

### 7.2.1 Problem Statement:

Build a Knowledge Flow model using Weka tool.

### 7.2.2 Description:

#### About Dataset used

The weather data is a small [open data set](#) with only 14 examples.

In RapidMiner it is named [Golf Dataset](#), whereas Weka has two data set: weather.nominal.arff and weather.numeric.arff

The dataset contains data about weather conditions are suitable for playing a game of [golf](#). the original dataset that only has 5 variables:

- 1.outlook
- 2.temperature
- 3.humidity
- 4.windy
- 5.play

#### About Arff:

An ARFF (Attribute-Relation File Format) file is an ASCII text file that describes a list of instances sharing a set of attributes.

ARFF files have two distinct sections. The first section is the **Header** information, which is followed the **Data** information.

The **Header** of the ARFF file contains the name of the relation, a list of the attributes (the columns in the data), and their types.

Lines that begin with a  are comments. The **@RELATION**, **@ATTRIBUTE** and **@DATA** declarations are case insensitive.

#### Knowledge Flow

The KnowledgeFlow presents a "data-flow" inspired interface to Weka. The user can select Weka components from a tool bar, place them on a layout canvas and connect them together in order to form a "knowledge flow" for processing and analyzing data. At present, all of Weka's classifiers and filters are available in the KnowledgeFlow along with some extra tools.

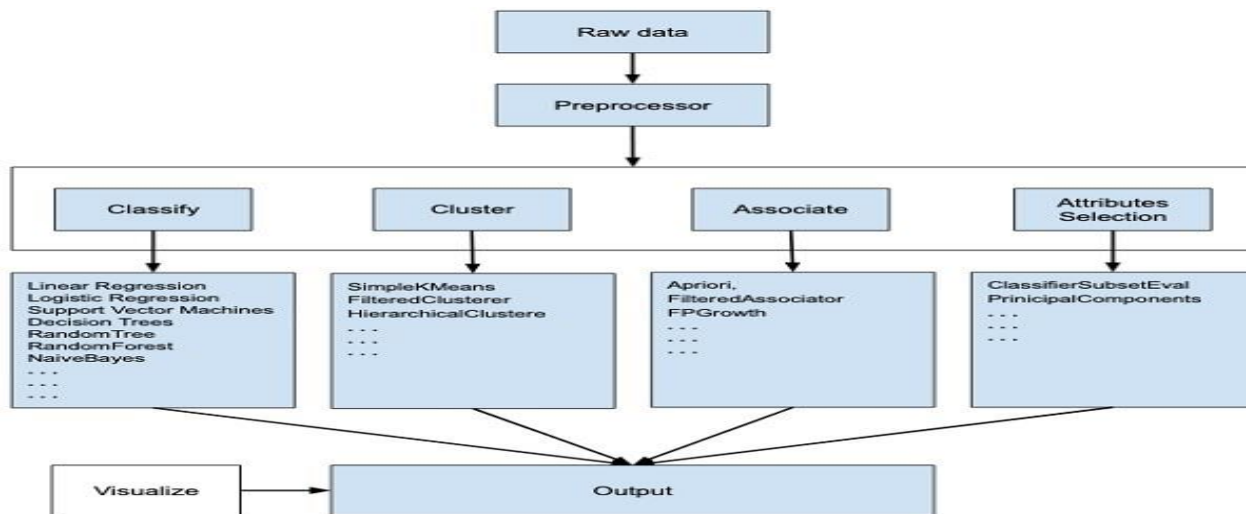
The KnowledgeFlow can handle data either incrementally or in batches (the Explorer handles batch data only). Of course learning from data incrementally requires a classifier that can be updated on an instance by instance basis. Currently in Weka there are five classifiers that can handle data incrementally: NaiveBayesUpdateable, IB1, IBk, LWR (locally weighted regression).

#### Features of the Knowledge Flow:

- intuitive data flow style layout
- process data in batches or incrementally
- process multiple batches or streams in parallel! (each separate flow executes in its own thread)
- chain filters together
- view models produced by classifiers for each fold in a cross validation
- visualize performance of incremental classifiers during processing (scrolling plots of classification accuracy, RMS error, predictions etc)

### About WEKA Software:

WEKA - an opensource software provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools so that you can develop machine learning techniques and apply them to real-world data mining problems. What WEKA offers is summarized in the following diagram –



If you observe the beginning of the flow of the image, you will understand that there are many stages in dealing with Big Data to make it suitable for machine learning –

First, you will start with the raw data collected from the field. This data may contain several null values and irrelevant fields. You use the data preprocessing tools provided in WEKA to cleanse the data.

Then, you would save the preprocessed data in your local storage for applying ML algorithms.

Next, depending on the kind of ML model that you are trying to develop you would select one of the options such as **Classify**, **Cluster**, or **Associate**. The **Attributes Selection** allows the automatic selection of features to create a reduced dataset.

Note that under each category, WEKA provides the implementation of several algorithms. You would select an algorithm of your choice, set the desired parameters and run it on the dataset.

Then, WEKA would give you the statistical output of the model processing. It provides you a visualization tool to inspect the data.

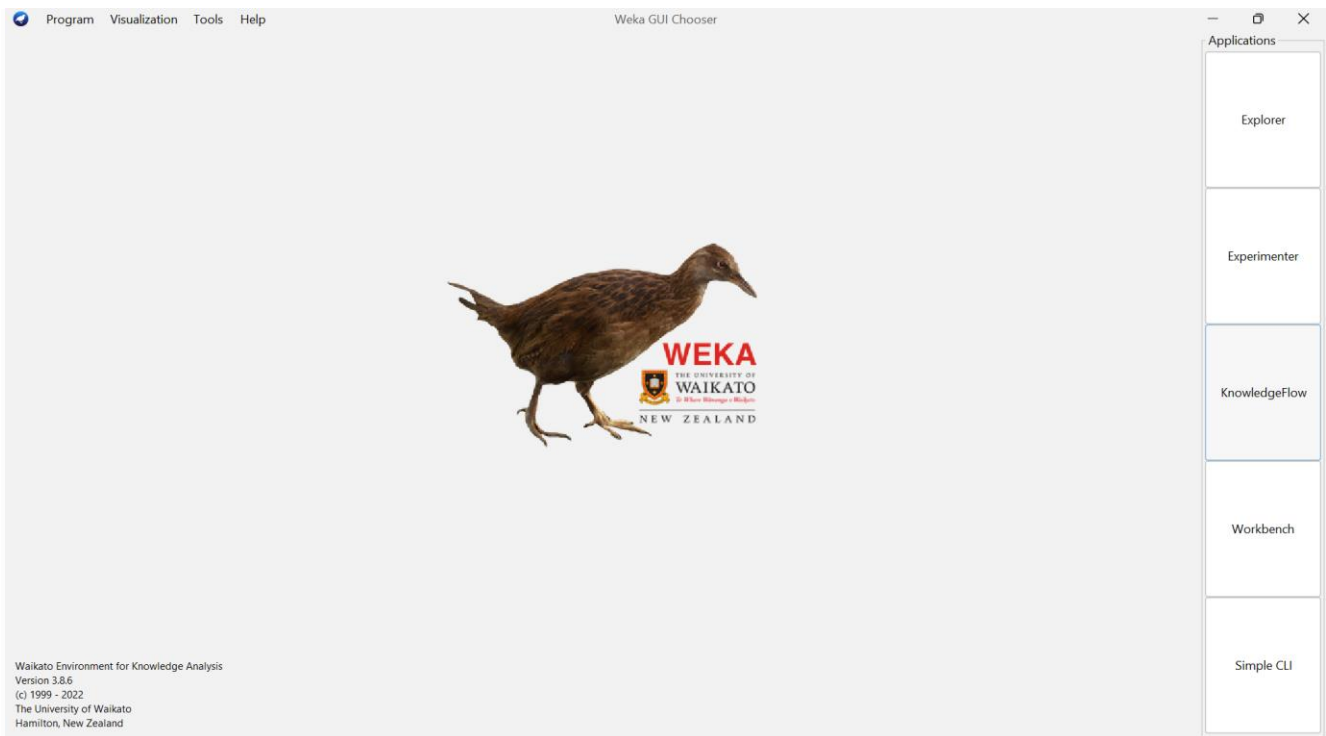
The various models can be applied on the same dataset. You can then compare the outputs of different models and select the best that meets your purpose.

Thus, the use of WEKA results in a quicker development of machine learning models on the whole.

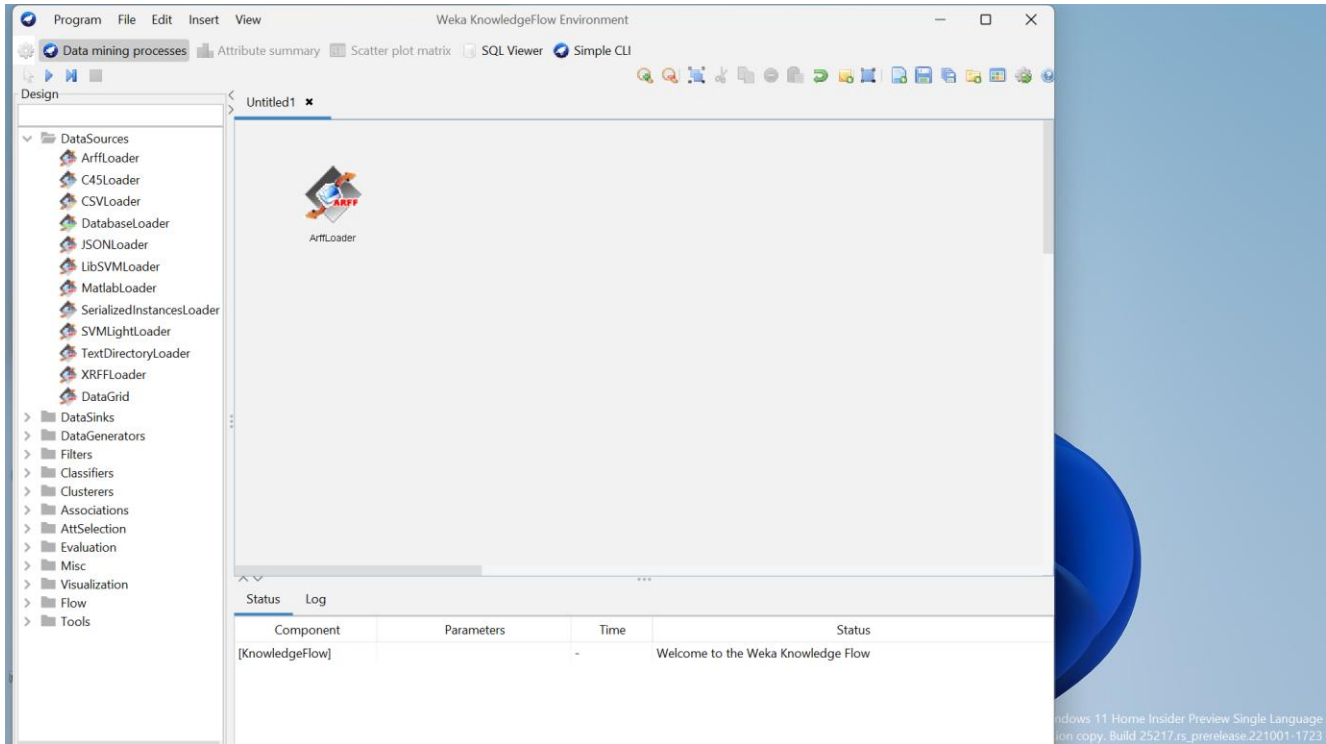
Now that we have seen what WEKA is and what it does, in the next chapter let us learn how to install WEKA on your local computer.

### **7.2.3 Steps to Build a Knowledge Flow model:**

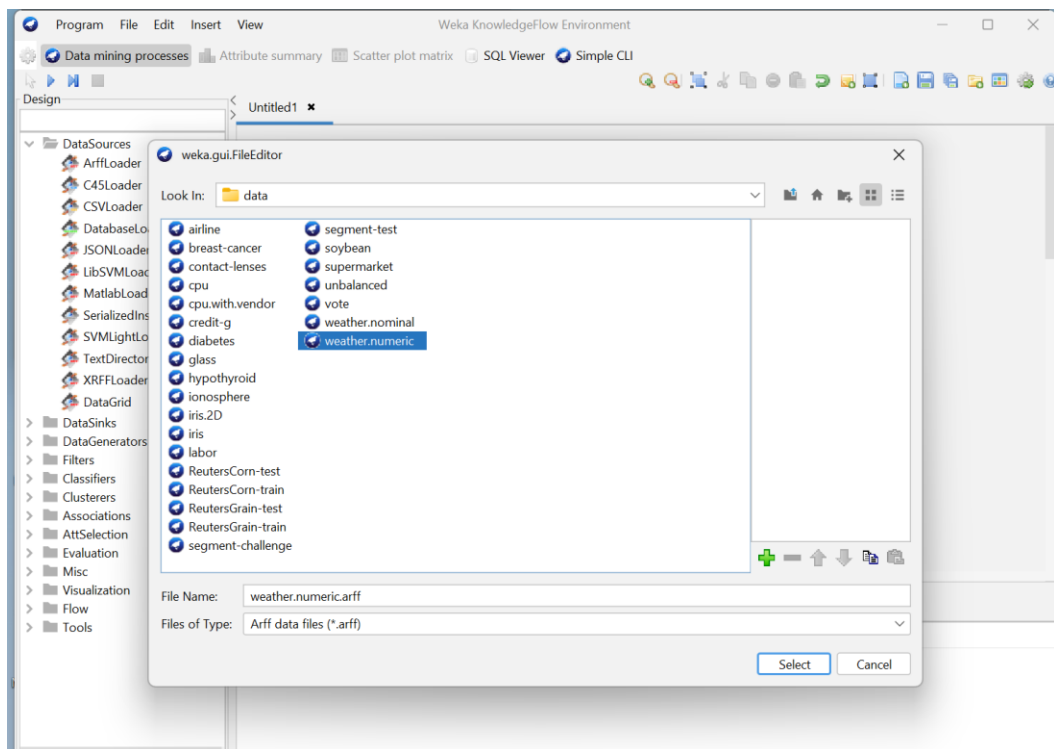
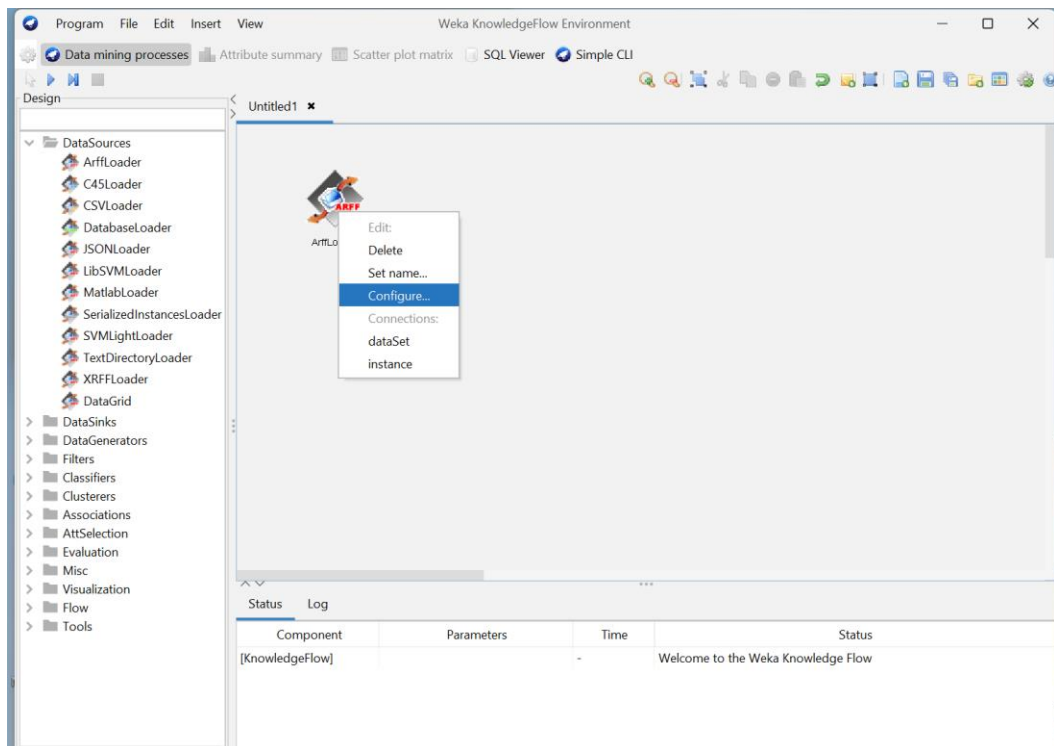
1. To Build the knowledge Flow model first open the weka application and select to open knowledge flow

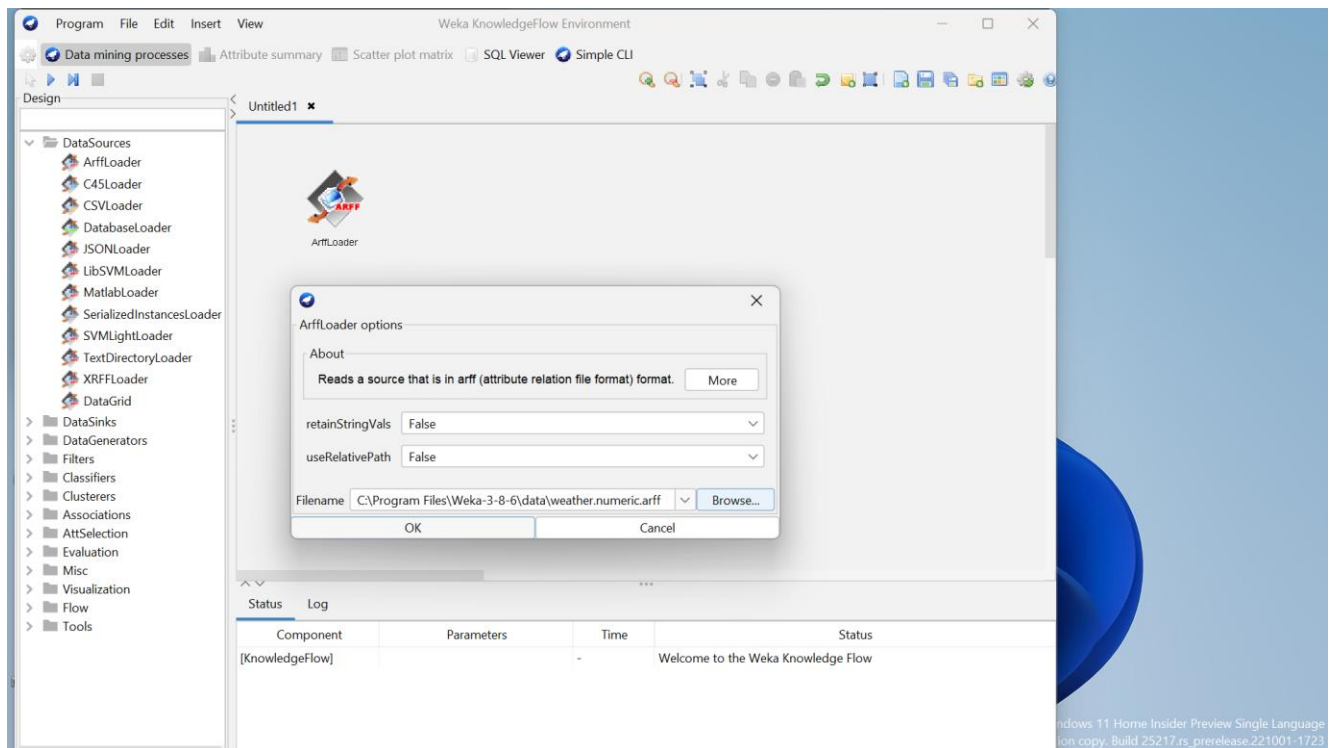


2. Now in the DataSources select ArffLoader then add that in the white screen which is appearing at the right side.

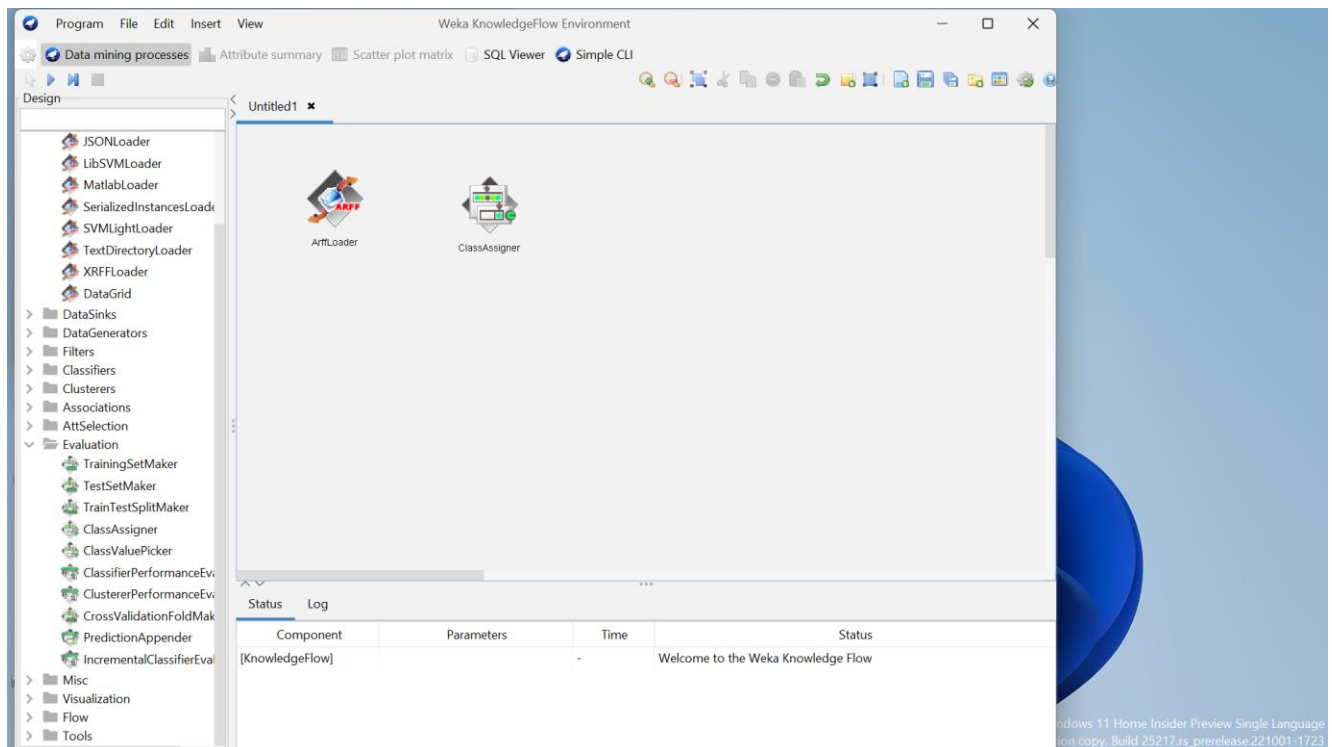


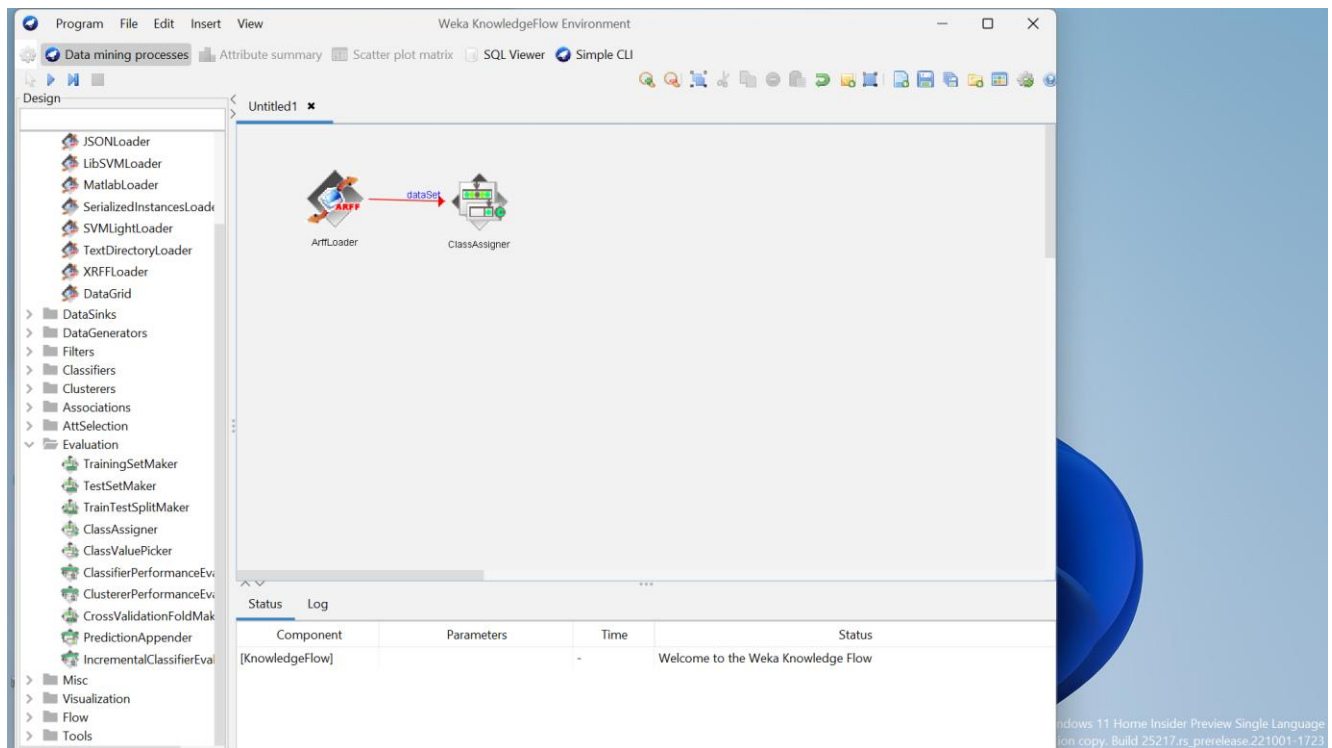
3. Right click on the Arffloader it will display some options from that select Configure to open Arff file from the browser and click on OK.



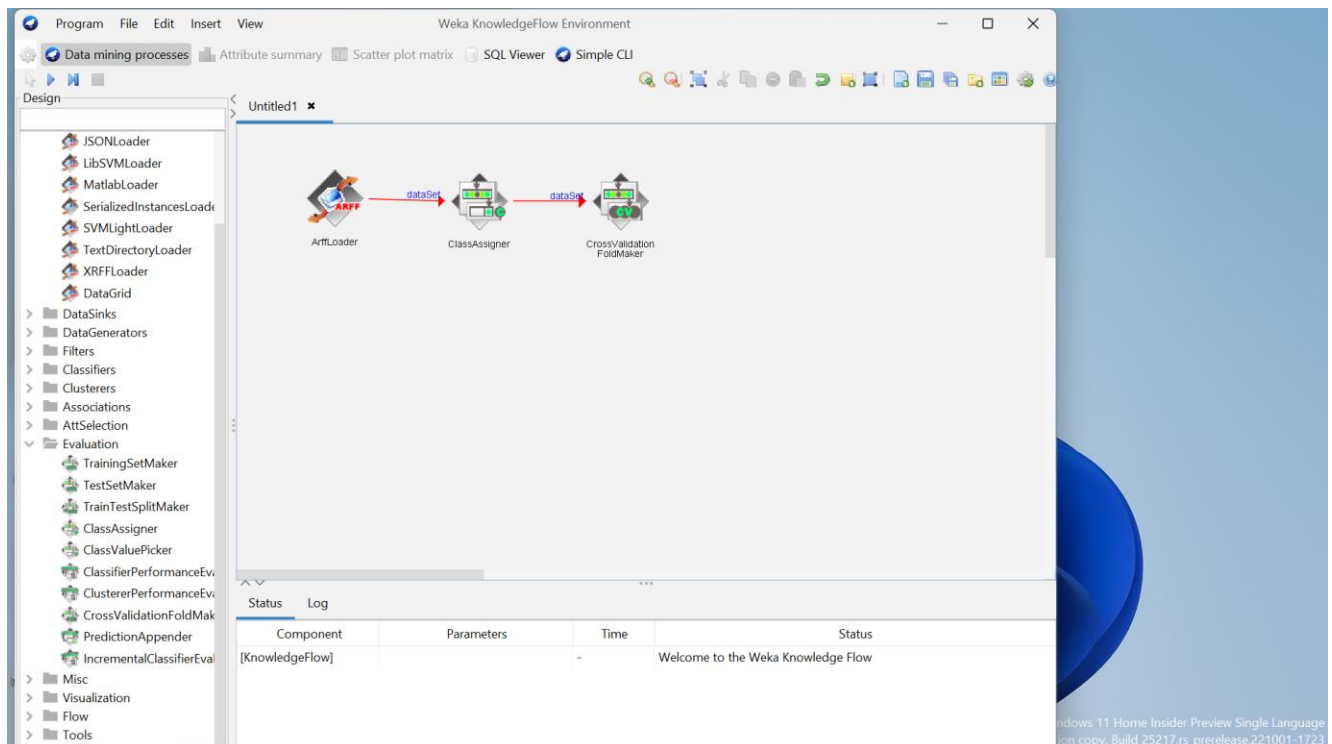


4. Now under Evaluation select Class Assigner to assign the class to the Arff file, again on the ArffLoader right click on that and select data set and drag it towards the ClassAssigner





5. Again under Evaluation select Cross Validation Flod Maker which is used to split the data into training set and test set.

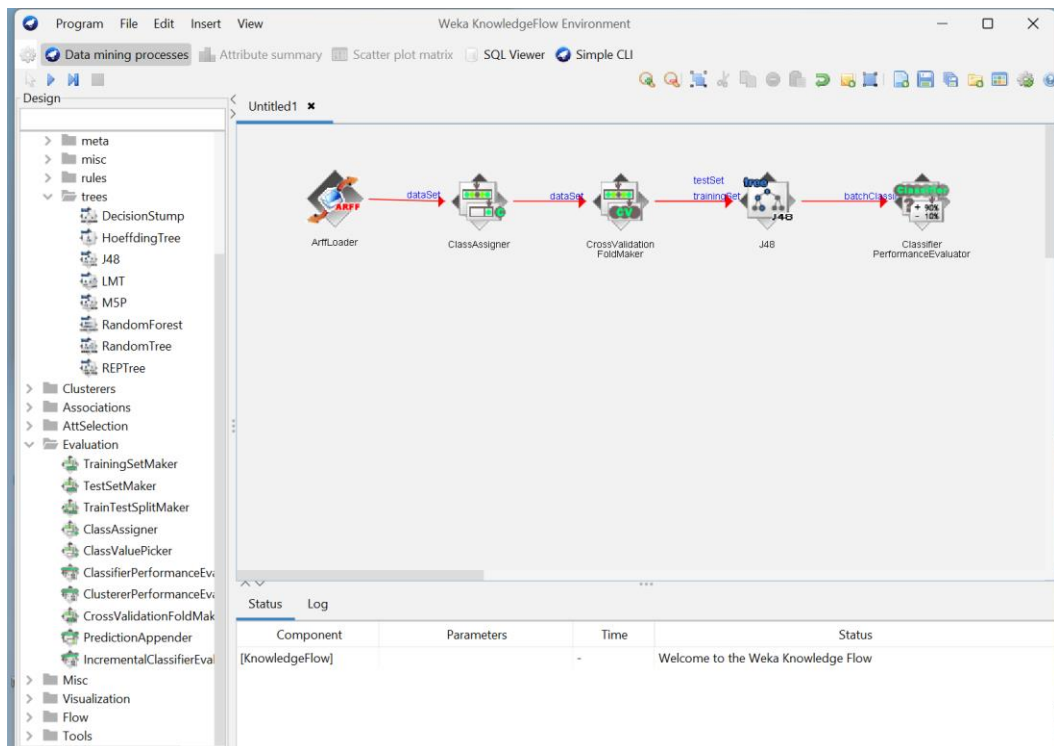
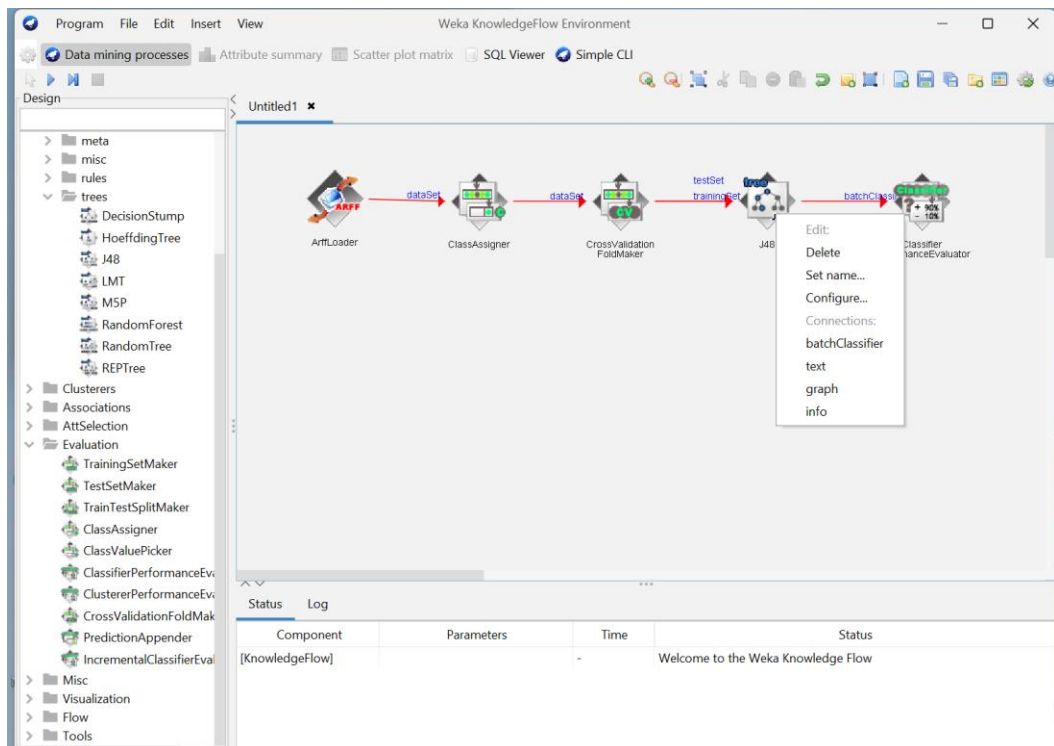


6. Now under Classifier select any classifier method like J48 which is used to classify the data set, Right click on cross validation flod maker and select both training set and test set drag them towards J48.

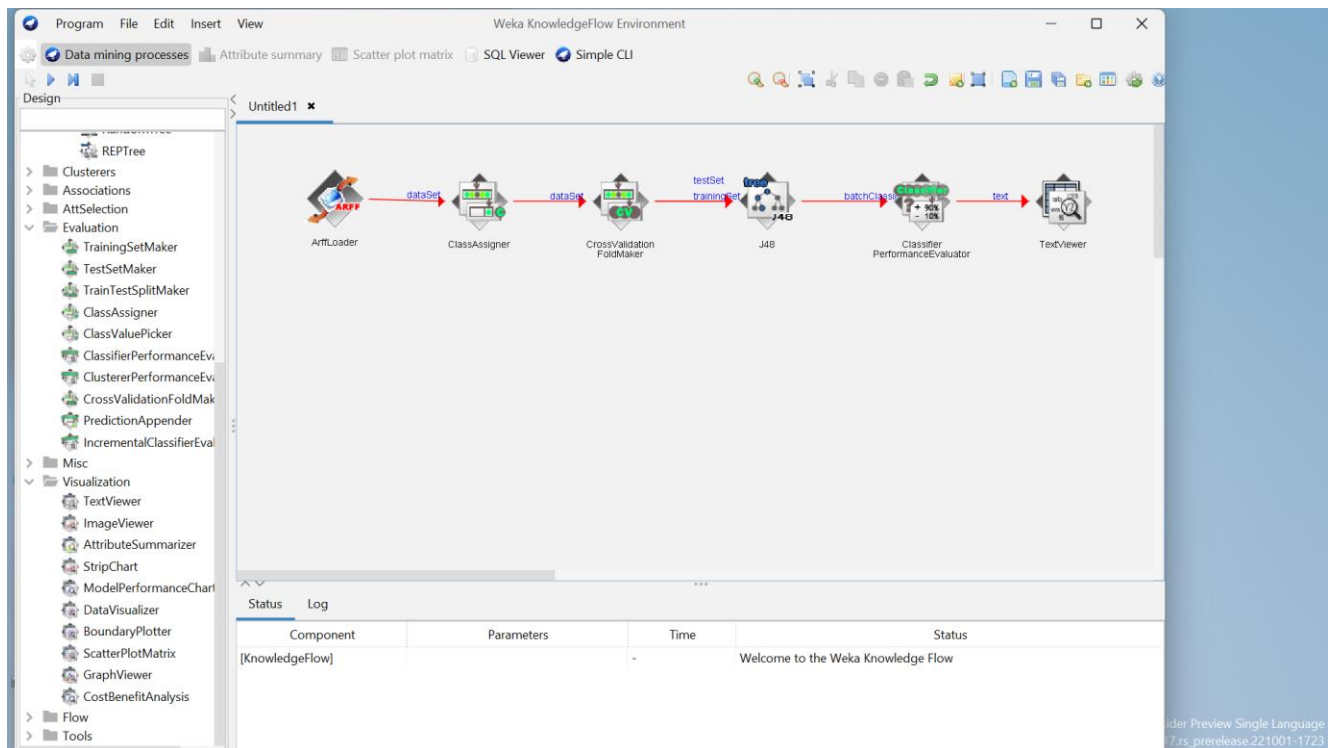
And under Evaluation select Classifier Performance Evaluator

Next right click on J48 and select batch Classifier and drag it towards Classifier performance Evaluator.

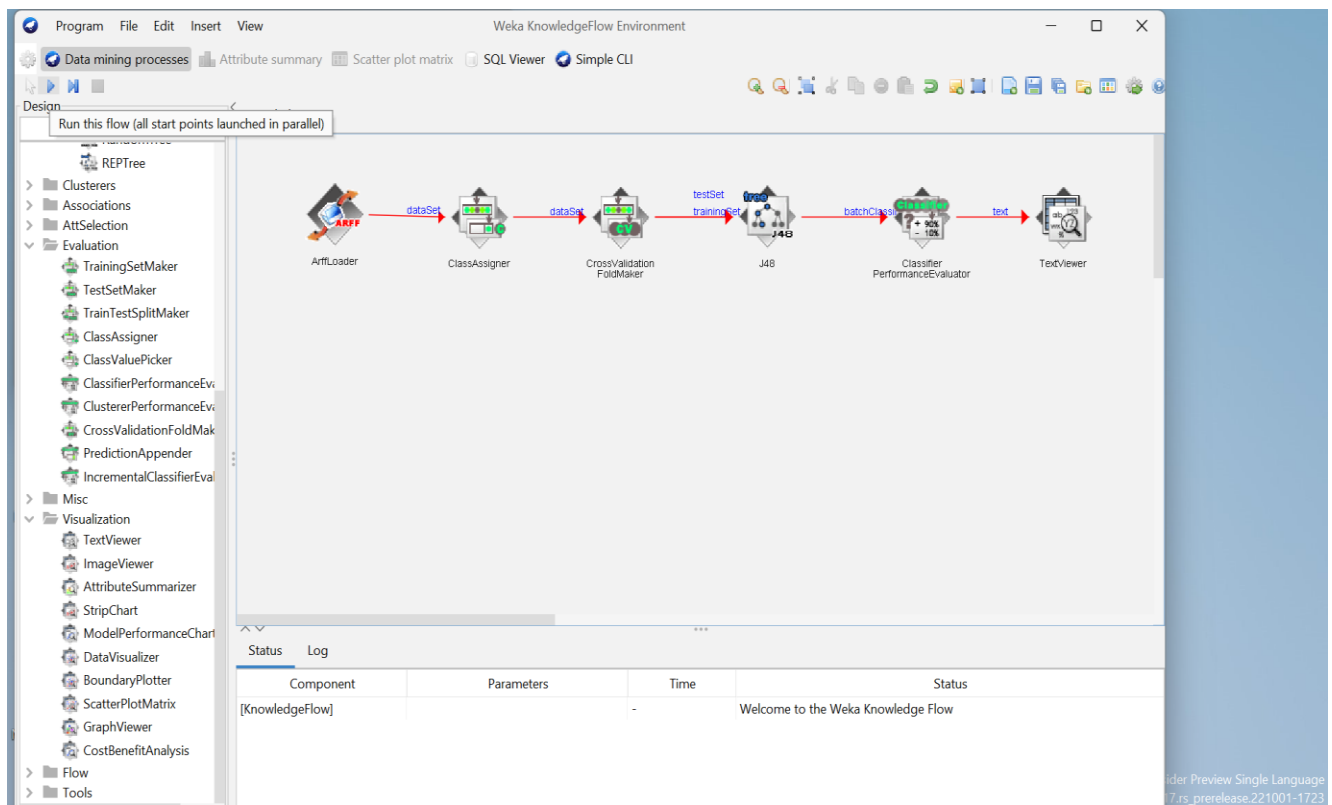




7. Now under Visualization select Text Viewer which is used to display the result, right click on classifier performance evaluator select text and drag it towards Text Viewer.

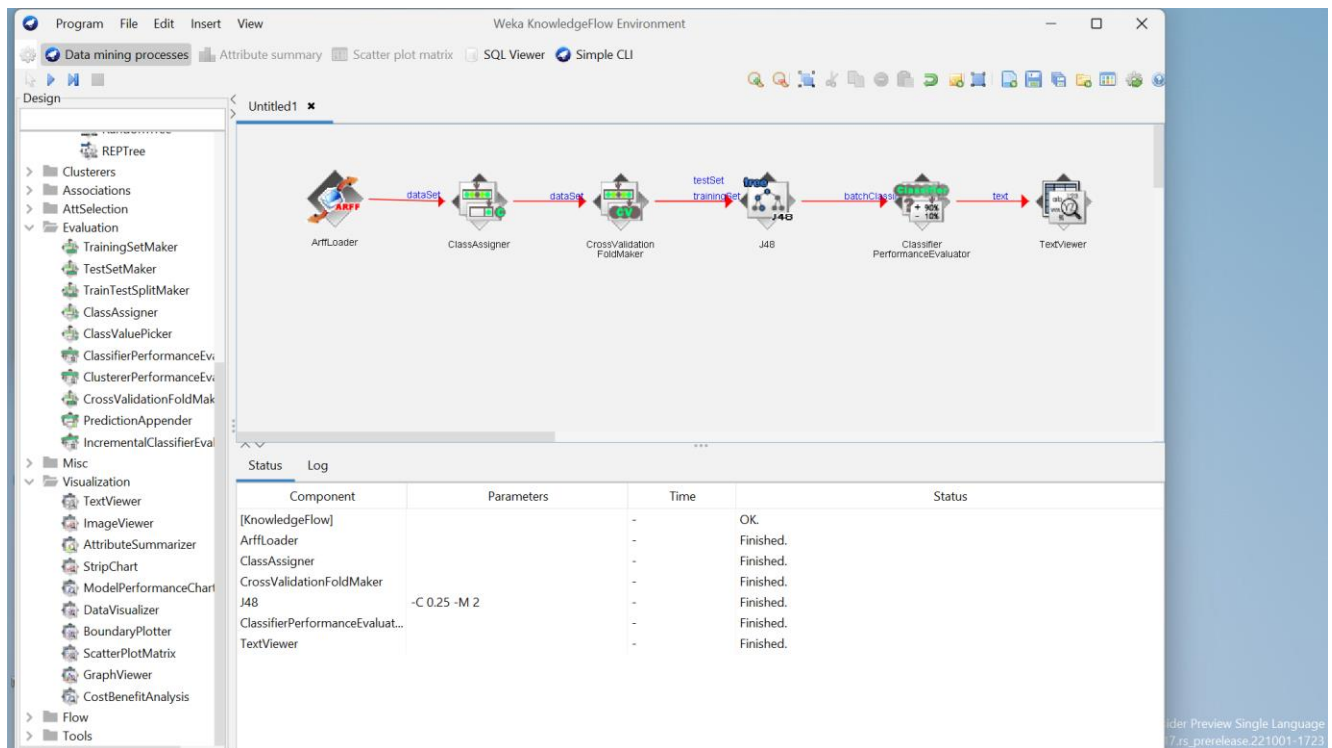


8. At the top left corner there is a button to run the flow click on run.

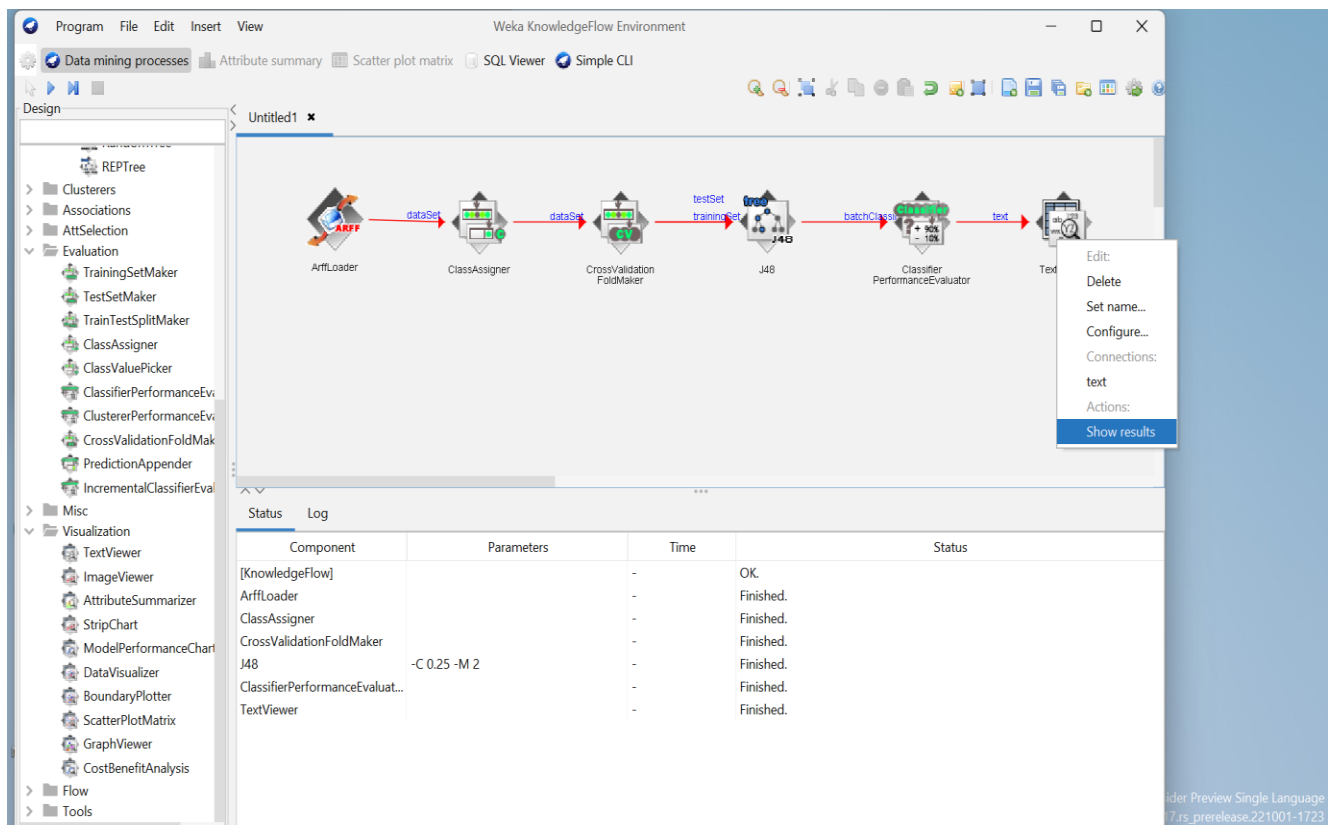


9. After that under status it will show status of different components that are completely finished or not.

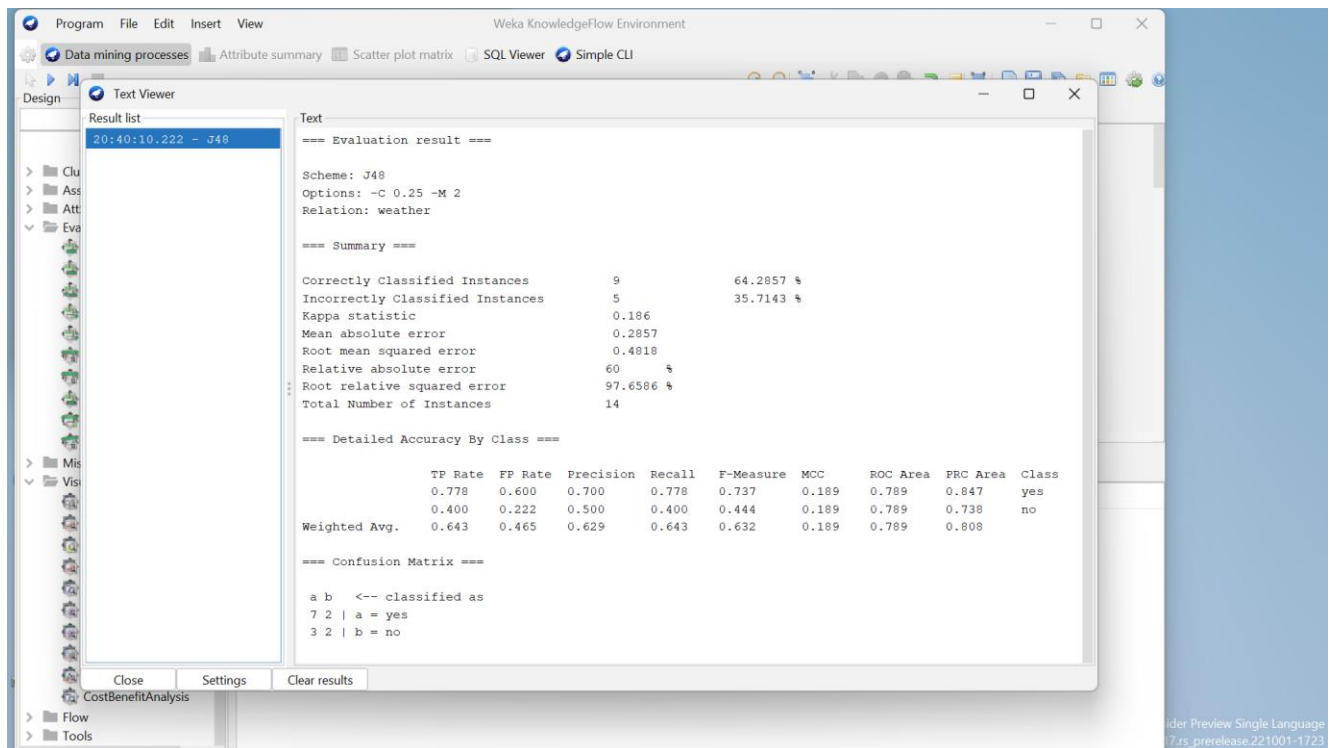




10. To see the result right click on Text Viewer under that select Show results.



11. Now the result is shown in the form of text.



## 7.2.4 Results and Discussion:

Therefor building a Knowledge Flow model using Weka tool is completed successfully.