketahui bahwa pelanggan yang berlangganan layanan telepon mendominasi keseluruhan pelanggan dan juga diketahui bahwa memiliki peluang yang tinggi untuk berpindah dari perusahaan, yaitu sebesar 90%. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *PaymentMethod*.



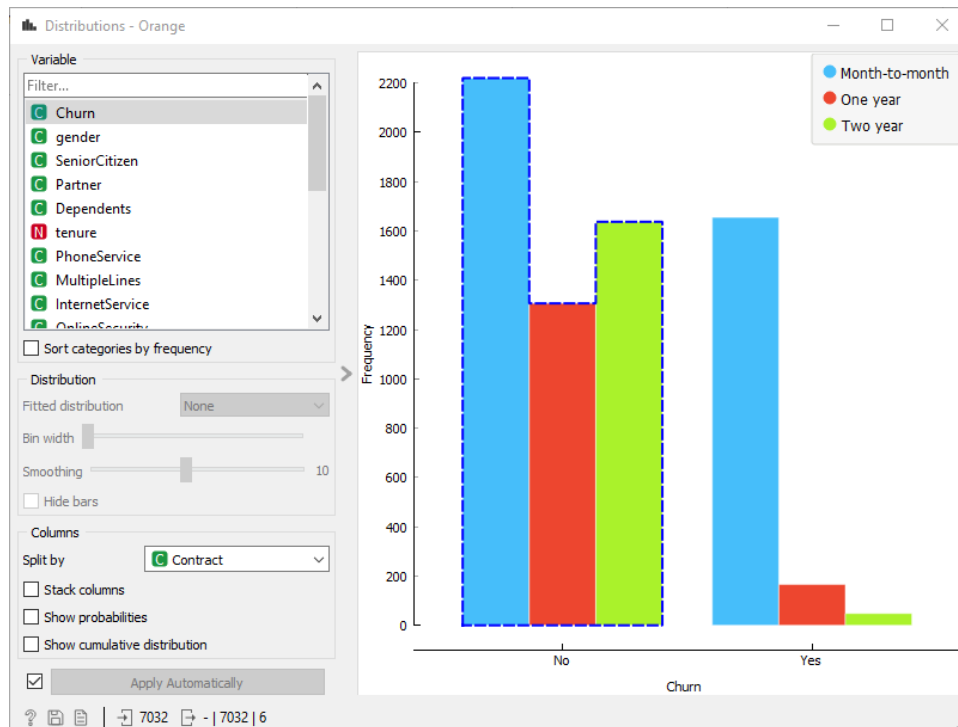
Gambar 2. 29 Data *Payment Method*

Dari data *PaymentMethod*, diketahui bahwa pelanggan yang menggunakan metode pembayaran cek elektronik lebih berpeluang untuk berpindah dari perusahaan, yaitu sebesar 57%. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *InternetService*.



Gambar 2. 30 Data *Internet Service*

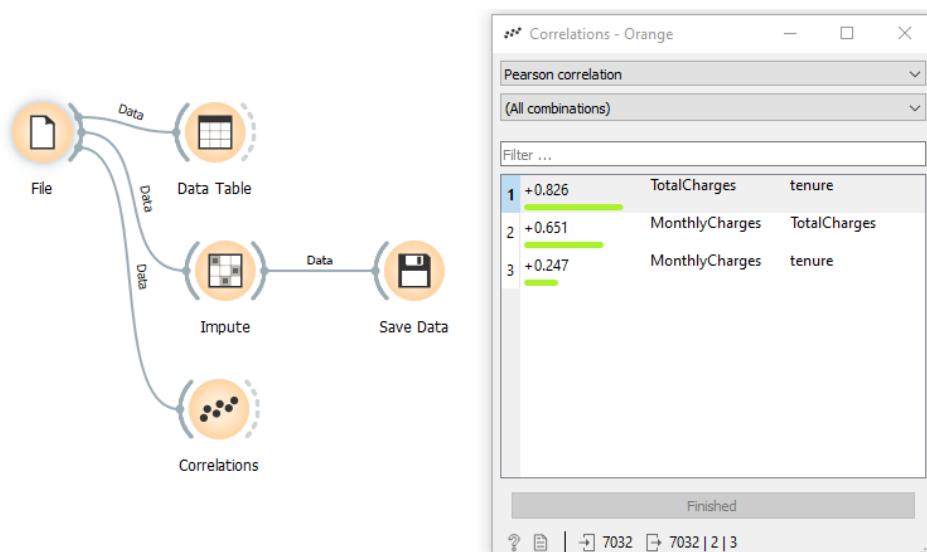
Dari data *InternetService*, diketahui bahwa pelanggan yang menggunakan Fiber Optic memiliki jumlah yang lebih tinggi untuk berpindah dari perusahaan, yaitu sebesar 69%. Dengan diketahuinya hal tersebut, layanan Fiber Optic ini dapat diperhatikan lebih lanjut dan dievaluasi oleh perusahaan dari segala sisi, melihat bahwa persaingan penggunaan Fiber Optic juga semakin tinggi. Selanjutnya, pada Gambar berikut menunjukkan persebaran data *Contract*.



Gambar 2. 31 Data *Contract*

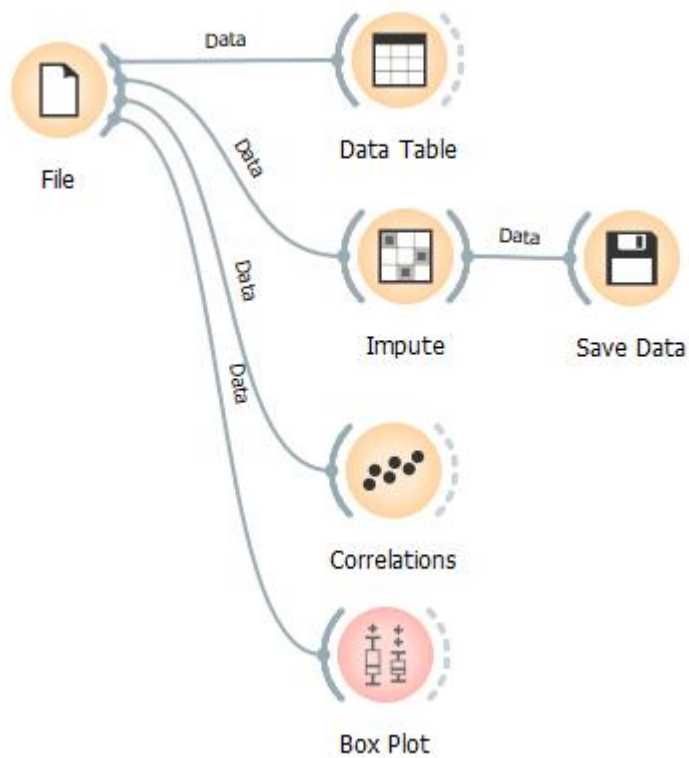
Dari data *Contract*, diketahui bahwa pelanggan yang memilih periode langganan perbulan memiliki kecenderungan lebih tinggi untuk berpindah dari perusahaan, yaitu sebesar 88%. Hasil ini dapat dipahami bahwa pelanggan yang berlangganan dalam hitungan bulan dapat berpindah kapan saja. Untuk mengatasi hal ini, promosi untuk program atau paket dalam periode waktu tertentu dapat diberikan agar pelanggan dapat lebih hemat dengan manfaat yang lebih menguntungkan.

Setelah dilihat persebaran data, kemudian dilihat juga hubungan antar data dengan menggunakan fitur *Correlation* untuk data *numerical* seperti pada Gambar di atas. Karena data yang diproses ini adalah data rasio, maka *Pearson correlation* digunakan dalam fitur ini. Pada Gambar berikut ditampilkan hubungan antar data dengan fitur *Correlations*.



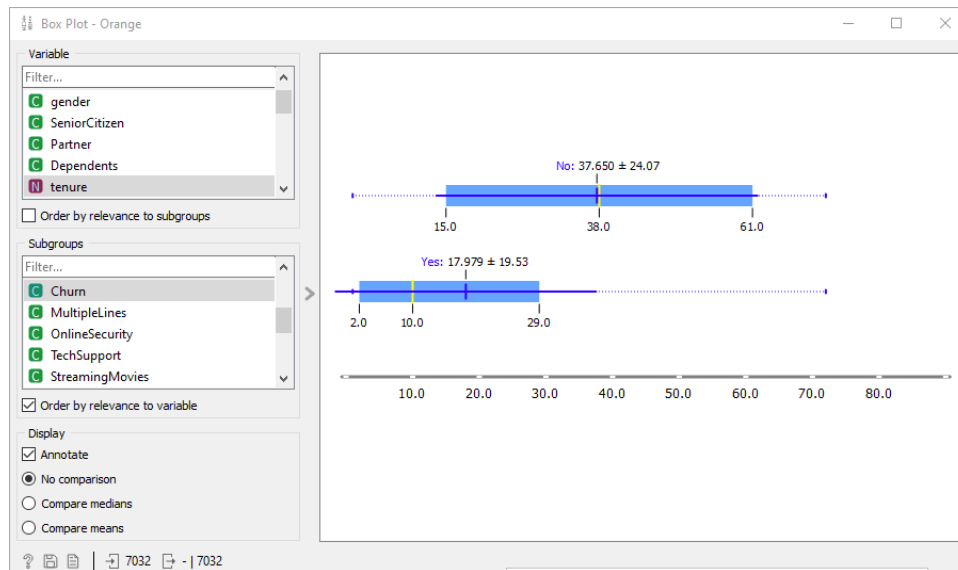
Gambar 2. 32 *Correlations*

Hasil dari fitur *Correlations* menunjukkan bahwa atribut *TotalCharges* memiliki hubungan yang kuat dengan atribut *tenure* (masa berlangganan pelanggan) dan juga atribut *MonthlyCharges*. Hal ini diketahui karena *TotalCharges* adalah total akumulatif dari *MonthlyCharges* dan *tenure*. Selanjutnya, hubungan data juga ditampilkan dengan menggunakan fitur *Box Plot* sebagai berikut.



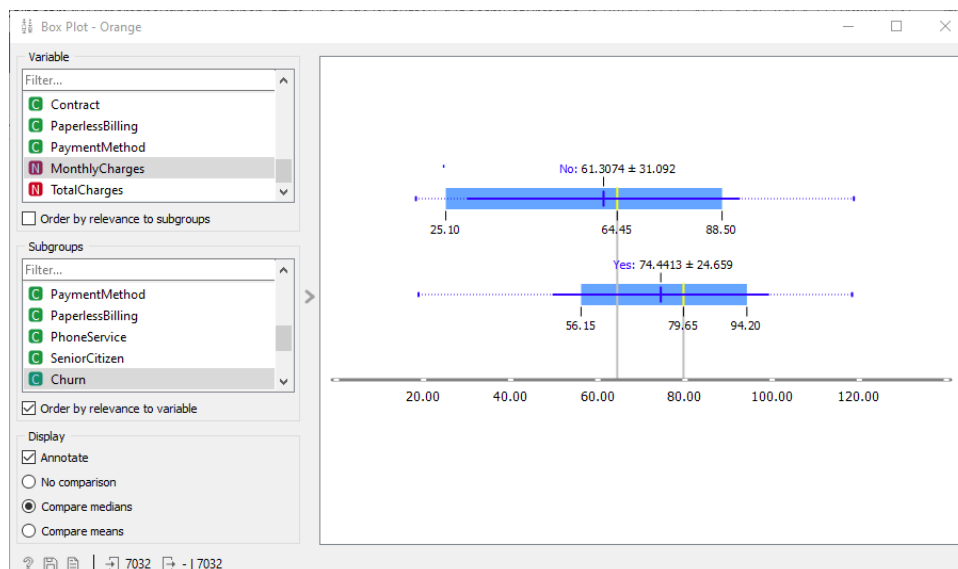
Gambar 2. 33 Hubungkan fitur *Box Plot*

Dengan melakukan klik ganda pada ikon *Box Plot*, kemudian ditampilkan halaman *Box Plot* sebagai berikut.



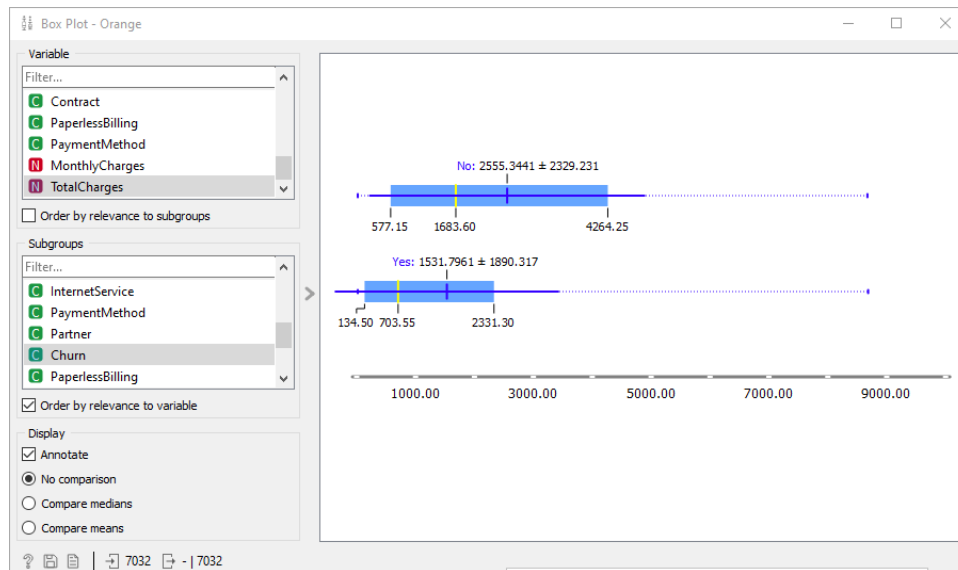
Gambar 2. 34 Data *tenure*

Dari data *tenure*, diketahui bahwa pelanggan dengan masa periode berlangganan yang rendah cenderung untuk berpindah dari perusahaan, sedangkan pelanggan dengan masa periode berlangganan yang lama cenderung untuk tetap berlangganan karena sudah menjadi pelanggan setia. Hasil yang didapatkan ini juga berkaitan dengan hasil pada data *Contract* yang menunjukkan bahwa pelanggan dengan periode langganan perbulan memiliki kecenderungan untuk berpindah dari perusahaan. Selanjutnya, pada Gambar berikut menunjukkan data *MonthlyCharges*.



Gambar 2. 35 Data *Monthly Charges*

Dari data *MonthlyCharge*, diketahui bahwa pelanggan dengan tagihan perbulan yang tinggi cenderung untuk berpindah dari perusahaan, sedangkan pelanggan dengan tagihan perbulan yang lebih rendah tetap bertahan dan tidak terindikasi *churn*. Untuk mengatasi hal ini, pemberian promosi khusus atau paket yang lebih hemat dapat diberikan untuk pelanggan agar tetap berlangganan. Selanjutnya, pada Gambar berikut menunjukkan data *TotalCharges*.

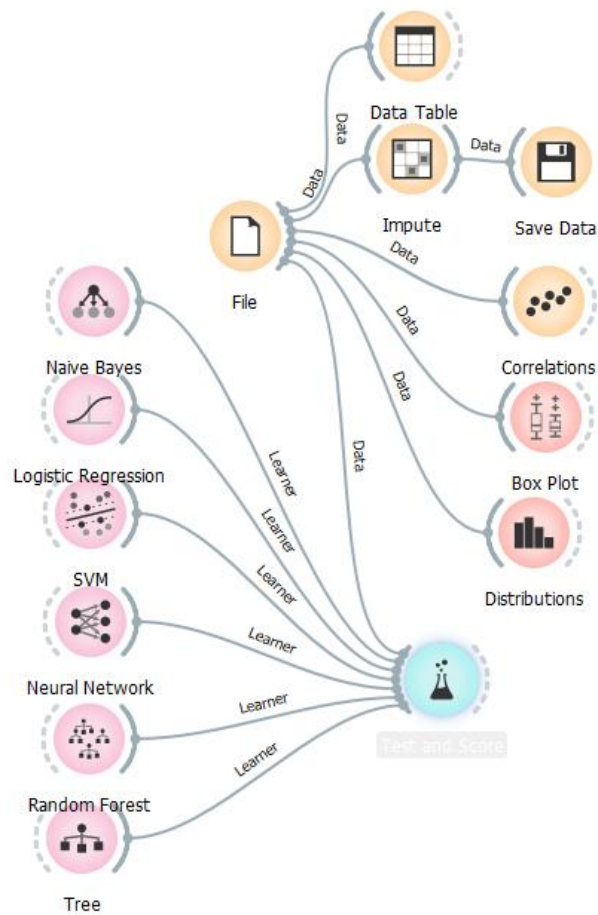


Gambar 2. 36 Data *Total Charges*

Dari data *TotalCharges*, diketahui bahwa pelanggan dengan total tagihan yang rendah cenderung untuk berpindah dari perusahaan, sedangkan pelanggan dengan total tagihan yang lebih tinggi tetap bertahan dan tidak terindikasi *churn*. Hal ini dapat dipahami karena berdasarkan data dan hasil temuan sebelumnya, diketahui bahwa sebagian besar pelanggan yang melakukan *churning* adalah pelanggan dengan masa periode berlangganan yang rendah, sehingga didapatkan data pelanggan dengan total tagihan yang rendah adalah yang terindikasi *churn*. Sementara itu, pelanggan dengan masa periode langganan yang tinggi dapat termasuk ke dalam kategori pelanggan dengan total tagihan yang tinggi juga.

2.1.5 Data Modelling

Dataset Telco-Customer-Churn ini adalah data *supervised* dengan atribut *Churn* sebagai target yang memiliki nilai *yes* dan *no*. Pemodelan untuk data *supervised* ini dapat menggunakan model klasifikasi. Untuk melakukan pemodelan data, fitur masing-masing model klasifikasi dapat dimasukkan ke dalam lembar kerja Orange sebagai berikut.



Gambar 2. 37 Pemodelan data

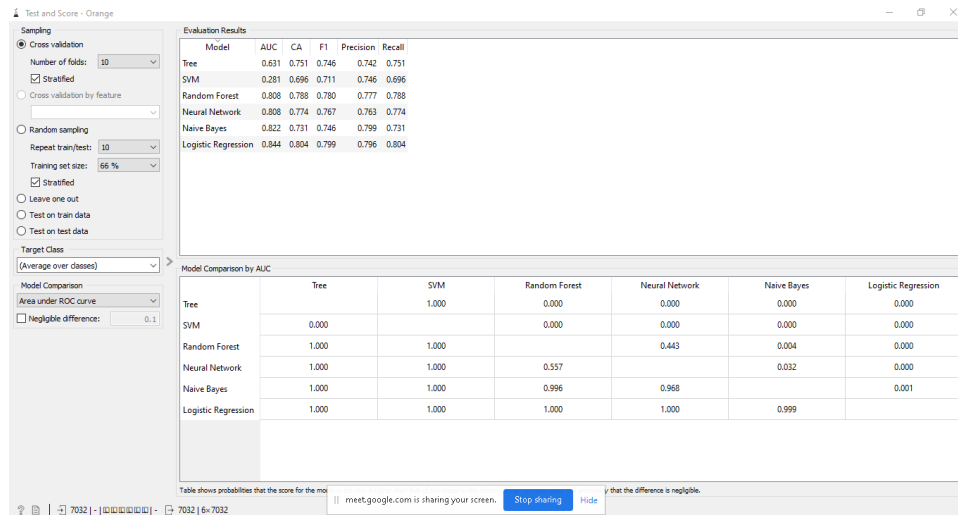
Pada Gambar di atas menunjukkan bahwa model klasifikasi yang digunakan untuk pemodelan data beserta alasan penggunaan metode ini adalah sebagai berikut:

- *Naive Bayes Classifier*, dipakai karena menjadi metode yang telah banyak digunakan oleh berbagai macam kasus klasifikasi dan hanya membutuhkan jumlah training data yang kecil.
- *Logistic Regression*, digunakan dengan alasan hanya membutuhkan waktu yang sedikit dalam prediksi modelnya dan metode ini tidak memiliki asumsi normalitas dan heteroskedastisitas atas variabel bebas yang digunakan dalam model sehingga tidak diperlukan uji asumsi klasik walaupun variabel independen
- *Support Vector Machine (SVM)*, metode ini merupakan metode yang peka terhadap pola data yang tidak beraturan dan juga relatif cepat dalam melakukan proses eksekusi modelnya.
- *Neural Network*, merupakan metode yang memiliki keunggulan dalam hal menyesuaikan pola data yang diberikan.
- *Random Forest*, memiliki tingkat akurasi yang cukup baik tetapi mengkonsumsi banyak waktu dalam eksekusi model.
- *Decision Tree*, menggunakan algoritma yang lebih simple dan juga waktu eksekusi model yang lebih singkat dibanding dengan metode lain.

Setelah itu, dilakukan evaluasi dengan menggunakan fitur *Test and Score* pada Orange.

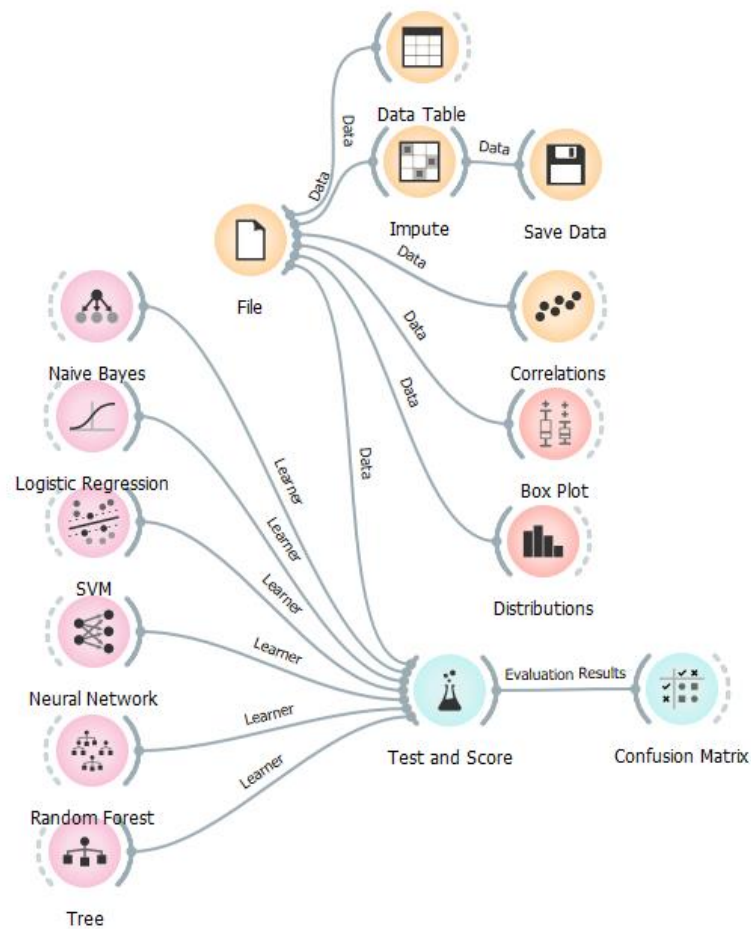
2.1.6 Results dan Analisis

Dari pemodelan data yang sudah dilakukan, kemudian dilakukan evaluasi dengan fitur *Test and Score* dan didapatkan hasil sebagai berikut.



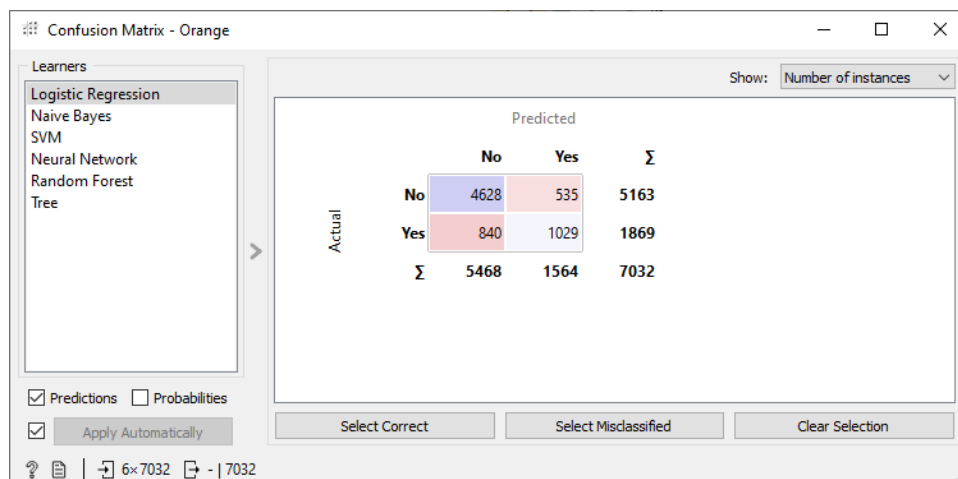
Gambar 2. 38 Hasil evaluasi

Pada Gambar di atas dapat dilihat pengaturan fitur *Test and Score* dengan menggunakan teknik *sampling k-fold cross validation* dengan nilai k sebesar 10. Hasil yang didapatkan menunjukkan bahwa *Logistic Regression* adalah algoritme yang mendapatkan hasil kinerja tertinggi dibandingkan dengan algoritme lainnya, yakni dengan nilai *precision* sebesar 0.79, *recall* sebesar 0.80, dan *F1 score* sebesar 0.79. Dari hasil perhitungan kinerja model tersebut, hasil pemodelan yang mendetail dapat dilihat dengan menggunakan fitur *Confusion Matrix* yang dihubungkan dengan ikon *Test and Score* seperti pada Gambar berikut.



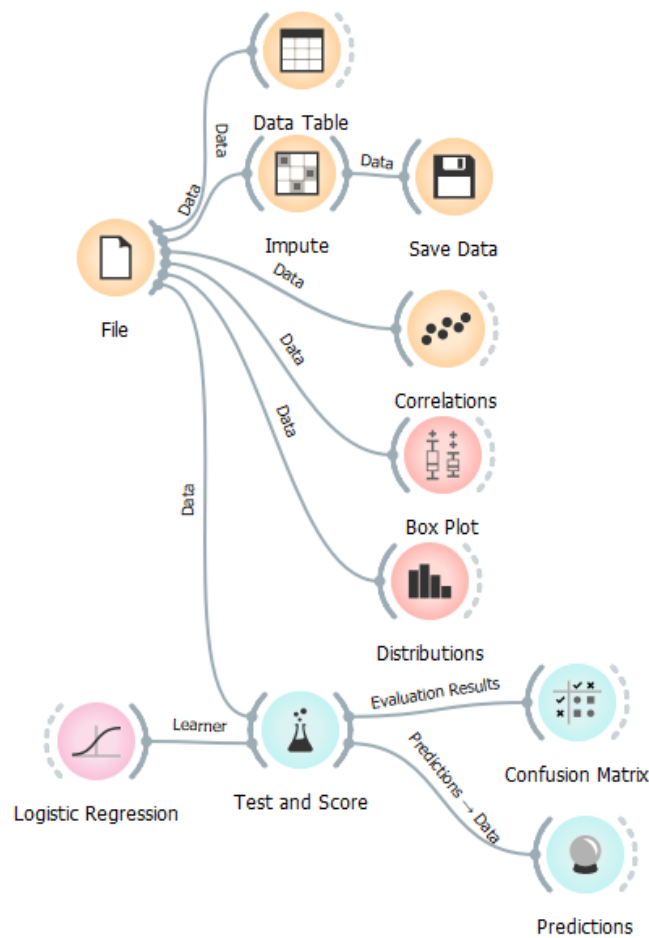
Gambar 2. 39 Hubungkan fitur *Confusion Matrix*

Setelah fitur *Confusion Matrix* dihubungkan dengan *Test and Score*, maka klik ganda pada ikon *Confusion Matrix* dan didapatkan sebagai berikut.



Gambar 2. 40 Hasil *Confusion Matrix*

Kemudian dilakukan prediksi dengan fitur predictions yang dihubungkan dengan fitur *Test and Score* dan mendapatkan hasil sebagai berikut.



Gambar 2. 41 Hubungkan fitur *Predictions*

Predictions - Orange

Show probabilities for

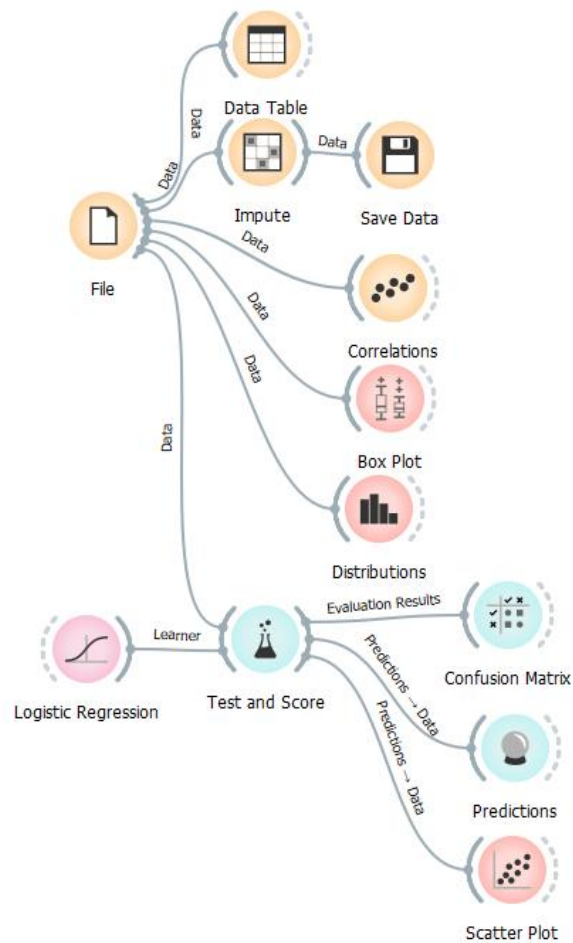
Churn	customerID	Logistic Regression	Logistic Regression (No)	istic Regression (Fold	gender	Senior/Citizen	Partner	Dependents	tenure
No	9763-GRSKD	No	0.803276	0.196724	1	Male	0	Yes	Yes	13.0
Yes	4190-MFLUW	No	0.837926	0.162074	1	Female	0	Yes	Yes	10.0
Yes	5407-CHPZW	Yes	0.496912	0.503088	1	Male	0	Yes	Yes	47.0
Yes	5380-WIKOV	Yes	0.398244	0.601756	1	Male	0	No	No	34.0
No	5948-UIZLF	No	0.669159	0.330841	1	Male	0	No	No	2.0
No	2954-PIBKO	No	0.993779	0.00622111	1	Female	0	Yes	Yes	69.0
No	4667-QONEA	No	0.982732	0.0172678	1	Female	1	Yes	Yes	60.0
No	3957-SQXML	No	0.963332	0.036667	1	Female	0	Yes	Yes	34.0
No	8028-PAVHQ	No	0.995307	0.0046928	1	Male	0	Yes	Yes	62.0
No	6416-PAWRK	No	0.523206	0.4767938	1	Female	0	No	No	46.0
No	2424-WHHPF	Yes	0.42811	0.57189	1	Male	1	No	No	1.0
Yes	8627-XJWR	Yes	0.430869	0.569131	1	Female	0	No	No	12.0
No	5236-SKJGO	No	0.961623	0.038377	1	Female	0	Yes	Yes	64.0
No	3930-ZGWVE	No	0.791546	0.208454	1	Male	0	No	No	1.0
No	2876-GZYVC	Yes	0.265918	0.734082	1	Female	0	No	No	13.0
No	3774-VBNXY	No	0.979966	0.02003373	1	Female	0	Yes	Yes	64.0
No	3780-VVMFA	No	0.558262	0.441738	1	Female	0	Yes	Yes	8.0
Yes	0887-MKZYI	No	0.576636	0.423364	1	Female	0	Yes	No	20.0
No	5997-OPVFA	No	0.979959	0.020041	1	Male	0	Yes	No	72.0
No	1139-WFSGR	No	0.678506	0.321494	1	Female	0	Yes	Yes	16.0
Yes	1918-ZBFOJ	Yes	0.465551	0.534449	1	Female	0	No	No	13.0
No	9142-KZXOP	No	0.840153	0.159847	1	Male	0	No	No	44.0
No	1347-KTTTA	No	0.90437	0.0956302	1	Male	0	Yes	No	64.0
No	2504-DSHHH	Yes	0.346429	0.653571	1	Male	1	Yes	No	23.0
No	0699-NDKJM	No	0.880084	0.119916	1	Female	0	Yes	No	57.0
No	5961-VUSRV	No	0.914391	0.085609	1	Female	0	No	No	12.0
Yes	2232-DMLXU	No	0.574124	0.425876	1	Female	0	Yes	No	1.0
No	0945-TSONK	No	0.961453	0.0385466	1	Female	0	Yes	Yes	47.0
No	1671-SEAPM	No	0.700407	0.299593	1	Female	0	No	No	11.0

Restore Original Order

7032 | 7032 |

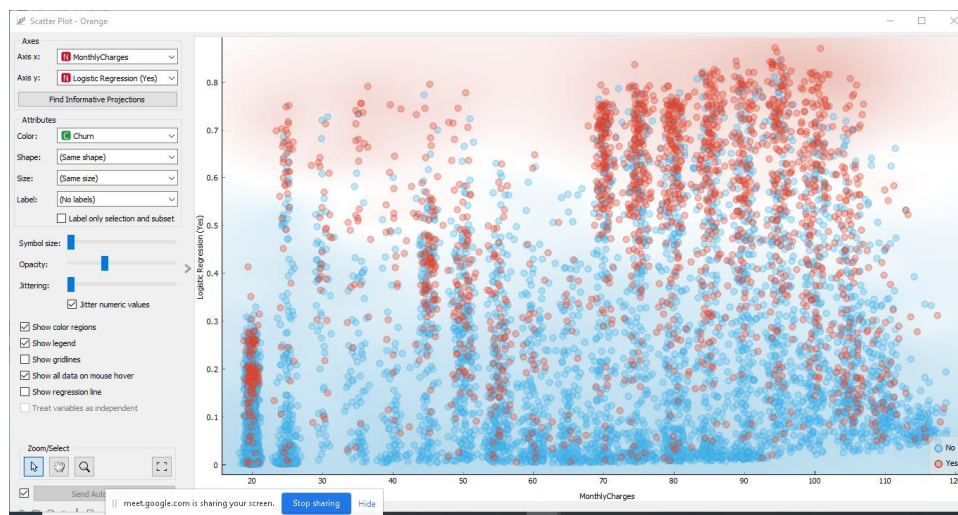
Gambar 2. 42 Hasil *Predictions*

Selanjutnya, untuk melihat pola dari hasil persebaran data dari pemodelan dapat menggunakan fitur *Scatter Plot* dan dihubungkan dengan fitur *Test and Score* sebagai berikut.



Gambar 2. 43 Hubungkan fitur *Scatter Plot*

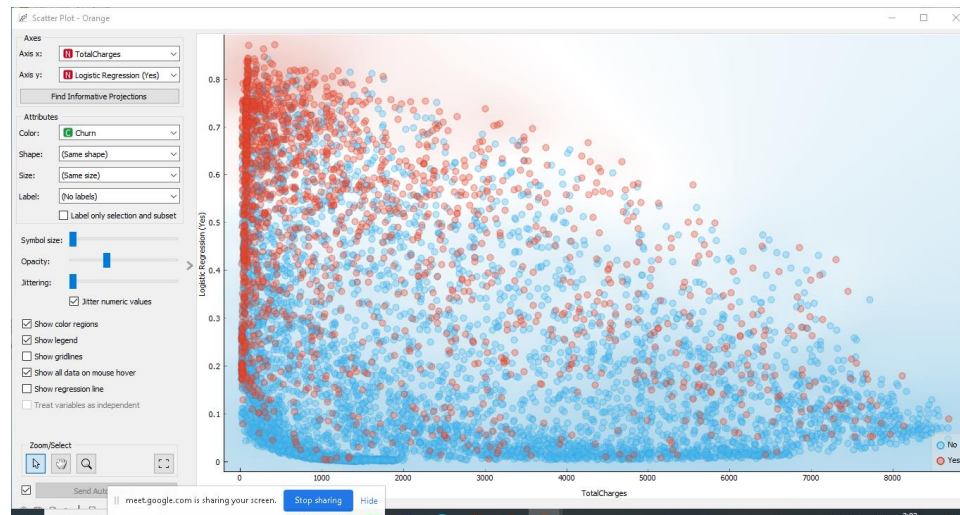
Setelah dihubungkan, klik ganda pada ikon *Scatter Plot* dan ditampilkan halaman seperti pada Gambar berikut.



Gambar 2. 44 Persebaran data *Monthly Charges*

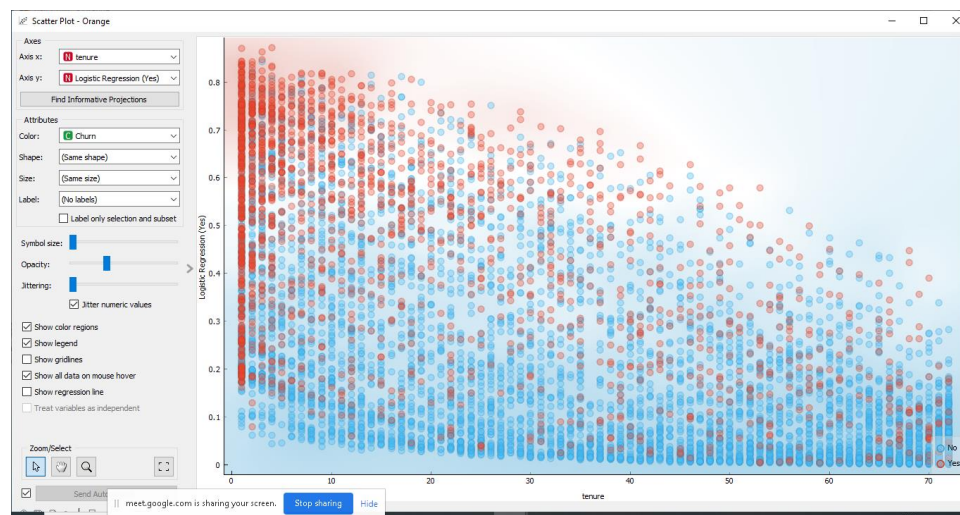
Gambar di atas menunjukkan data *MonthlyCharges* dan diketahui bahwa pelanggan dengan tagihan perbulan yang tinggi memiliki potensi yang tinggi untuk berpindah dari perusahaan,

yakni yang digambarkan dengan titik merah. Hal ini dapat diatasi dengan memberikan program atau paket berlangganan khusus yang disesuaikan dengan profil para pelanggan, sehingga titik merah dapat berkurang. Selanjutnya, persebaran data *TotalCharges* ditampilkan pada Gambar berikut.



Gambar 2. 45 Persebaran data *Total Charges*

Gambar di atas menunjukkan bahwa pelanggan dengan total tagihan yang rendah memiliki potensi yang tinggi untuk berpindah dari perusahaan. Hal ini berkaitan erat dengan *tenure* atau masa periode berlangganan pelanggan yang rendah juga, seperti pada Gambar berikut.



Gambar 2. 46 Persebaran data *tenure*

Gambar di atas menunjukkan bahwa pelanggan dengan total tagihan yang rendah memiliki potensi yang tinggi untuk berpindah dari perusahaan. Hal ini berkaitan dengan data total tagihan, dimana pelanggan dengan total tagihan yang rendah merupakan pelanggan dengan masa periode berlangganan yang tidak lama. Kedua data ini saling berkaitan dan menunjukkan bahwa pelanggan-pelanggan tersebut memiliki kecenderungan yang tinggi untuk berpindah dari perusahaan. Hal ini dapat diatasi dengan melakukan evaluasi dari segala sisi sesuai dengan temuan-temuan yang didapatkan sebelumnya.

