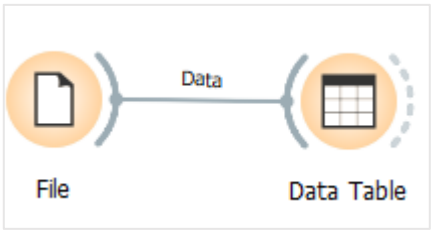
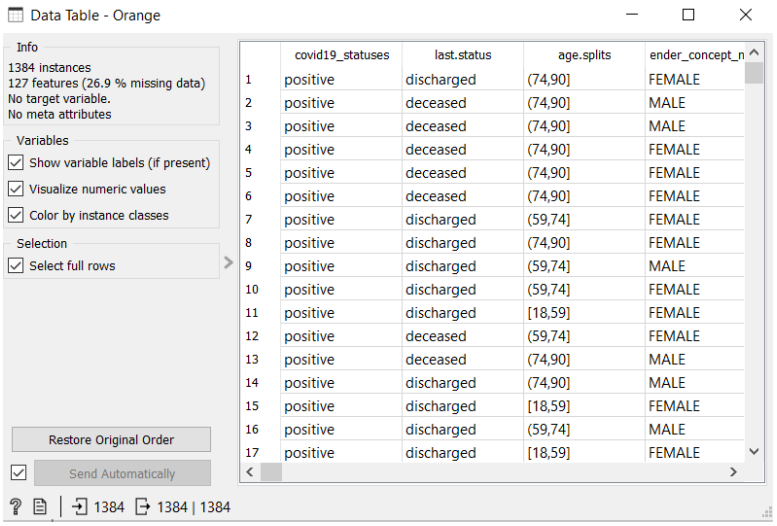
1. Gathering the data collection

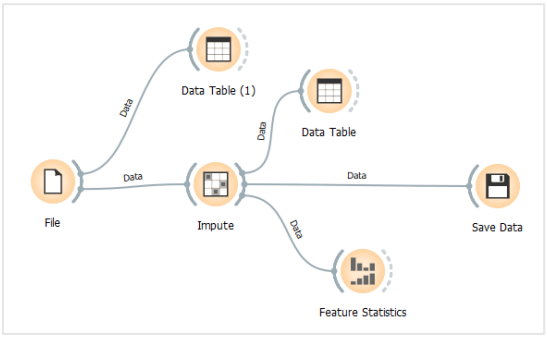


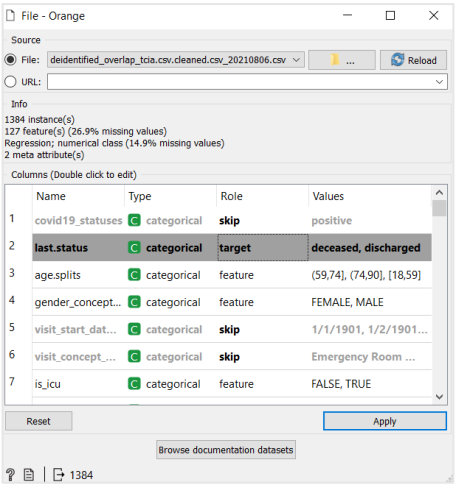


	covid19_statuses	last.status	age.splits	ender_concept_n
1	positive	discharged	(74,90]	FEMALE
2	positive	deceased	(74,90]	MALE
3	positive	deceased	(74,90]	MALE
4	positive	deceased	(74,90]	FEMALE
5	positive	deceased	(74,90]	FEMALE
6	positive	deceased	(74,90]	FEMALE
7	positive	discharged	(59,74]	FEMALE
8	positive	discharged	(74,90]	FEMALE
9	positive	discharged	(59,74]	MALE
10	positive	discharged	(59,74]	FEMALE
11	positive	discharged	[18,59]	FEMALE
12	positive	deceased	(59,74]	FEMALE
13	positive	deceased	(74,90]	MALE
14	positive	discharged	(74,90]	MALE
15	positive	discharged	[18,59]	FEMALE
16	positive	discharged	(59,74]	MALE
17	positive	discharged	[18,59]	FEMALE

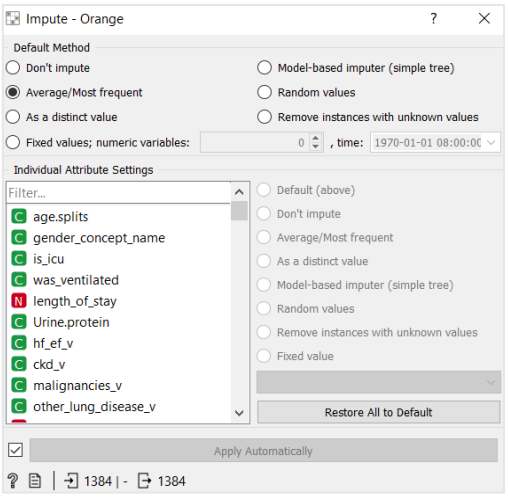
- By linking file data to the Data Table widget, can better understand the data and select the attributes as a wish. As a result, there are 26.9% missing values, 127 characteristics, and 1383 instances in total.

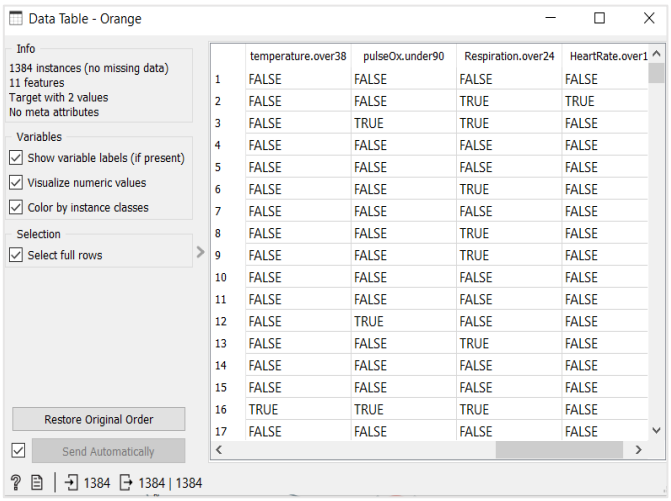
2. Data Cleaning

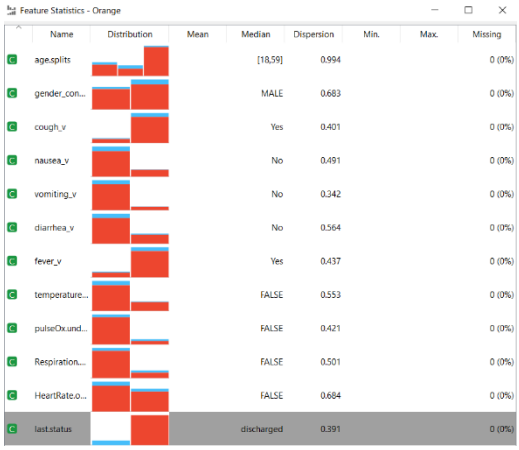


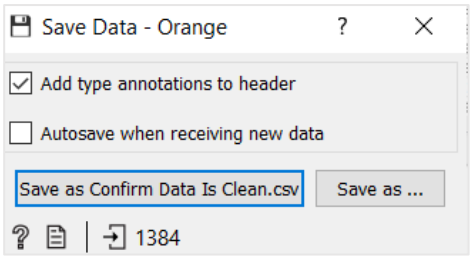


- Use for selected or drop any features that aren't relevant in the File widget for preparing datasets, and choose target variable in the column role. Then, using the Imput widget, replace the noisy or missing data with "Average/Most Frequent" and restore the data.



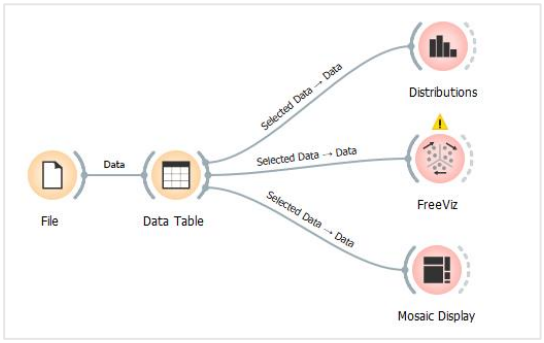


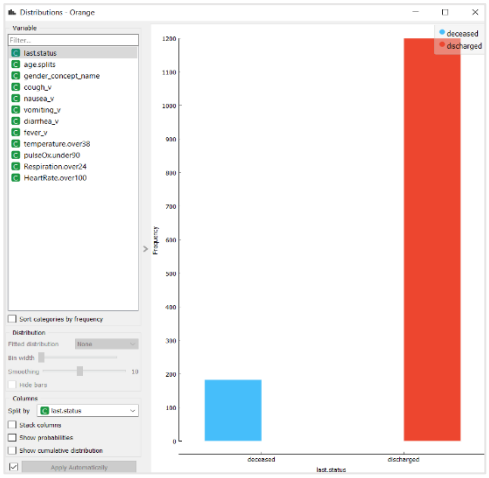


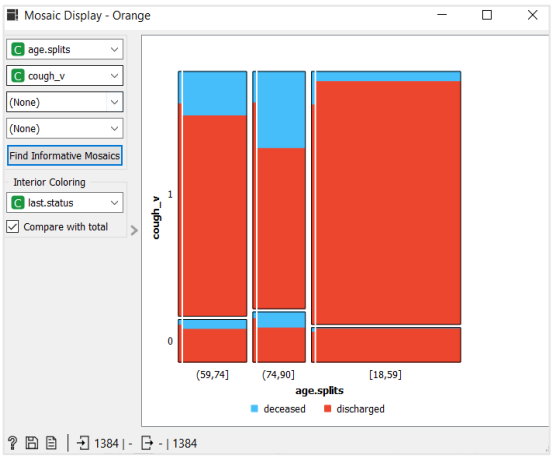


- After utilising the Impute Widget, the Feature Statistics widget displays that no missing data was found.
- After ensuring that the data is clean and that the target variable has been selected, save the new data collection using the Save Data widget so that it can be utilised in the next step.

3. Data Preparation



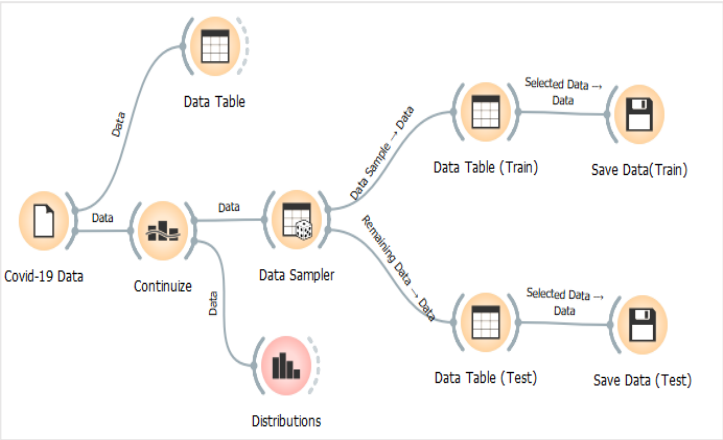


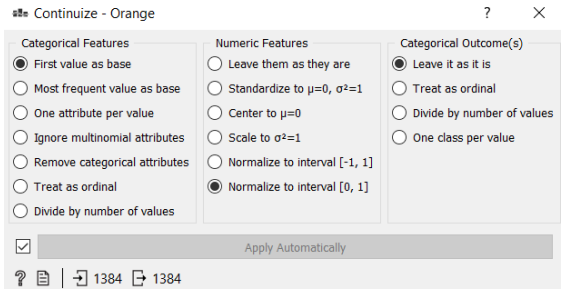


- The Distribution widget, FreeViz widget, and Mosaic Display widget can be used to visualise the features specified for a better understanding.
- The display from the FreeViz widget shows there is no missing data with 11 features selected and 1384 instances. The target with two values represents the last status feature, which can be "Discharged" or "Deceased."
- The Feature Statistics widget provides an explanation of the visualisation and allows to examine the distribution, meaning, media, and dispersion of each feature name. Then, in the next step, connect the Impute widget to the Save Data widget.

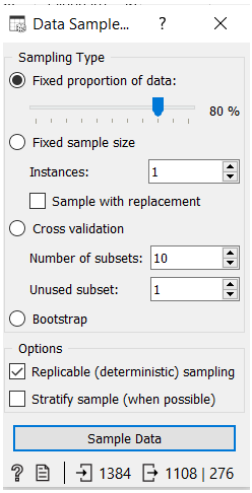
4. Choosing A Model For Training and Testing Data

a) Splitting the data



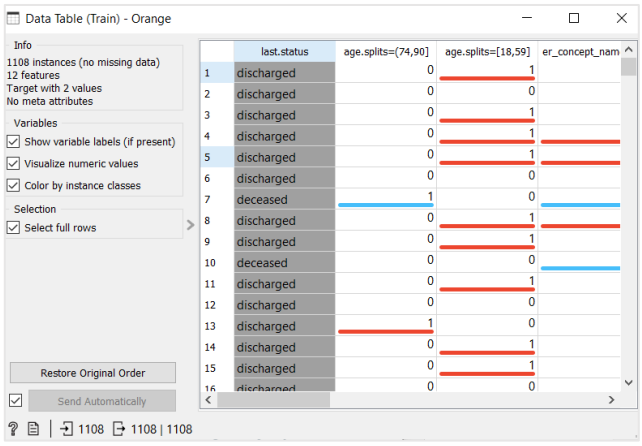


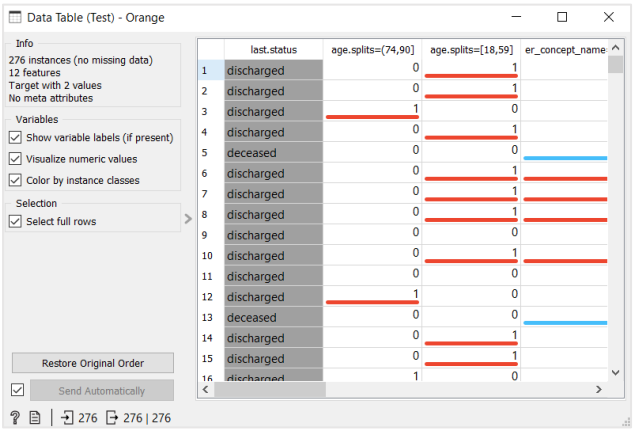
- First and foremost, the dataset must be split. In this stage, convert a categorical feature to a numerical one, such as [Yes:1, No:0]. By clicking "Normaliza to interval [0,1], can alter the Continuize widget.



-

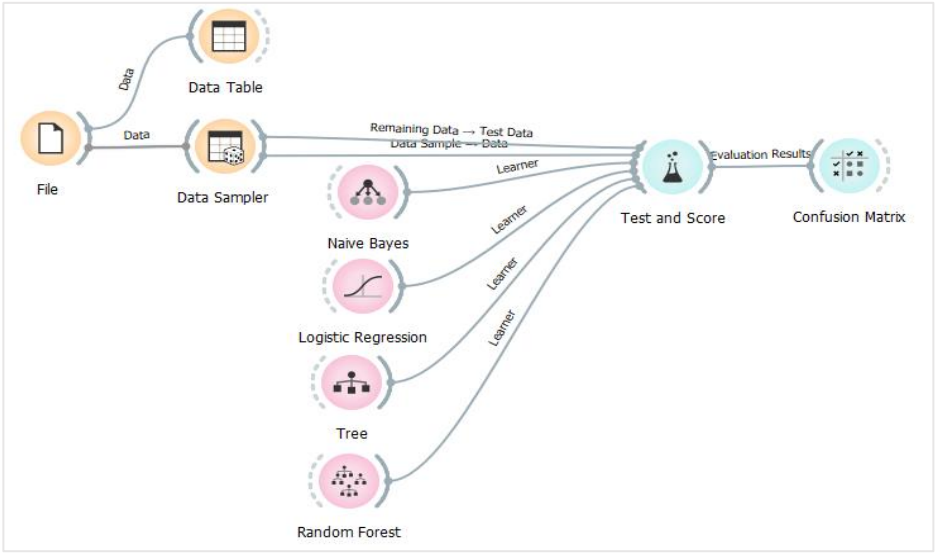
- Then divide the data into 80:20 training and testing data (1108) (276).

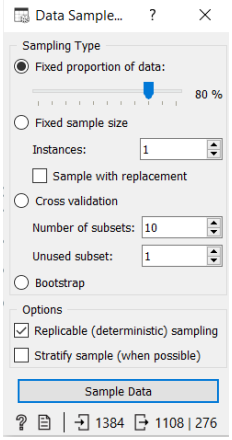




- For some of the features, each figure shows that the training data, which is 1108, and the testing data, which is 276 have been transformed into numerical form. The data can then be saved in the Save Data widget and used for prediction.

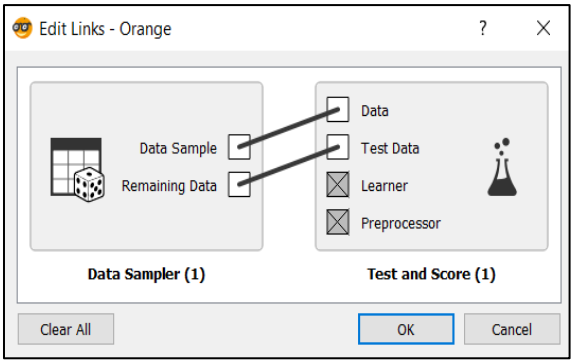
b) Choose the model to evaluation





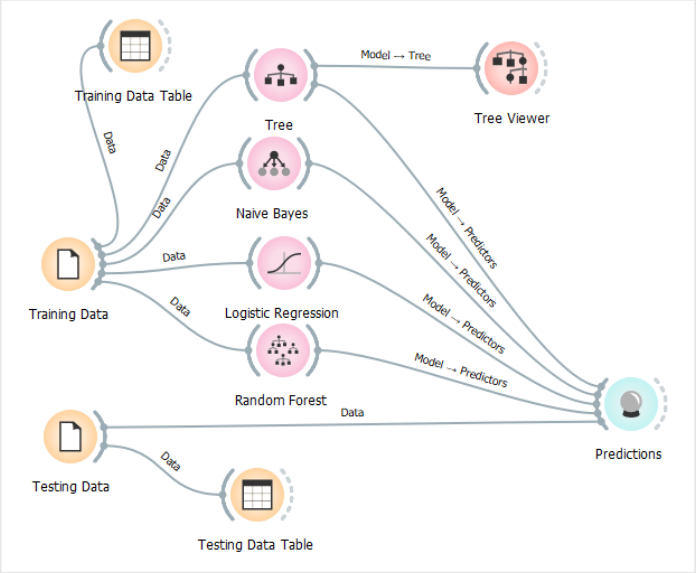
- Alternatively, the Data Sampler widget performs the same function as sklearn's train test split.
- In the Confusion Matrix widget, four classifications were connected to the Test and Score widget to analyse the results and compare the classification prediction.

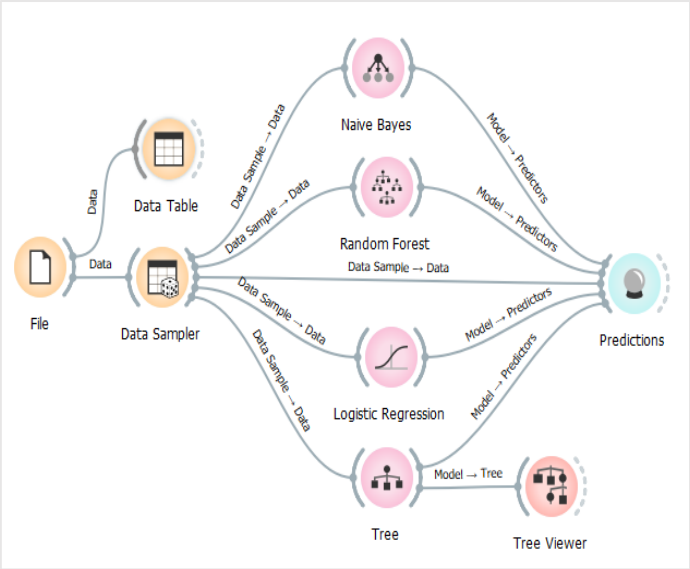
- Cross-validation has been implemented in the Orange Test & Score widget, which is used for evaluating model performance, preventing overfitting and providing a better indicator of how well the model would perform on unseen data.
- Cross-validation is a widely used statistical method for measuring machine learning model performance (accuracy) and preventing overfitting in a prediction model, especially when data is limited.



- In the Data Sampler widget, Boostrap was utilised as a Random Forest for categorization. Data Sample -> Data (Train Data: 80) and Data Sample ->Test Data were used in this project study (Test Data: 20).
- After sending the supervised machine learning models to the Test & Score widget with the Train and Test samples, the results of the models may be viewed in the table within the Test & Score widget. However, Test on test data and Test on train data must be clicked before watching the evaluation outcome because there are additional possibilities for evaluation. Then, in Edit Links, must link the relevant Data Sampler with the appropriate Test and Score.

5. Prediction





- There are two types of predictions that can be made: separating Training Data and Testing Data widgets or utilising the Data Sampler widget. Both have the same outcome.

https://www.analyticsvidhya.com/blog/2020/11/predicting-employee-attrition-using-orange-ows-visual-programming-software/

https://towardsdatascience.com/data-science-made-easy-data-modeling-and-prediction-using-orange-f451f17061fa

http://docs.biolab.si/orange/2/widgets/rst/evaluate/predictions.html

https://phoenixnap.com/kb/handling-missing-data-in-python

https://www.analyticsvidhya.com/blog/2021/05/dealing-with-missing-values-in-python-a-complete-guide/

https://github.com/chhayac/Machine-Learning-Notebooks

https://jakevdp.github.io/PythonDataScienceHandbook/03.04-missing-values.html

https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=89096912

https://www.meta.org/papers/using-different-machine-learning-models-to/34542195