



Klasifikasi pada Orange


Dr.Indra, M.T.I

Outline

- Sumber dataset
- Praktikum Klasifikasi dengan Orange
(orange.biolab.si)

Sumber Data

data.jakarta.go.id




**JAKARTA
open data**
Berbagi Data untuk Transparansi

Temukan Data Seluruh Unit Kerja Pemerintah Provinsi DKI Jakarta


DATA | ORGANISASI | TOPIK | VISUALISASI | INFOGRAFIS | TENTANG

Temukan **4695** data di sini.


Data Tempat Ibadah 




PENDIDIKAN




KESIHATAN



KEUANGAN
DAERAH



KEPENDUDUKAN



PERHUBUNGAN

Sumber data

data.jakarta.go.id/dataset?q=Data+Tempat+Ibadah

Dinas Pariwisata Da... (9)

Biro Umum dan Admin... (8)

Dinas Kesehatan (7)

Biro Pendidikan dan... (7)

Kota Administrasi J... (6)

Dinas Komunikasi, I... (6)

Show More Organisasi

Topik

Pariwisata & kebuda... (9)

Kesehatan (6)

Penanggulangan Bencana
(5)

Data Penyebaran Tempat Ibadah Di Kota Administrasi Jakarta Timur

Dataset ini berisi mengenai Data Penyebaran Tempat Ibadah Di Kota Administrasi Jakarta Timur yang meliputi masjid, musholla, gereja, vihara dan pura. Penjelasan mengenai...

CSV

Data Penyebaran Tempat Ibadah Di Wilayah Kota Administrasi Jakarta Barat

data ini merupakan kumpulan data penyebaran tempat ibadah di Wilayah Kota Administrasi Jakarta Barat. Data ini berisikan variabel : kecamatan kelurahan...

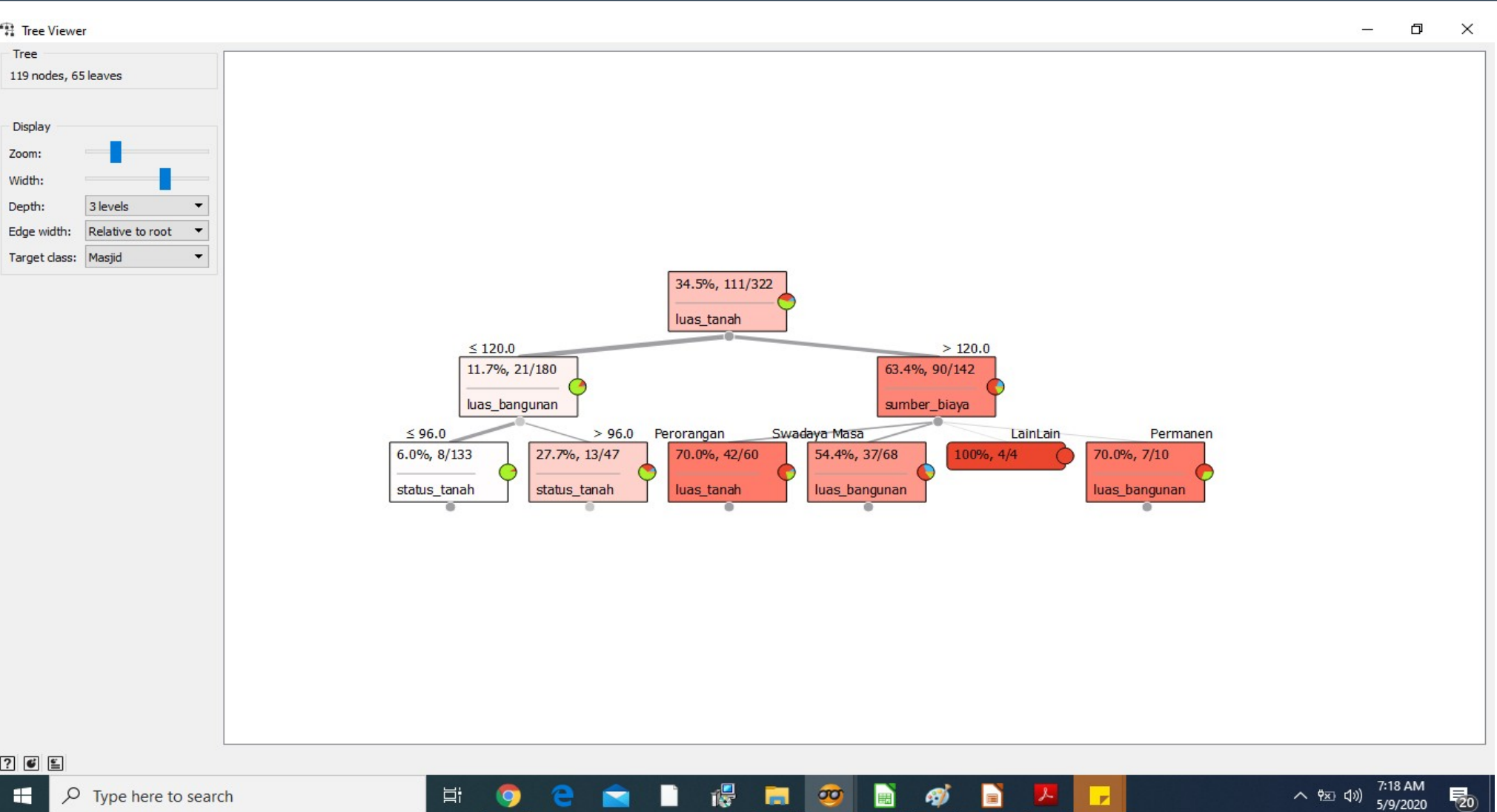
CSV

Data Penyebaran Tempat Ibadah Di Kota Administrasi Jakarta Utara

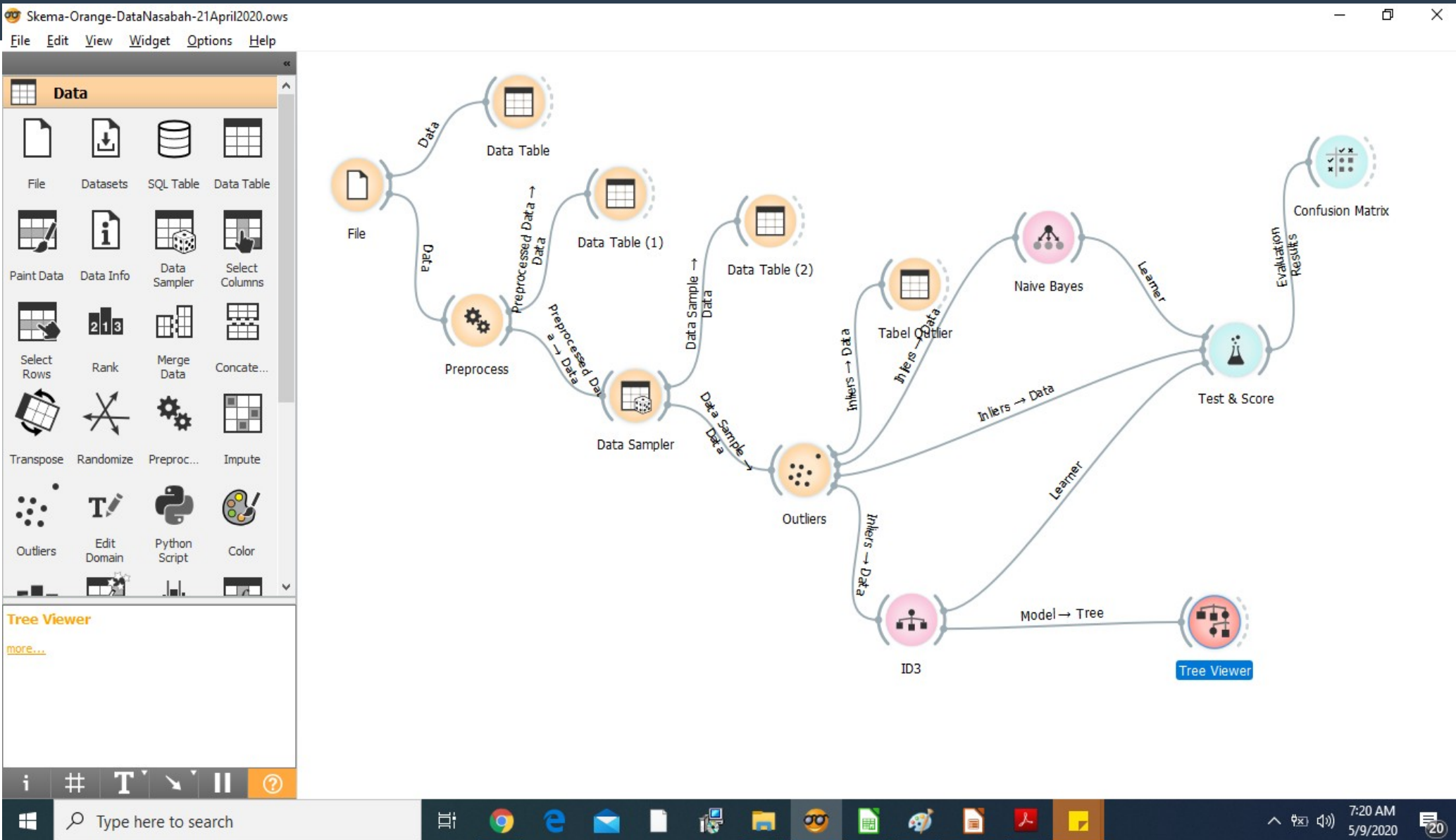
Dataset ini berisi mengenai Data Penyebaran Tempat Ibadah Di Kota Administrasi Jakarta Utara yang meliputi masjid, musholla, gereja, vihara dan pura. Penjelasan mengenai...

CSV

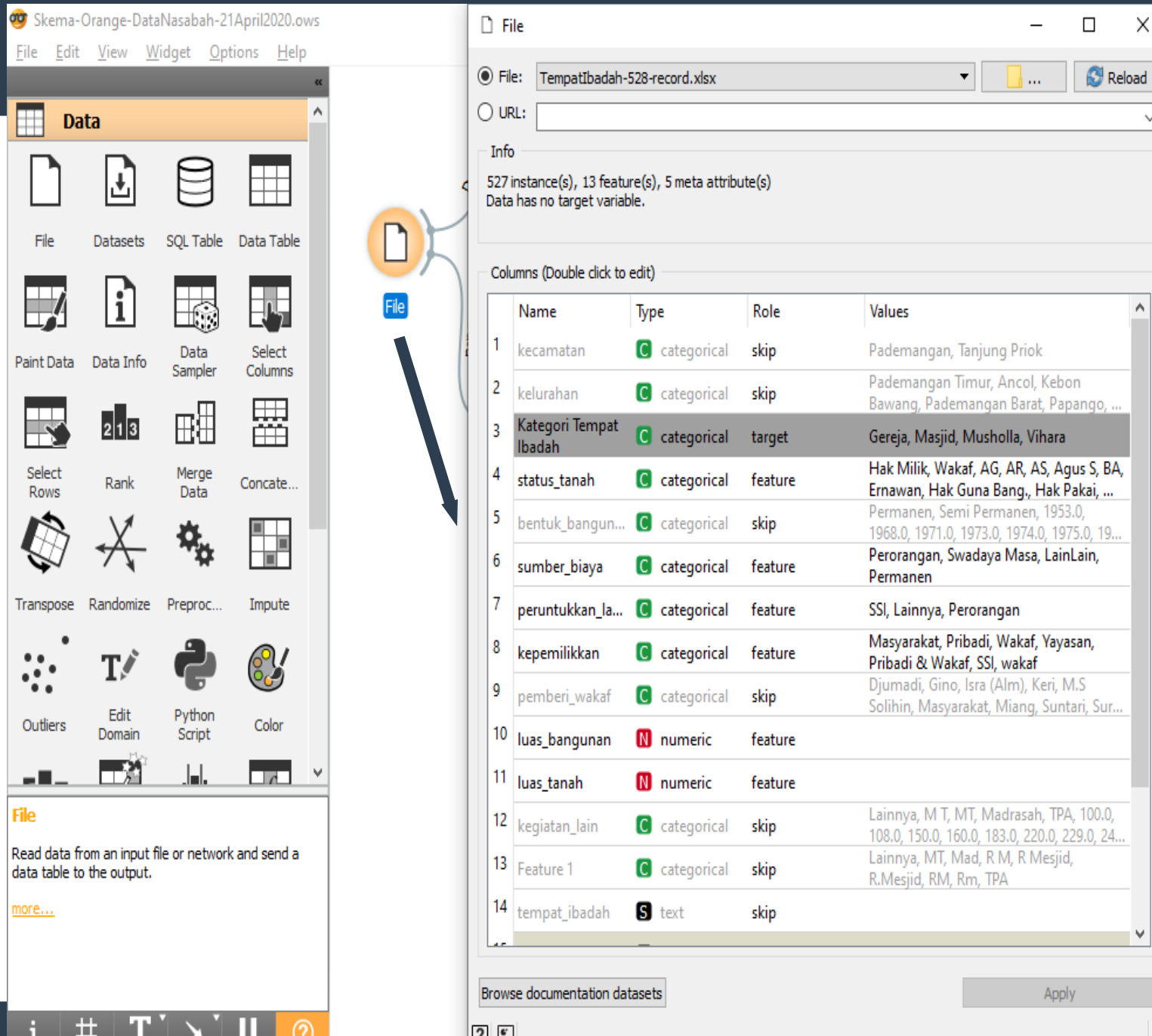
ID3 menggunakan Orange



Skema Orange



Widget File di Orange



Skema-Orange-DataNasabah-21April2020.ows

File Edit View Widget Options Help

Data

File Datasets SQL Table Data Table

Paint Data Data Info Data Sampler Select Columns

Select Rows Rank Merge Data Concat...

Transpose Randomize Preproc... Impute

Outliers Edit Domain Python Script Color

File

Read data from an input file or network and send a data table to the output.

[more...](#)

File

File: TempatIbadah-528-record.xlsx

URL:

Info

527 instance(s), 13 feature(s), 5 meta attribute(s)
Data has no target variable.

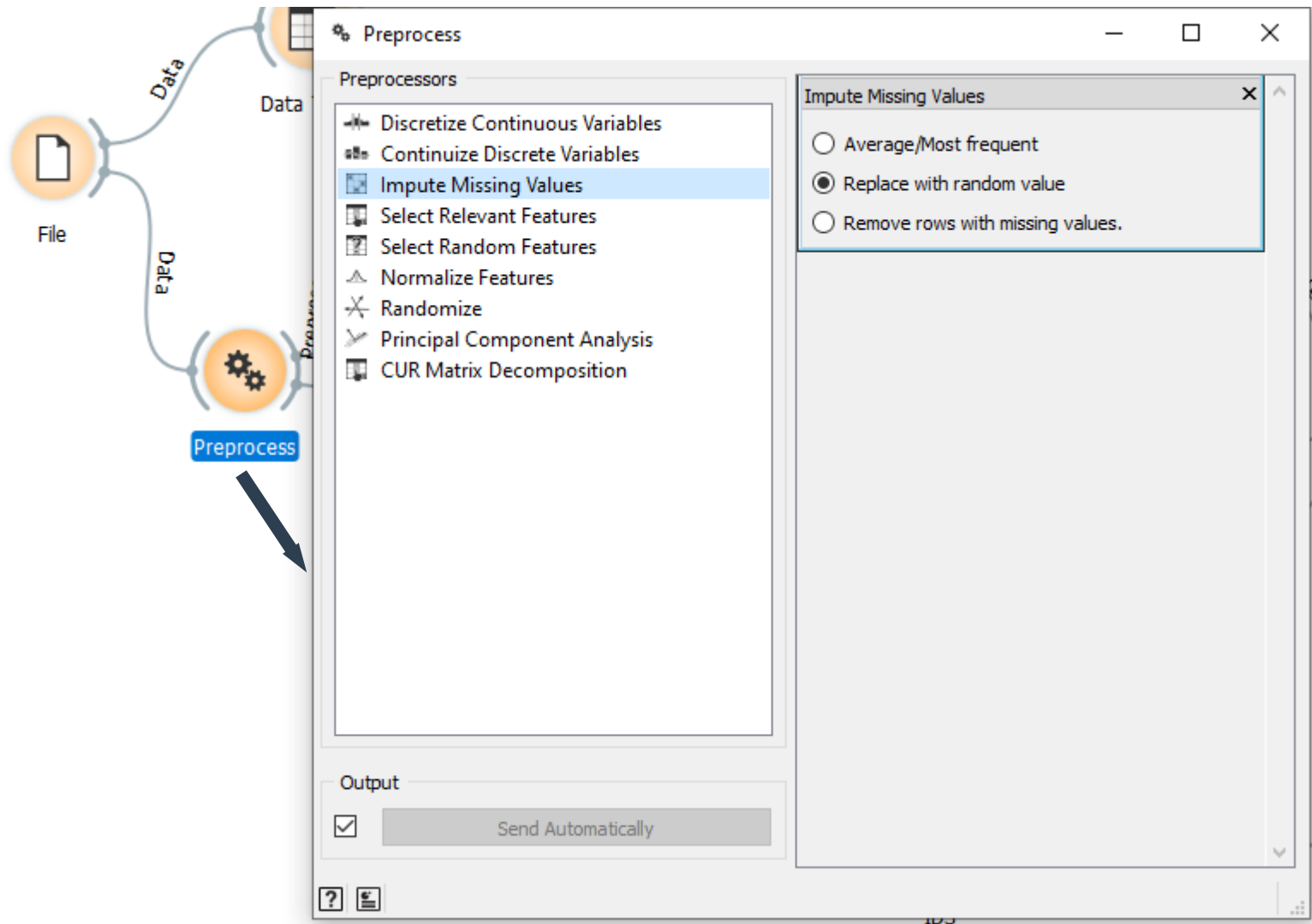
Columns (Double click to edit)

	Name	Type	Role	Values
1	kecamatan	C categorical	skip	Pademangan, Tanjung Priok
2	kelurahan	C categorical	skip	Pademangan Timur, Ancol, Kebon Bawang, Pademangan Barat, Papango, ...
3	Kategori Tempat Ibadah	C categorical	target	Gereja, Masjid, Musholla, Vihara
4	status_tanah	C categorical	feature	Hak Milik, Wakaf, AG, AR, AS, Agus S, BA, Ernawan, Hak Guna Bang., Hak Pakai, ...
5	bentuk_bangun...	C categorical	skip	Permanen, Semi Permanen, 1953.0, 1968.0, 1971.0, 1973.0, 1974.0, 1975.0, 19...
6	sumber_biaya	C categorical	feature	Perorangan, Swadaya Masa, LainLain, Permanen
7	peruntukkan_la...	C categorical	feature	SSI, Lainnya, Perorangan
8	kepemilikan	C categorical	feature	Masyarakat, Pribadi, Wakaf, Yayasan, Pribadi & Wakaf, SSI, wakaf
9	pemberi_wakaf	C categorical	skip	Djumadi, Gino, Isra (Alm), Keri, M.S Solihin, Masyarakat, Miang, Suntari, Sur...
10	luas_bangunan	N numeric	feature	
11	luas_tanah	N numeric	feature	
12	kegiatan_lain	C categorical	skip	Lainnya, M T, MT, Madrasah, TPA, 100.0, 108.0, 150.0, 160.0, 183.0, 220.0, 229.0, 24...
13	Feature 1	C categorical	skip	Lainnya, MT, Mad, R M, R Mesjid, R.Mesjid, RM, Rm, TPA
14	tempat_ibadah	S text	skip	

Browse documentation datasets

Apply

Preprocess



Preprocess

- **Impute Missing Values**

- Average/Most Frequent

- Mengganti atribut yang null/kosong dengan nilai rata-rata keseluruhan value pada atribut dengan nilai continue
 - Mengganti dengan nilai yang paling sering muncul untuk atribut bernilai diskrit

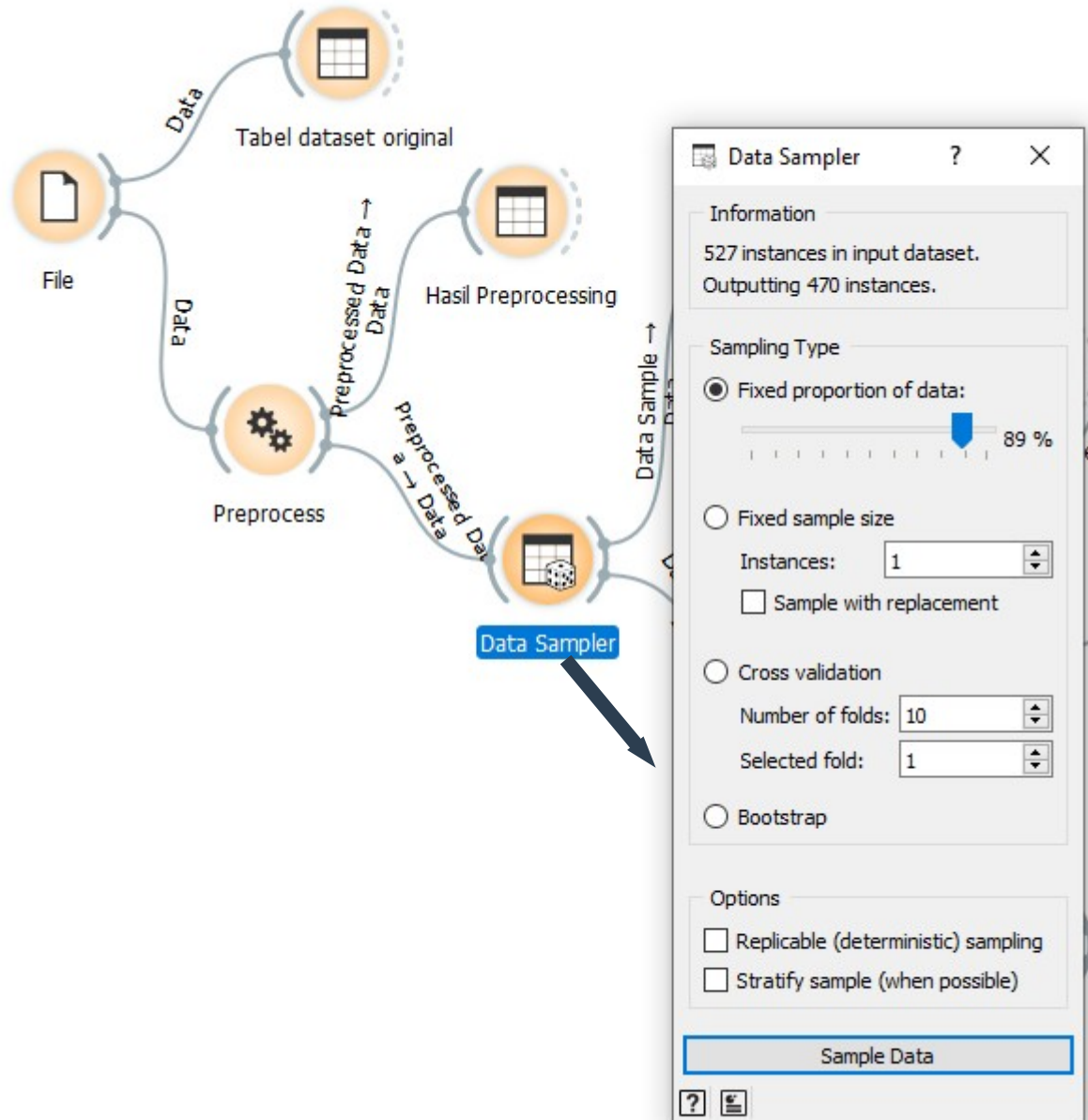
- Replace with Random Value

- Mengganti atribut yang null dengan nilai acak dalam kisaran variabel atribut tersebut

- Remove rows with missing values

- Menghapus record jika ada atribut yang null

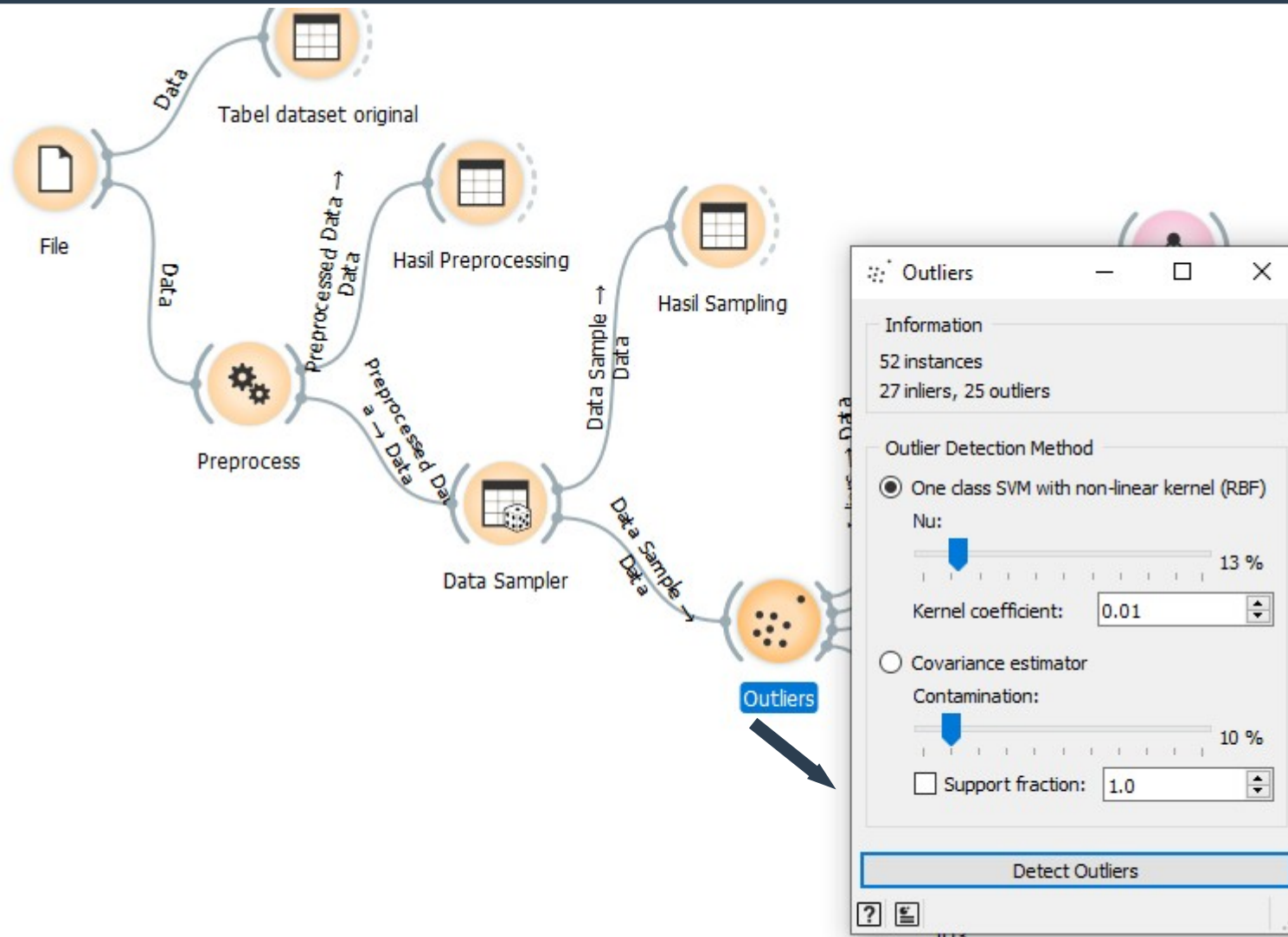
Data Sampler



Data Sampler

- Mengambil sampel dari jumlah populasi berdasarkan teknik sampling:
 - Fixed Proportion of Data: Mengambil $k\%$ dari total keseluruhan data (Populasi); misal: 70% dari 300 data
 - Fixed Sample Size : Mengambil sampel sejumlah nilai yang sudah ditentukan (Misal 100 dari 300 data)
 - Cross Validation: membagi data menjadi beberapa partisi (folds) dan memilih jumlah folds serta folds mana yang digunakan dalam sampel
 - Bootstrap: mengambil sampel dari persamaan statistik Bootstrap

Outlier



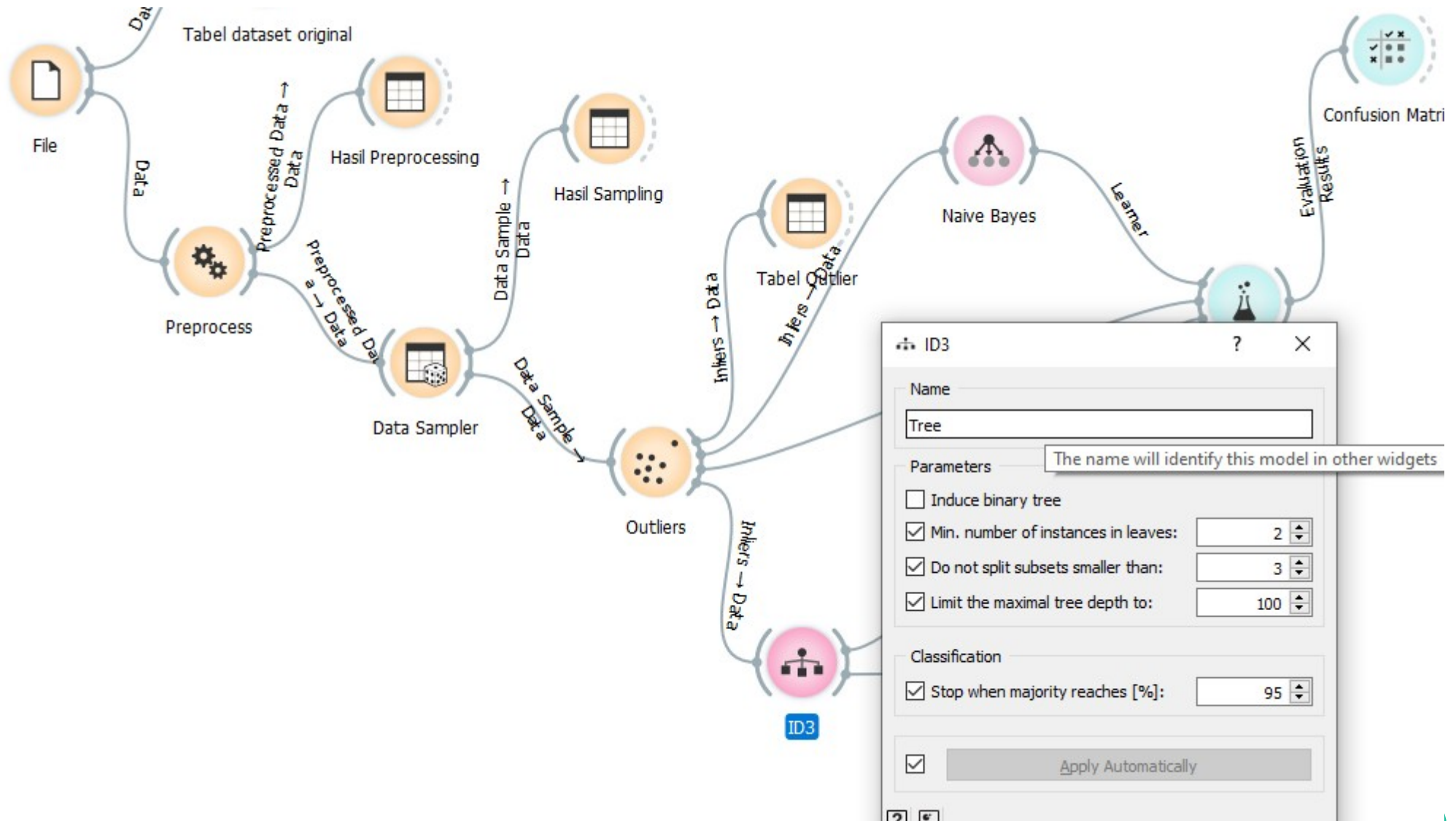
Outlier

- Widget untuk mendeteksi outliers dengan mendeteksi jarak antar elemen atau record
- Outliers: sekumpulan data yang diindikasikan outliers
- Inliers: sekumpulan data yang tidak diindikasikan outliers

Deteksi Outliers

- **SVM (Multiple kernel)**
 - One class SVM with non-linear kernel (RBF) menggunakan non-Gaussian distributions
- **Elliptical Envelope**
 - Covariance Estimator menggunakan distribusi Gaussian

Tree



Tree

- **Induce binary:** membuat cabang maksimal 2
- **Min.number of instances in leaves**
 - jika diceklist, algoritme tidak akan pernah membangun cabang yang akan menempatkan kurang dari jumlah contoh pelatihan yang ditentukan ke salah satu cabang.
- **Do not split subset smaller than**
 - Algoritme untuk membatasi dalam membagi node dengan jumlah instance lebih sedikit dari yang sudah ditentukan.
- **Limit the maximal tree depth to**
 - membatasi kedalaman pohon klasifikasi ke jumlah level node yang ditentukan.

Pengujian Klasifikasi

Test & Score

Sampling

Click on the table header to select shown columns

Ok, got it

☐ Cross validation

Number of folds: 20

☒ Stratified

☐ Cross validation by feature

☒ Random sampling

Repeat train/test: 2

Training set size: 80 %

☒ Stratified

☐ Leave one out

☐ Test on train data

☐ Test on test data

Target Class

Masjid

Method	AUC	CA	F1	Precision	Recall
Tree	0.803	0.731	0.653	0.579	0.750
Naive Bayes	0.779	0.723	0.633	0.574	0.705

