AI for Future Workforce

Machine Predictive Maintenance Classification

intel digital readiness

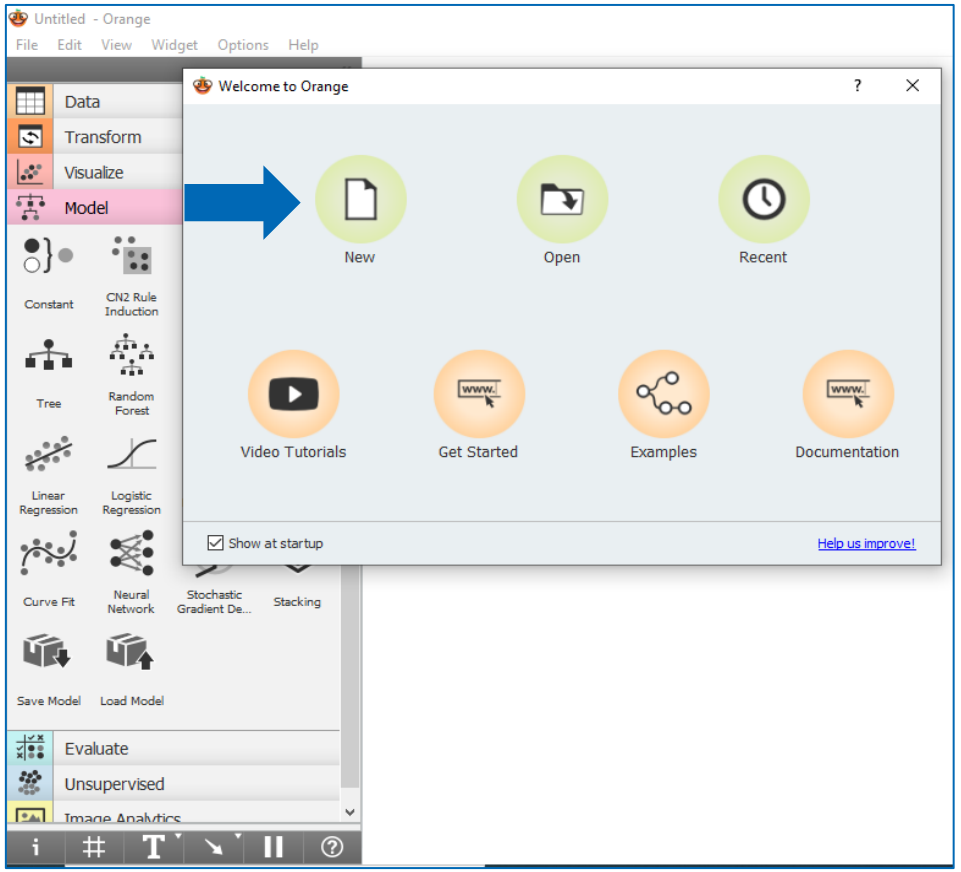
Use Case Implementation using Orange Data Mining

intel digital readiness

Use Case Implementation using Orange Data Mining

Create a New Project

- Welcome to Orange Page will pop up.
- Click on New to open a new.

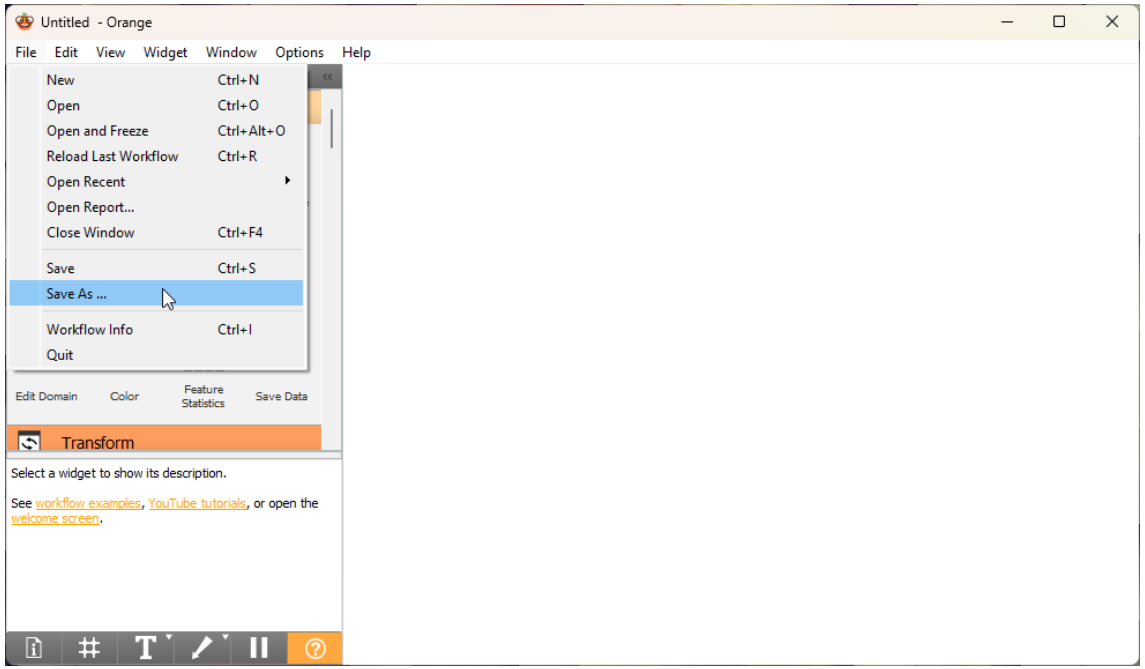


Use Case Implementation using Orange Data Mining

Create a New Project

Upload Dataset

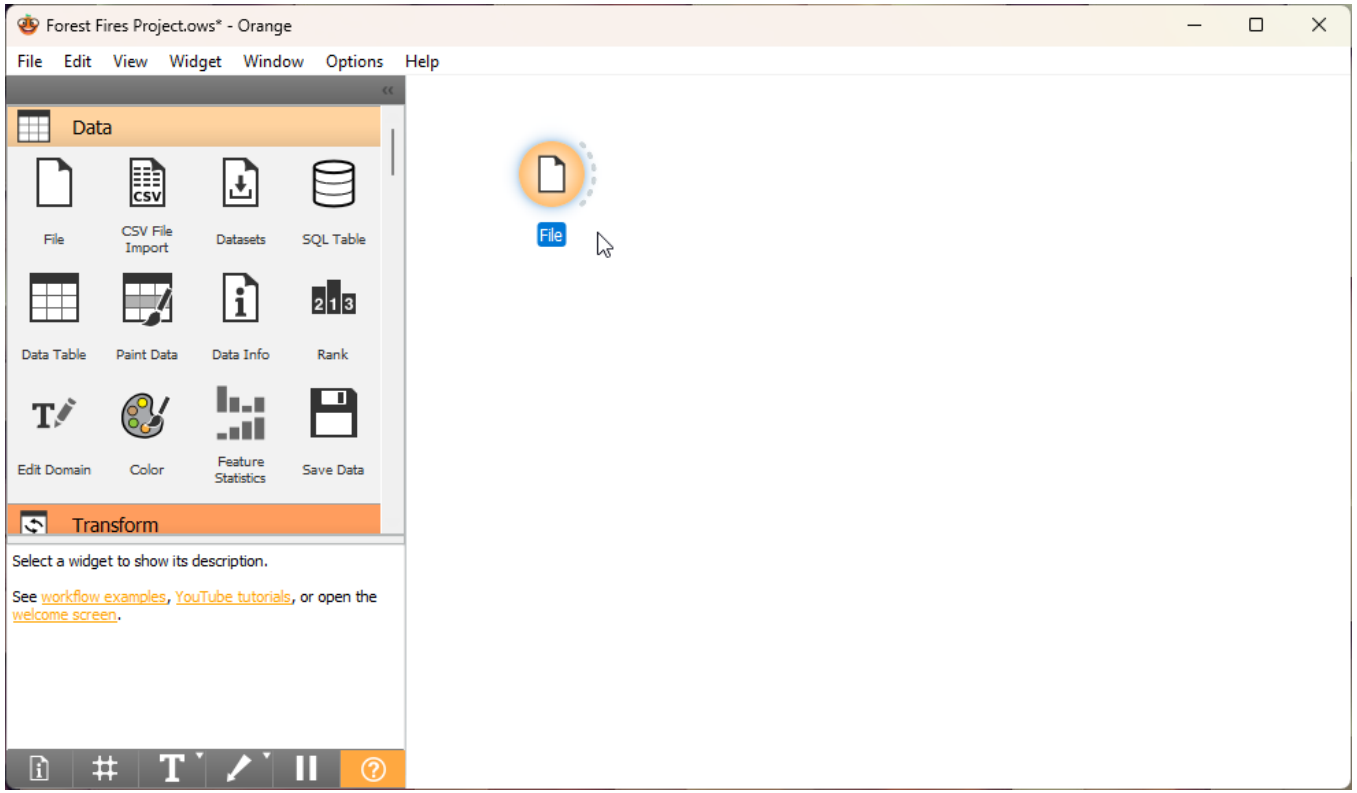
- Goto File and click Save As option to save your workflow



Use Case Implementation using Orange Data Mining

Data Acquisition

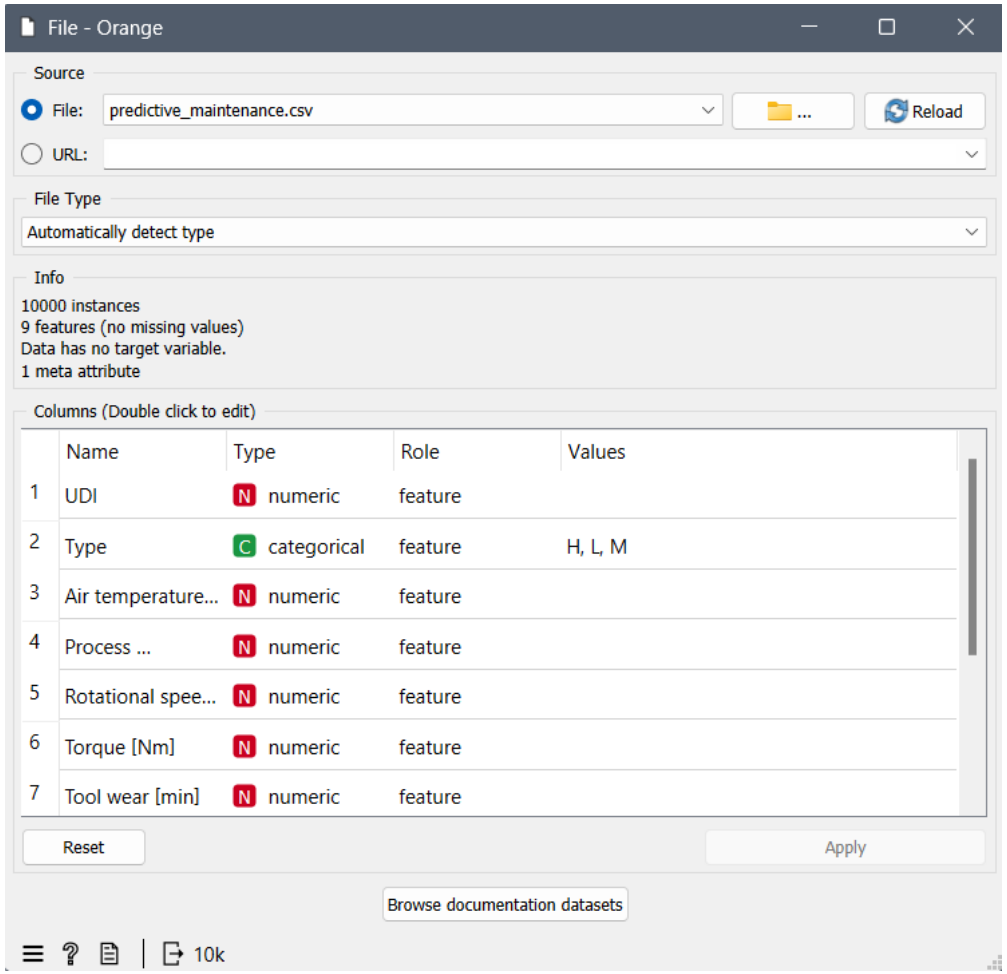
- Upload Dataset
  - Use the File widget (from Data category) for importing the dataset in the Orange Workflow.



Use Case Implementation using Orange Data Mining

Data Acquisition

- Upload Dataset
  - Double click on File widget and browse for the dataset – predictive_maintenance.csv
  - The dataset has 10k Rows & 10 Columns

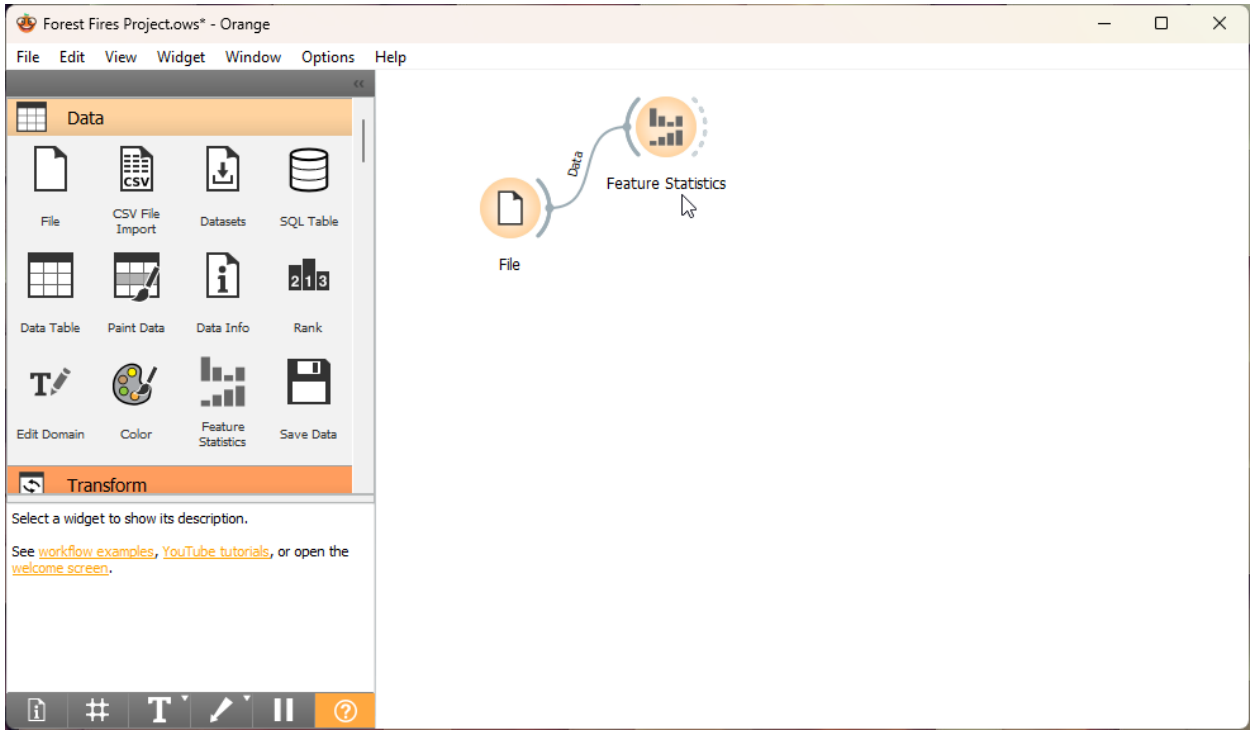


Use Case Implementation using Orange Data Mining

Data Exploration

Descriptive Summary

- We will use Feature Statistics widget to explore the dataset based on the descriptive summary of the columns
- Connect Feature Statistics with the File widget.

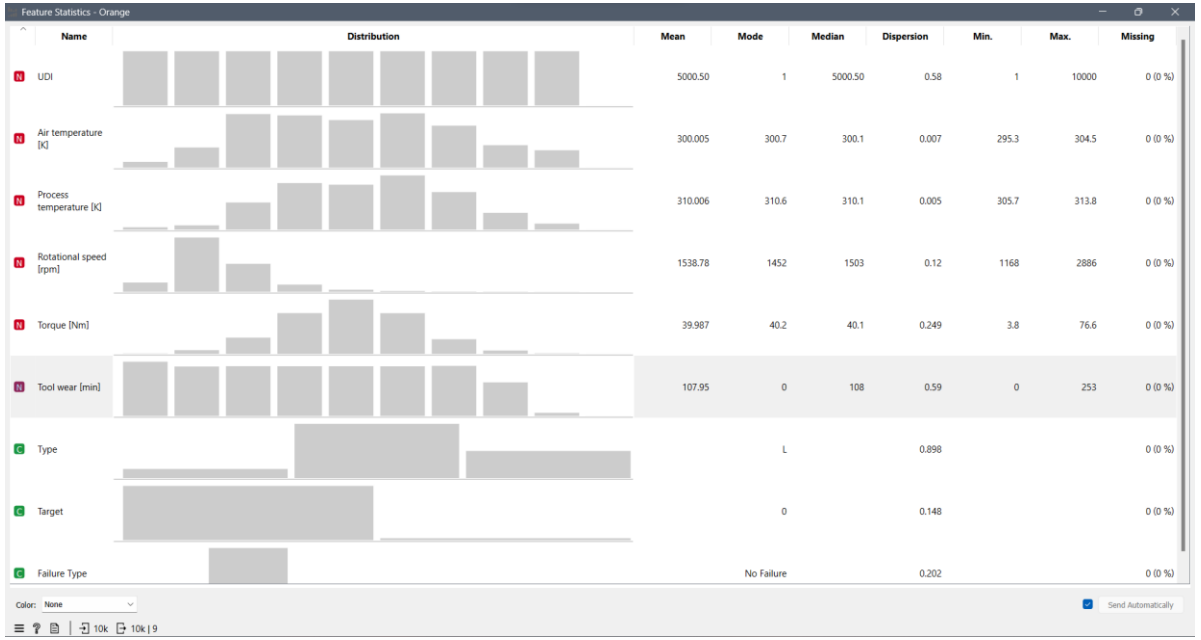


Use Case Implementation using Orange Data Mining

Data Exploration

Descriptive Summary

- The Feature Statistics widget gives us various statistical measures which clarify the data.
- Can you find out if the Failure Type column has any missing values?

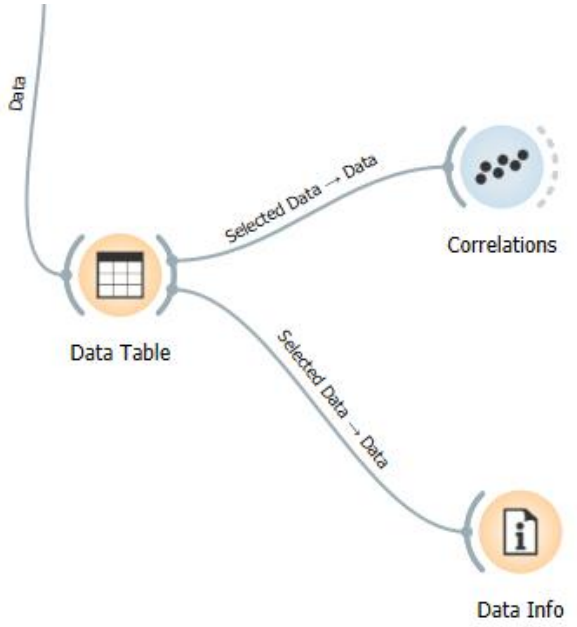


Use Case Implementation using Orange Data Mining

Data Exploration

Correlations

- We will now use Correlations widget and understand the linear relationship between the columns
- Connect Correlations widget with the File widget.

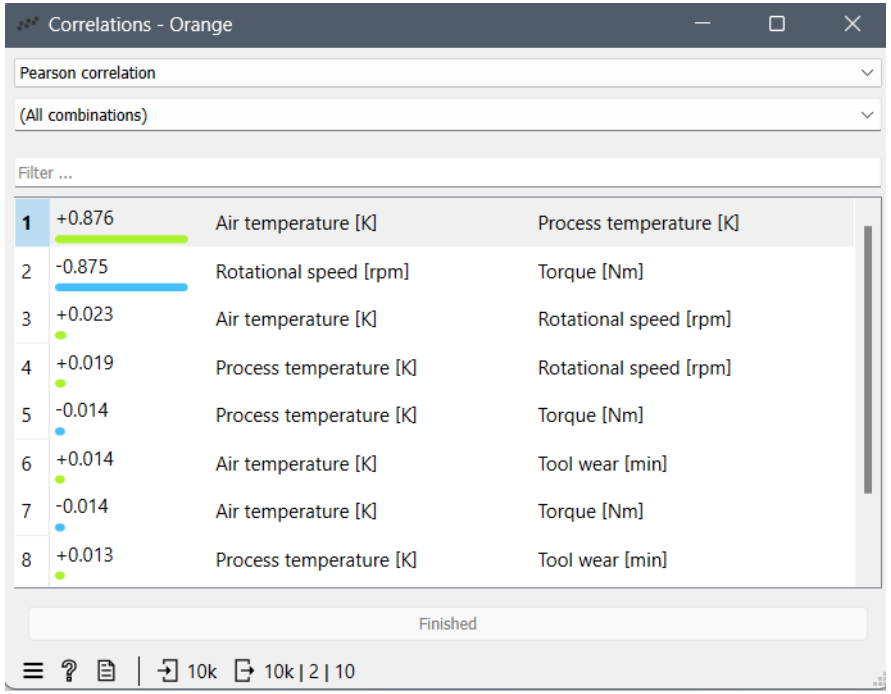


Use Case Implementation using Orange Data Mining

Data Exploration

Correlations

- 두 열 사이의 피어슨 상관 계수는 -1에서 +1 사이입니다.
- 상관 관계가 없다는 것은 상관 계수 값이 0일 때를 의미합니다.
- 양의 상관 관계는 상관 계수 값이 0보다 크고 1보다 작거나 같은 경우입니다.
- 음의 상관 관계는 상관 계수 값이 0보다 작고 -1보다 크거나 같은 경우입니다.
- 두 열 사이의 높은 상관 관계 값은 두 열이 독립적이지 않음을 의미합니다 (한 열의 값이 다른 열에 따라 변합니다).

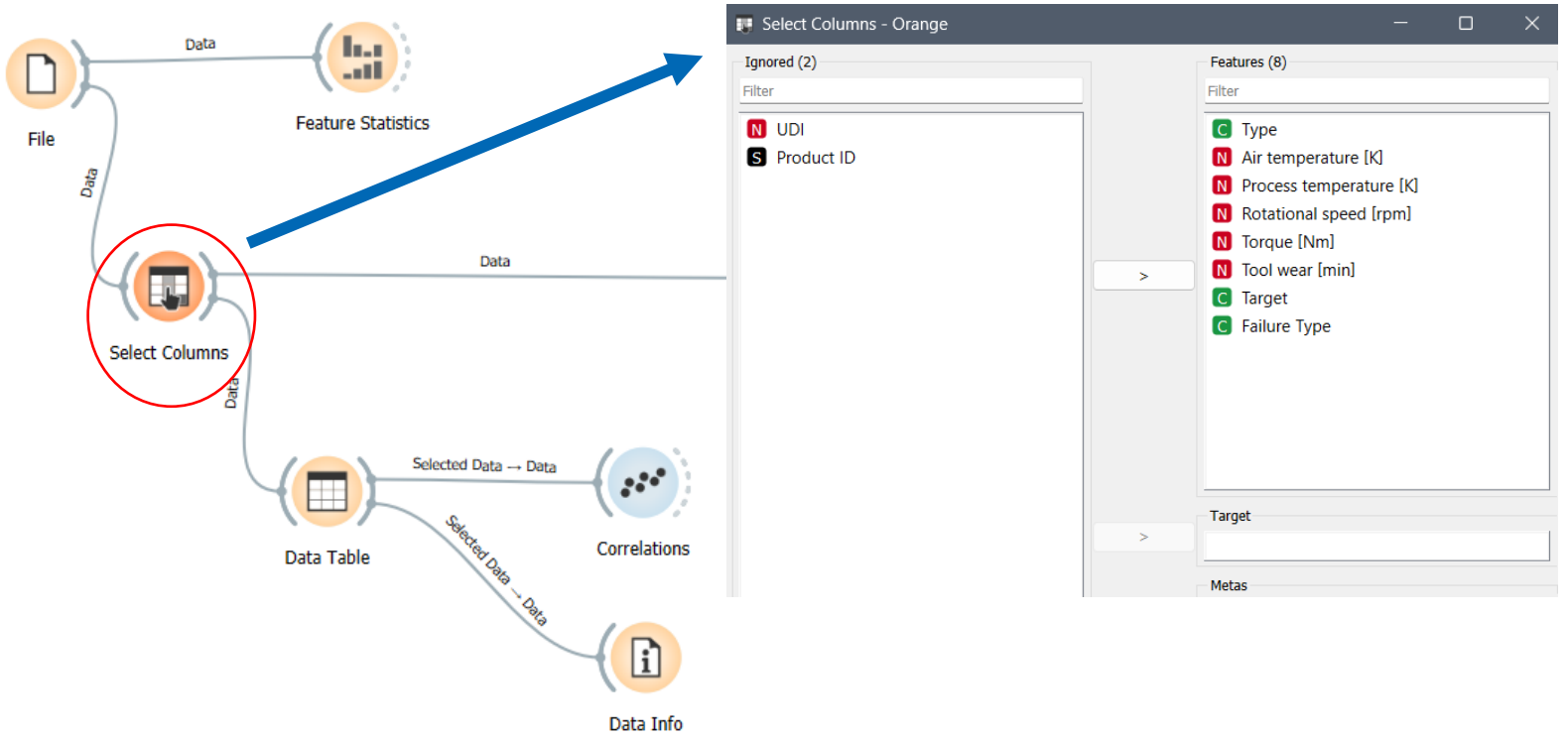


Use Case Implementation using Orange Data Mining

Data Pre-processing

Dropping the Columns

- Columns are dropped that do not contribute to the decision-making ability of ML models.

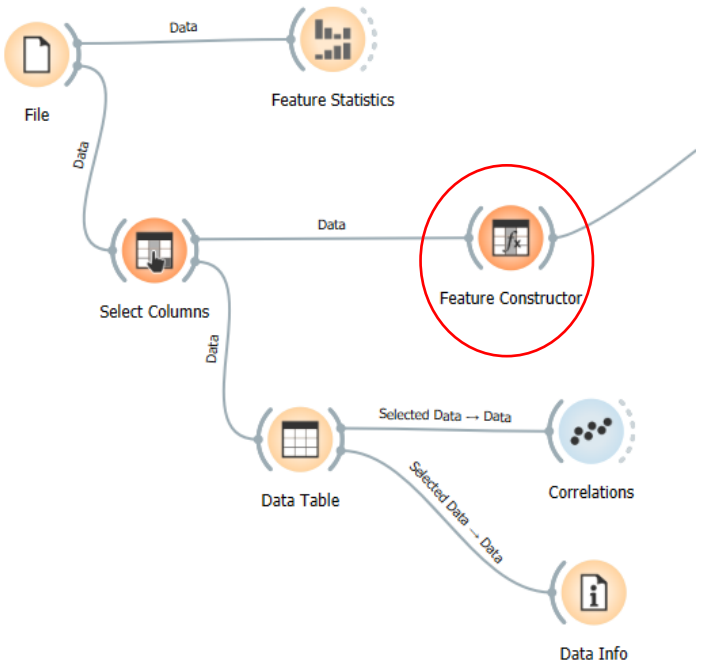


Use Case Implementation using Orange Data Mining

Data Pre-processing

Formula (Feature Constructor) Widget - Transform

- 머신 러닝 알고리즘은 입력 특징의 스케일에 민감합니다. 값의 범위가 큰 변수는 학습 과정을 지배하여 최적이 아닌 결과를 초래할 수 있습니다. 따라서 변환이 필요합니다

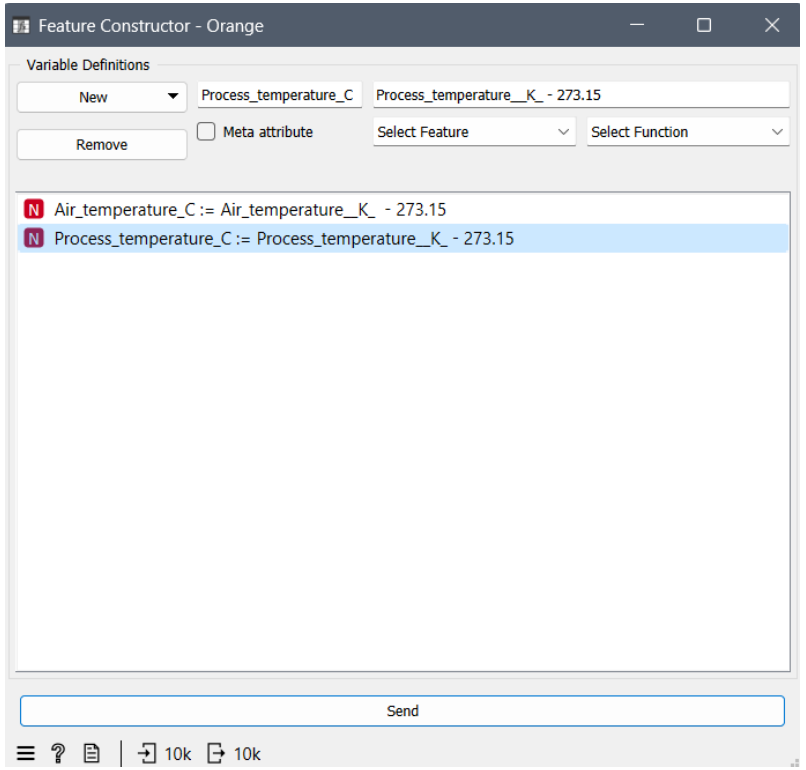


Use Case Implementation using Orange Data Mining

Data Pre-processing

Formula (Feature Constructor) Widget - Transform

- Orange의 'Formula' 위젯은 사용자 지정 표현식 및 계산을 통해 기존 특징을 기반으로 새로운 특징을 생성하여 데이터셋을 향상시킬 수 있게 합니다.

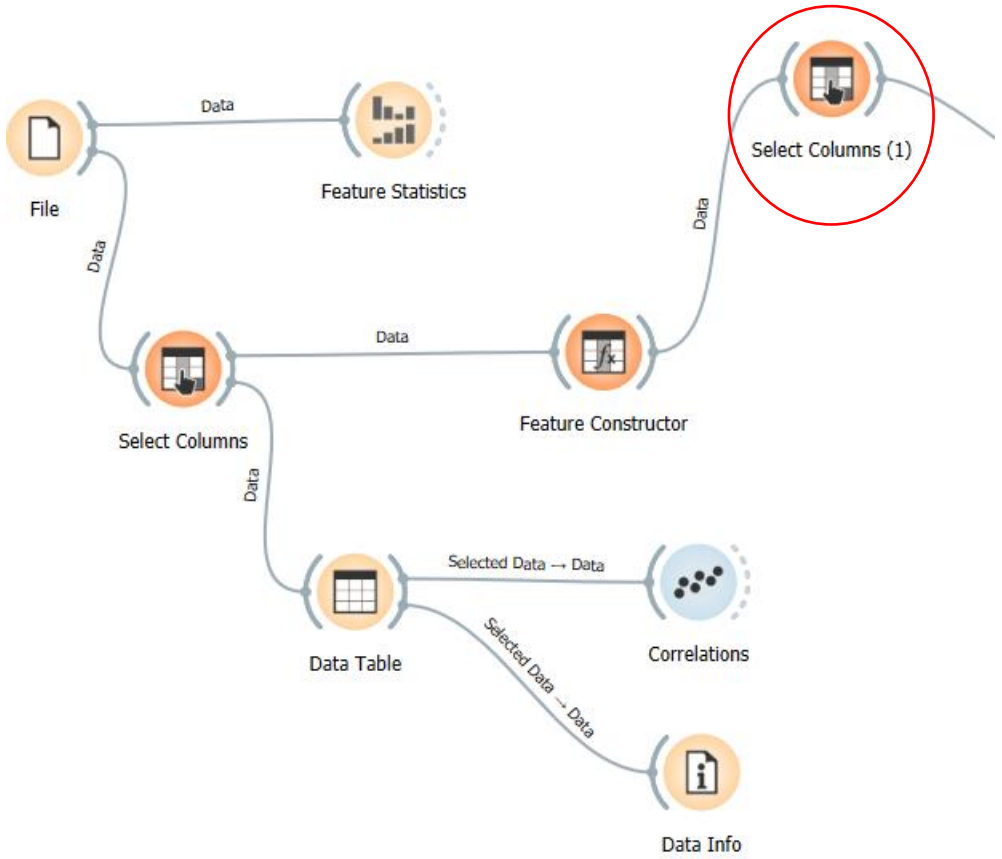


Use Case Implementation using Orange Data Mining

Data Pre-processing

Select Columns

- Choose the Select Columns widget to add it to your work area.
- We will use this widget to set up the fields we want to use in our analyses.

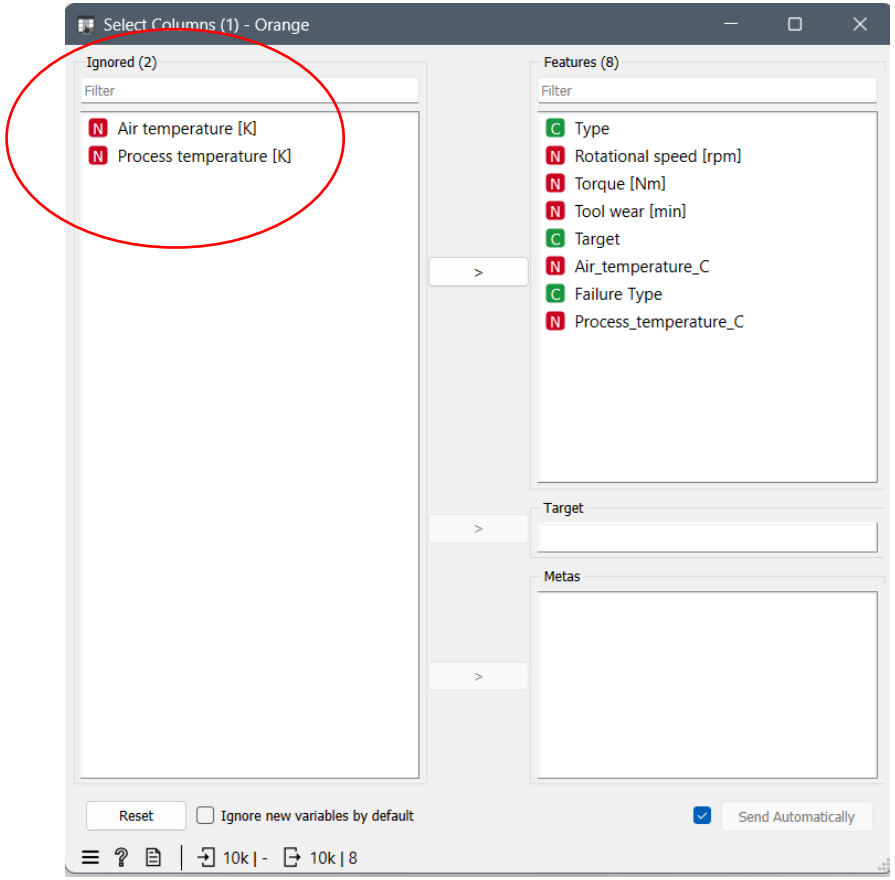


Use Case Implementation using Orange Data Mining

Data Pre-processing

Select Columns

- Dropping columns Air temperature [K], and Process temperature [K] as we have converted and added Air_temperature_C and Process_temperature_C.

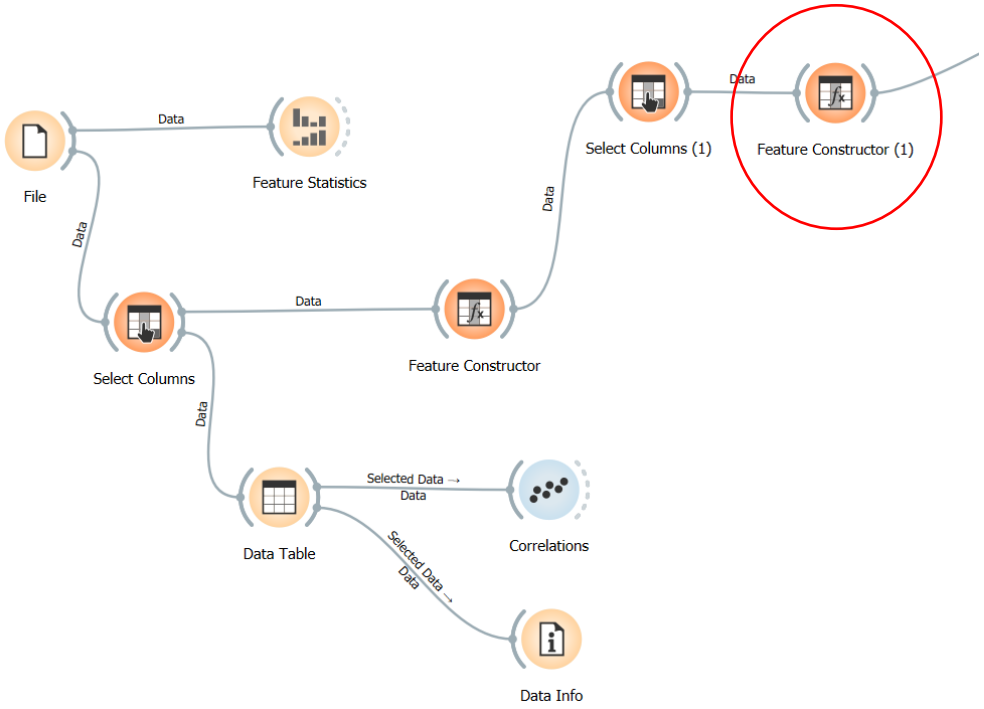


Use Case Implementation using Orange Data Mining

Feature Engineering

Formula (Feature Constructor) Widget - Transform

- The "Formula“ widget is a powerful component that allows users to create new features or attributes from existing data.

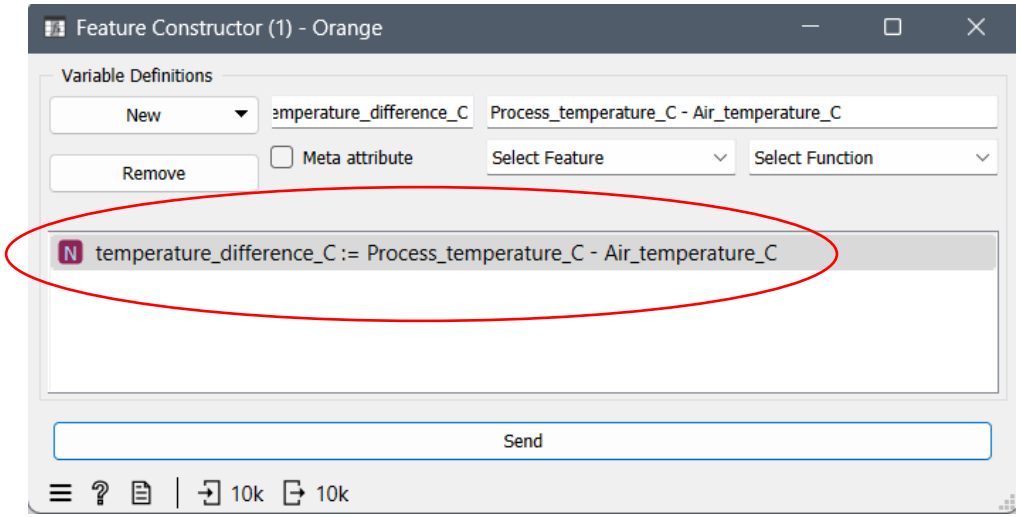


Use Case Implementation using Orange Data Mining

Feature Engineering

Formula (Feature Constructor) Widget - Transform

- Add a dependent feature named temperature_difference_C, which is the difference between Process_temperature_C and Air_temperature_C.

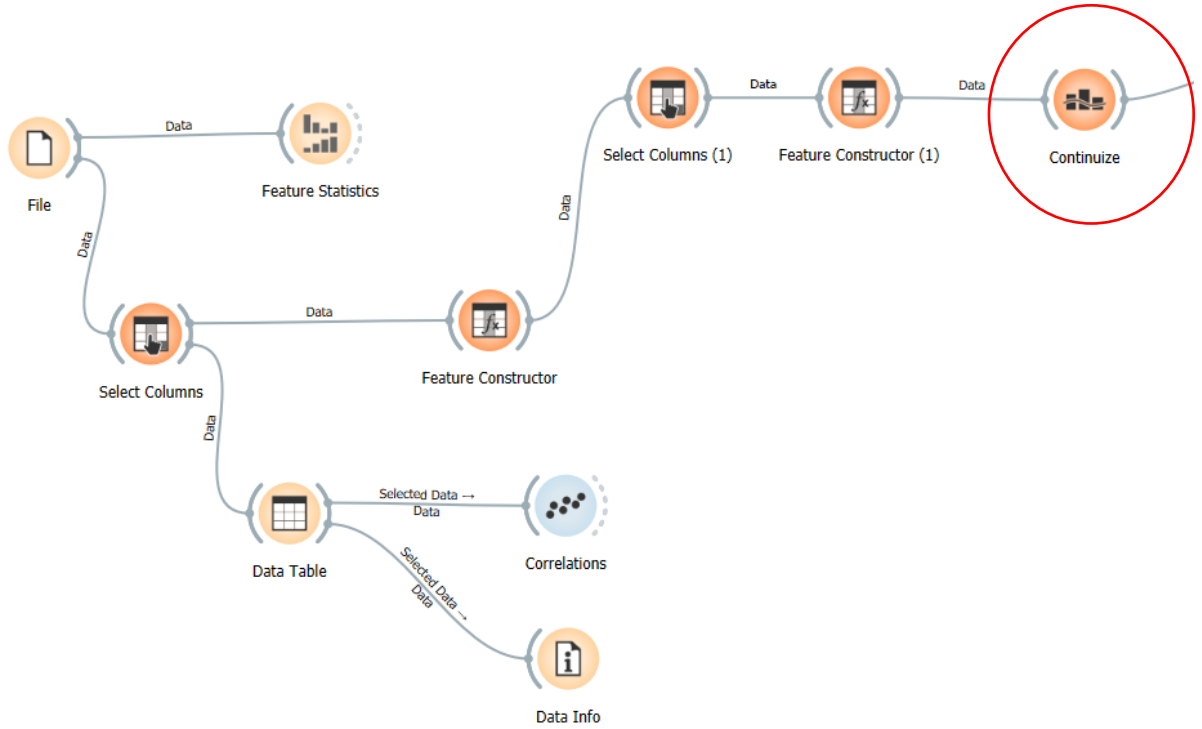


Use Case Implementation using Orange Data Mining

Feature Engineering

Continuize - Transform

- The "Continuize" widget in Orange is a powerful data preprocessing tool designed to handle categorical variables and transform them into continuous features.

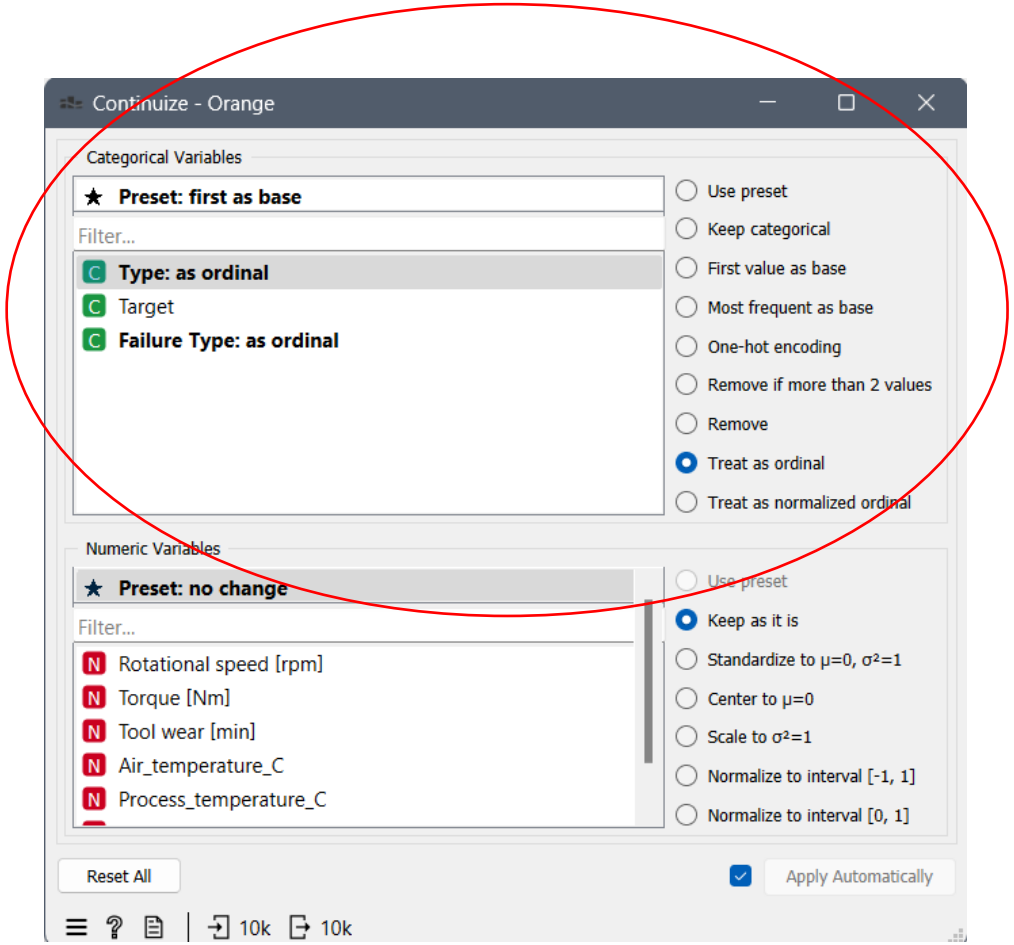


Use Case Implementation using Orange Data Mining

Feature Engineering

Continuize

- Label Encoding for "Type" and "Failure Type" features.

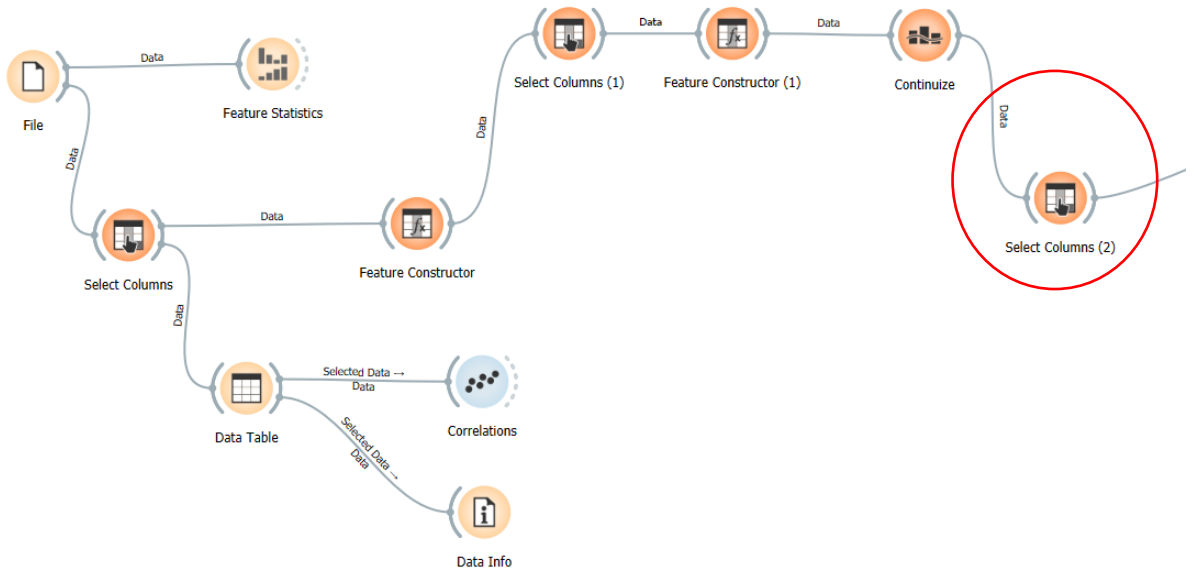


Use Case Implementation using Orange Data Mining

Feature Engineering

Select Columns

- Choose the Select Columns widget to add it to your work area.
- We will use this widget to set up the fields we want to use in our analyses.

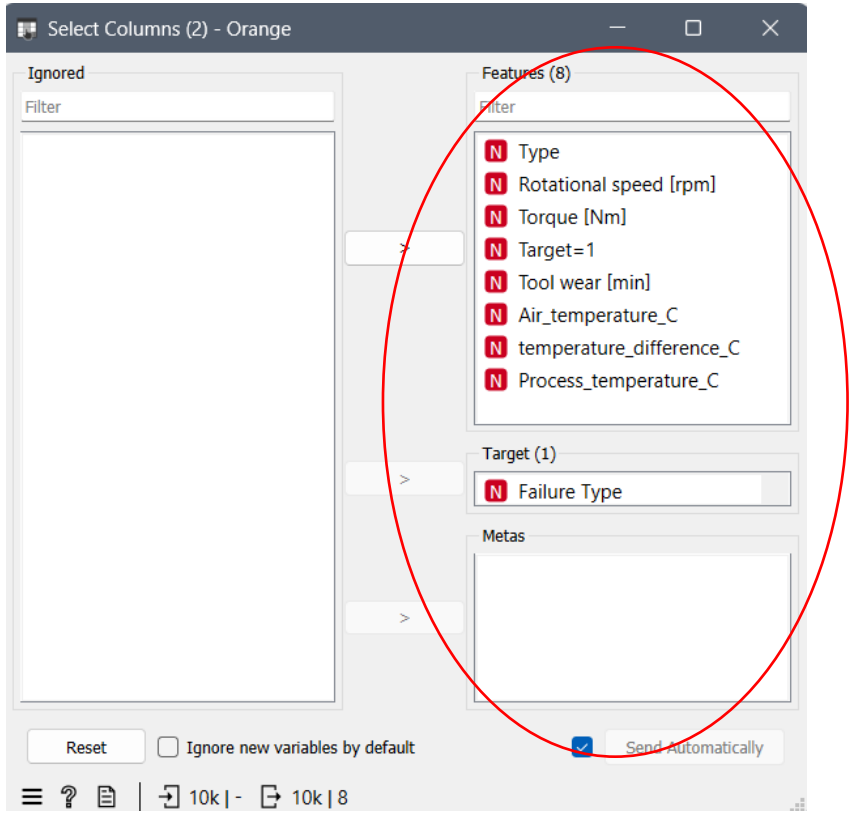


Use Case Implementation using Orange Data Mining

Feature Engineering

Select Columns

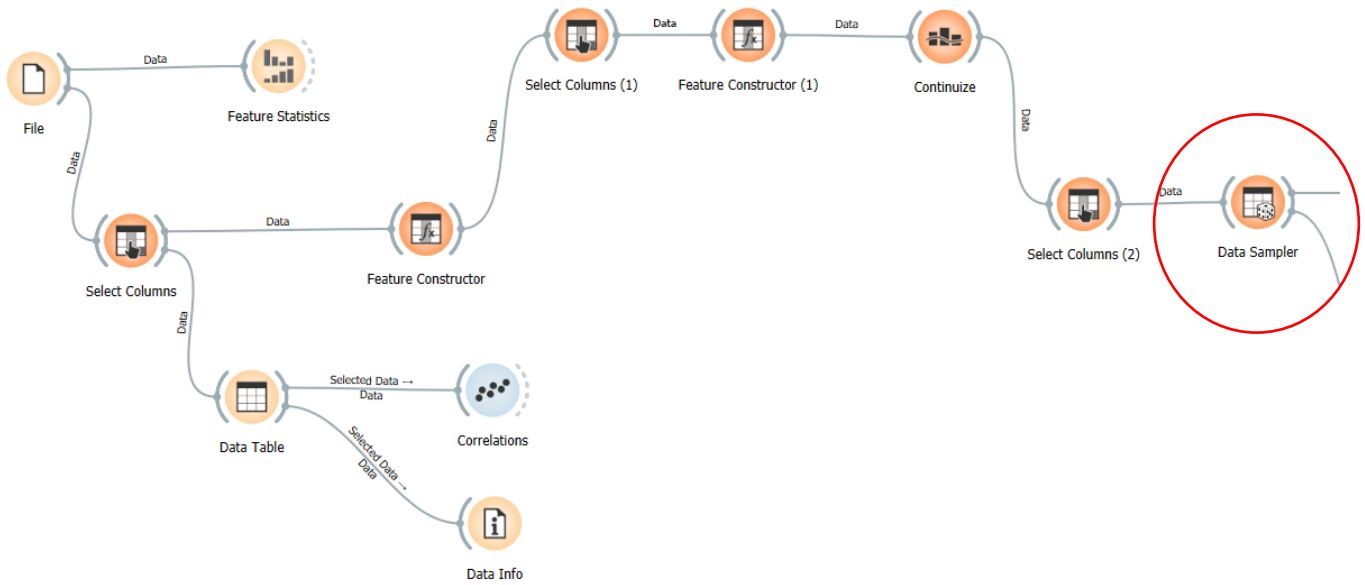
- Setting “Failure Type” column as the Target, and all other columns as the features.



Use Case Implementation using Orange Data Mining

Data Splitting → Train & Test

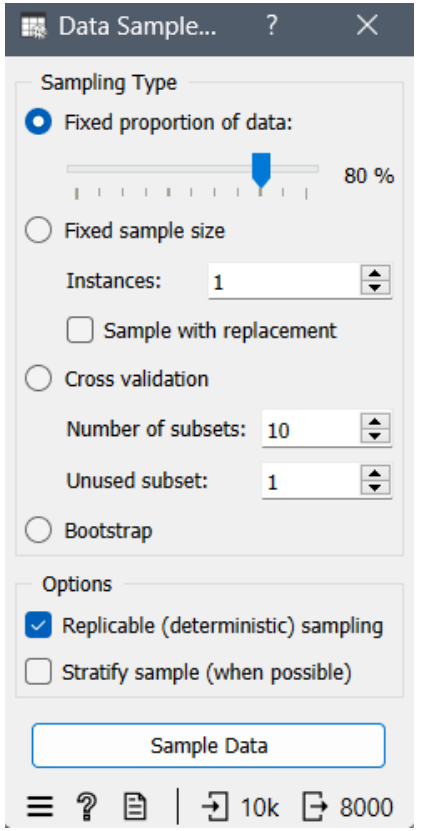
- Sampling- Data Sampler - Transform
  - Data Sampler widget must be used to select a sample of data based on our requirements.
  - Connect Data Sampler widget with Select Columns widget.



Use Case Implementation using Orange Data Mining

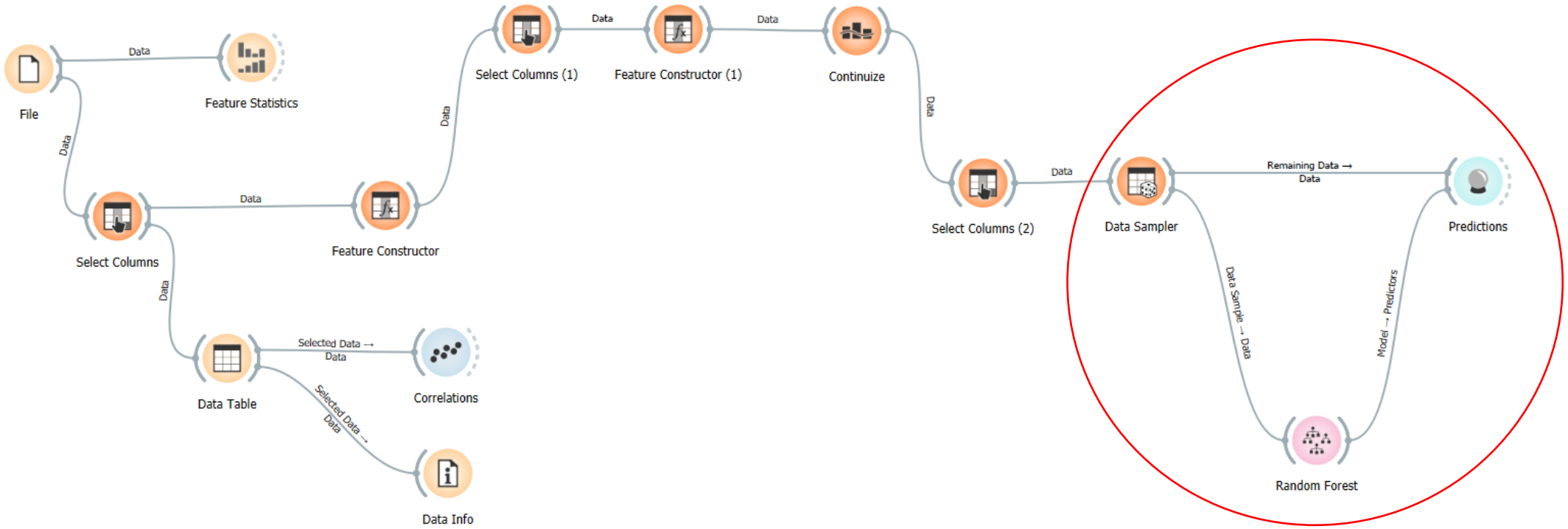
Data Splitting → Train & Test

- Sampling
  - Double click the Data Sampler widget and check the settings.
  - Our requirement is to sample the data in a ratio of 80%-20%
  - 80% of the data will be used for Training and the 20% for Testing.



Use Case Implementation using Orange Data Mining

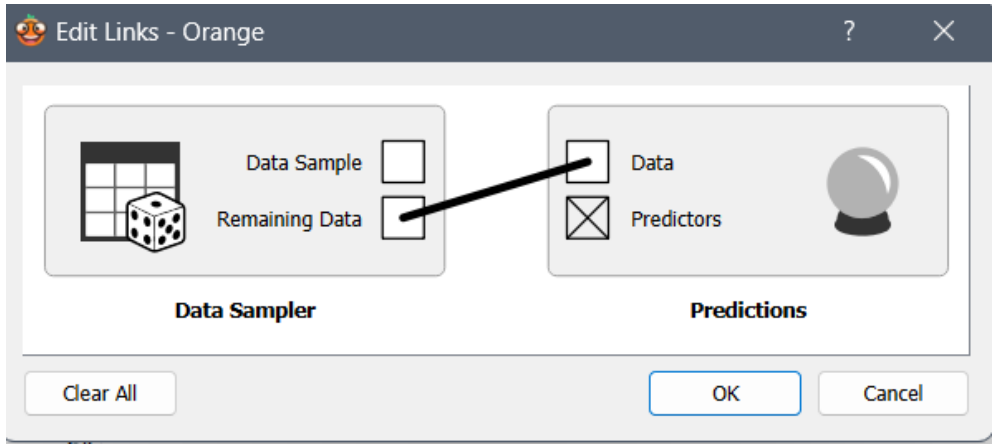
Data Splitting → Train & Test



Use Case Implementation using Orange Data Mining

Data Splitting → Train & Test

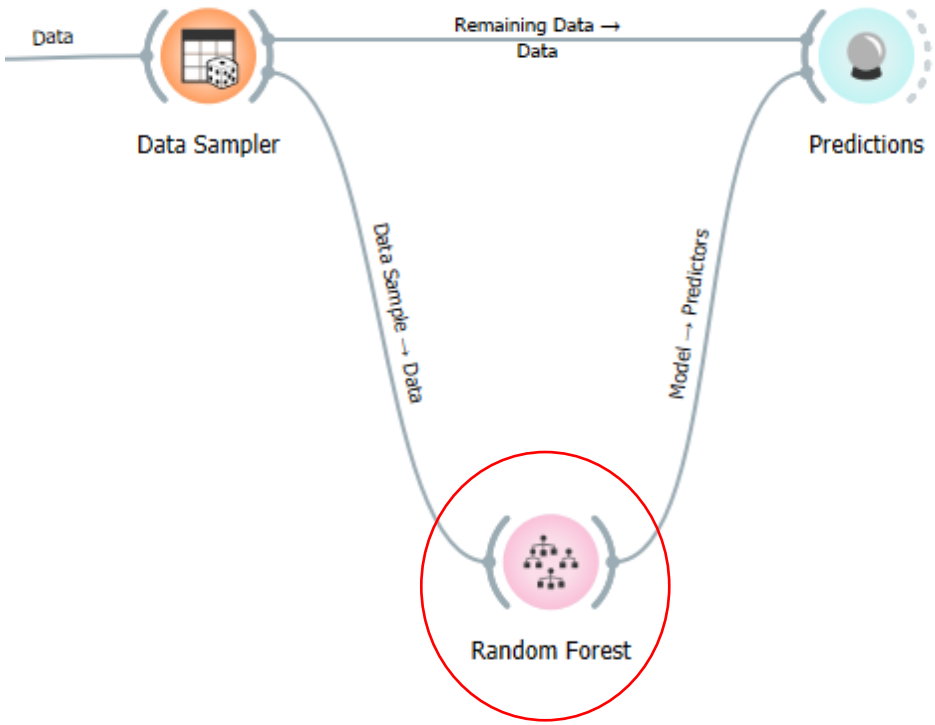
- Sampling
  - Double click on the connection line between the Data Sampler widget & Predictions widget
  - We will add the connection here from Remaining Data → Data of Predictions widget to make predictions out of the test data.



Use Case Implementation using Orange Data Mining

Model Building

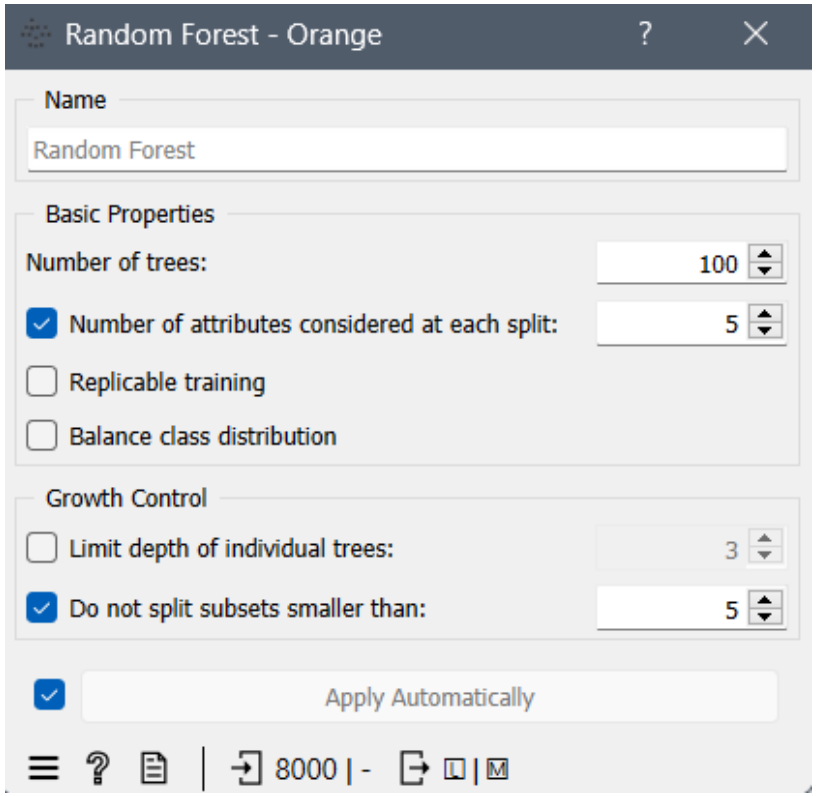
- Random Forest
  - Use Random Forest widget and connect it with Predictions widget and Data Sampler widget.
  - This will apply Random Forest model and we will be able to find the model scores.



Use Case Implementation using Orange Data Mining

Model Building

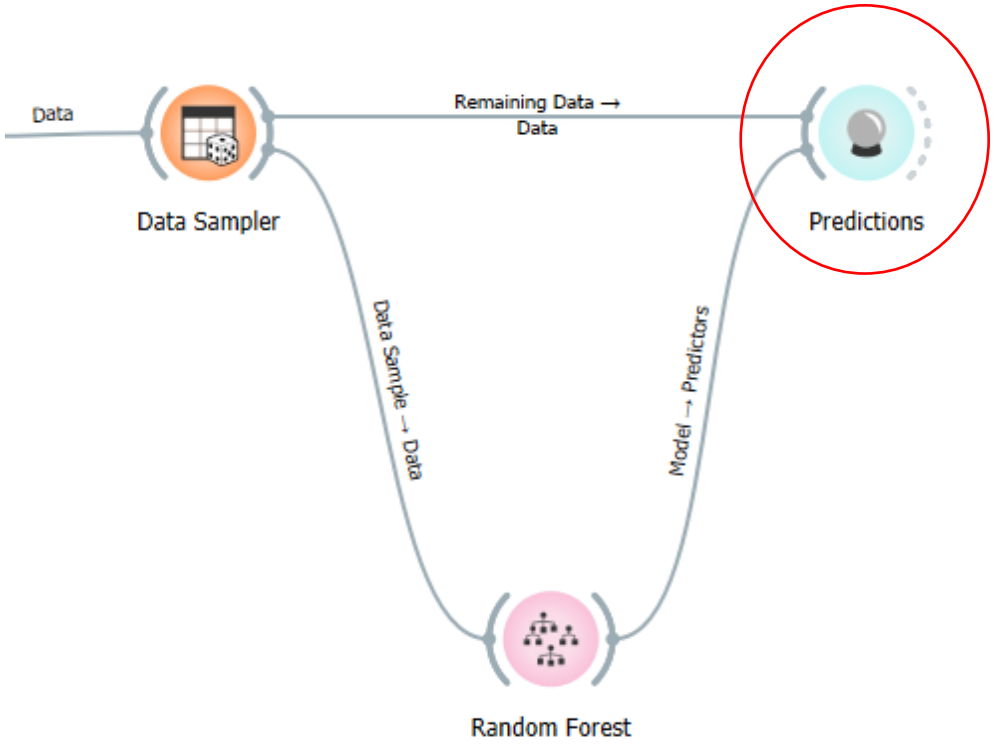
- Random Forest
  - Need to attach Data Sample to the Random Forest widget for training and Remaining Data to the Predictions widget so as make predictions.



Use Case Implementation using Orange Data Mining

Model Evaluation

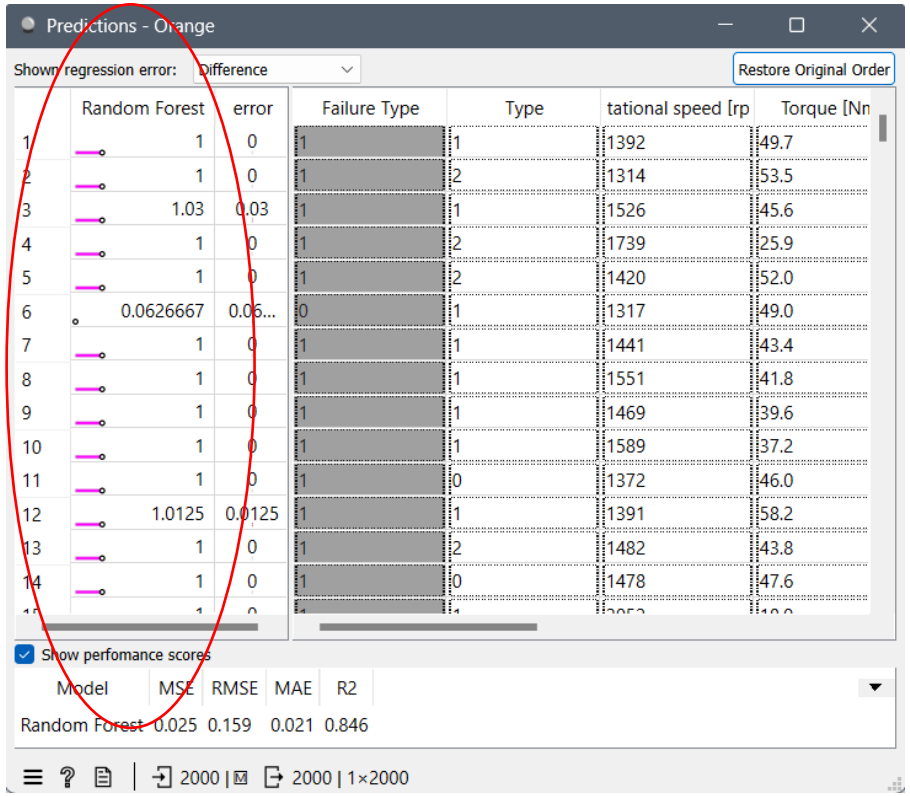
- Predictions
  - The Predictions widget in Orange is a powerful tool that allows users to make predictions using machine learning models.



Use Case Implementation using Orange Data Mining

Model Evaluation

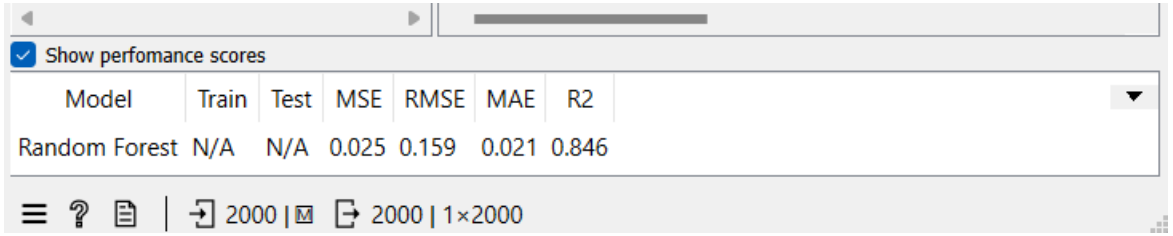
- Predictions
  - Here are the model predictions on test dataset [Remaining Data] with 2k rows.



Use Case Implementation using Orange Data Mining

Model Evaluation

- Random Forest
  - Double click on Predictions widget to the model scores.
  - A high RMSE means the model is bad
  - A high R2 means the model is good.




intel®
digital
readiness

Categorical Variables (범주형 변수)

범주형 변수에 대해 여러 가지 변환 옵션을 제공합니다:

1.Use preset (프리셋 사용):

1. 설정된 프리셋을 사용할지 여부를 선택합니다.

2.Keep categorical (범주형 유지):

1. 범주형 변수를 그대로 유지합니다. 별도의 변환 없이 사용됩니다.

3.First value as base (첫 번째 값을 기준으로):

1. 첫 번째 범주를 기준으로 하고, 나머지 범주를 이 기준에 대해 비교합니다.

4.Most frequent as base (가장 빈도가 높은 값을 기준으로):

1. 데이터에서 가장 자주 나타나는 범주를 기준으로 설정합니다.

5.One-hot encoding (원-핫 인코딩):

1. 각 범주를 이진 벡터로 변환합니다. 예를 들어, 3개의 범주가 있다면 각각의 범주는 [1, 0, 0], [0, 1, 0], [0, 0, 1]로 변환됩니다.

6.Remove if more than 2 values (2개 이상의 값이 있으면 제거):

1. 2개 이상의 범주형 값을 가지는 변수는 제거됩니다.

7.Remove (제거):

1. 선택된 범주형 변수를 데이터에서 완전히 제거합니다.

8.Treat as ordinal (서열형으로 처리):

1. 범주형 변수를 서열형 변수로 처리합니다. 즉, 범주들 사이에 순서가 있다고 가정합니다.

9.Treat as normalized ordinal (정규화된 서열형으로 처리):

1. 서열형 변수로 처리하되, 이를 0과 1 사이의 값으로 정규화합니다.

Numeric Variables (수치형 변수)

수치형 변수를 어떻게 변환할지 선택할 수 있는 옵션들입니다:

1.Use preset (프리셋 사용):

1. 설정된 프리셋을 사용할지 여부를 선택합니다.

2.Keep as it is (그대로 유지):

1. 수치형 변수를 그대로 유지합니다. 별도의 변환 없이 사용됩니다.

3.Standardize to $\mu=0$, $\sigma^2=1$ (표준화):

1. 평균을 0으로, 분산을 1로 하는 표준화 변환을 적용합니다.

4.Center to $\mu=0$ (평균을 0으로 중심화):

1. 변수의 평균을 0으로 맞춥니다. 단위 분산으로 변환하지는 않습니다.

5.Scale to $\sigma^2=1$ (분산을 1로 스케일):

1. 분산을 1로 맞춥니다. 평균은 변화시키지 않습니다.

6.Normalize to interval [-1, 1] (구간 [-1, 1]로 정규화):

1. 수치형 변수를 -1과 1 사이의 값으로 정규화합니다.

7.Normalize to interval [0, 1] (구간 [0, 1]로 정규화):

1. 수치형 변수를 0과 1 사이의 값으로 정규화합니다.