Learning Analytics Tool

Daevesh Singh

Department of Educational Technology

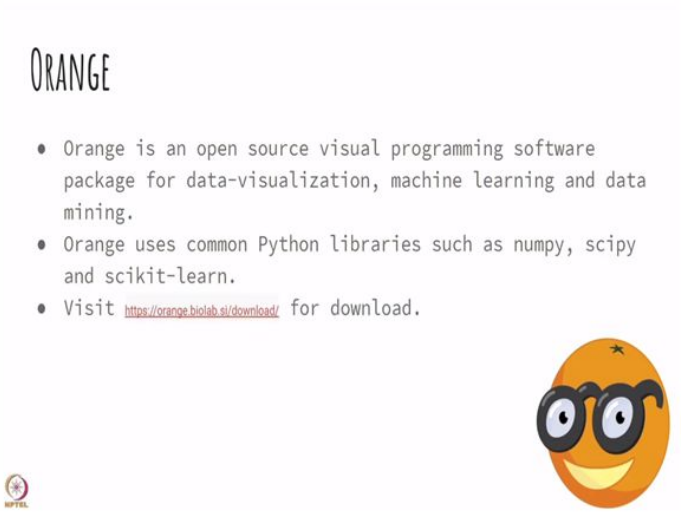
Indian Institute of Technology, Bombay

Orange – Data Mining Fruitful and Fun

(Refer Slide Time: 00:16)

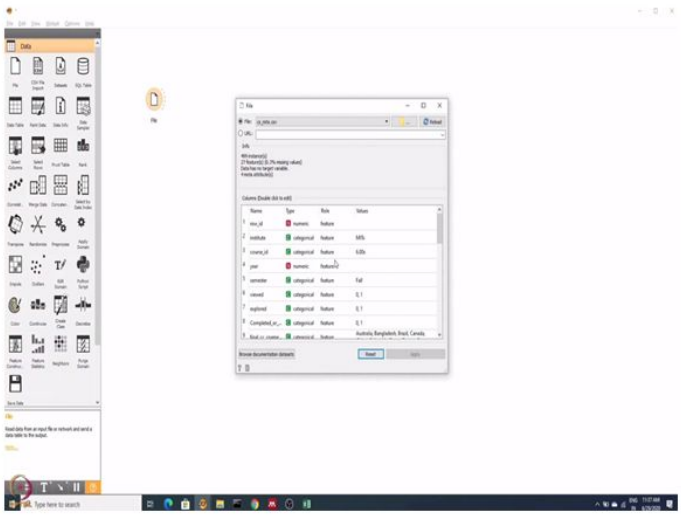
Hello Everyone. I am Daevesh, a Research Scholar at the Educational Department of IIT Bombay. Today, we will talk about the tool called Orange.

(Refer Slide Time: 00:27)



Orange is an open-source Visual Programming software package for Data Visualization, Machine Learning and Data Mining. Orange uses common Python libraries such as NumPy, SciPy, and Scikit-learn. You can download orange by visiting this particular link.

(Refer Slide Time: 00:47)



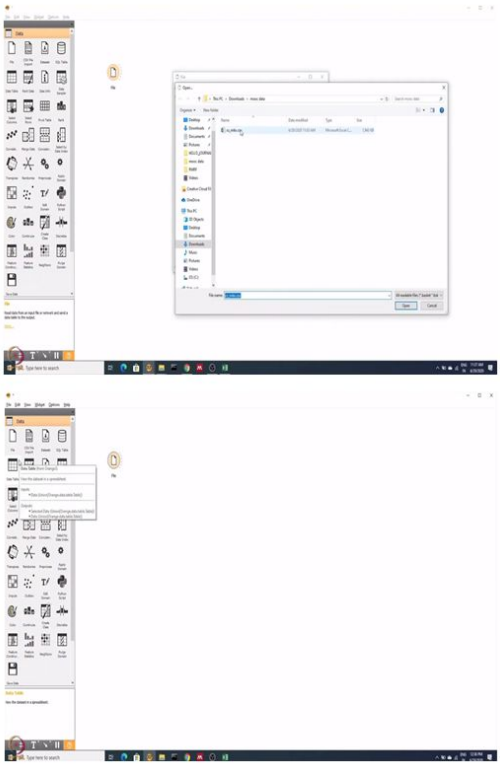
This is the welcome screen of Orange. The white blank space here is called the canvas. These items on the left are called widgets, these widgets are actually the computational units of Orange. They help in reading the data, processing it and visualizing it and building predictive models. The widgets have input output channels. You just have drag and drop it over the canvas. So this

particular widget has only the output channel. A widget, in particular, can have input, output or both channels.

Now, these widgets have been mainly classified into 5 categories. The first category is Data. The second category is Visualize. It also has a couple of widgets. Third, being Model, fourth is Evaluate and fifth is Unsupervised. These widgets actually communicate with each other with the help of these Input-Output channels.

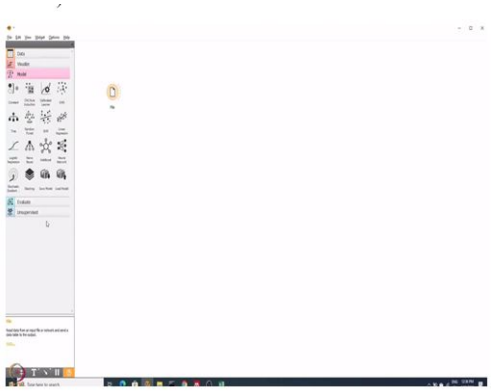
Now, if we double click the file icon, a dialogue box will appear. This gives us a choice to use our own dataset or we can also use the dataset that is provided by Orange. This also provides us with a description of the various features that are involved. So, these are the names of the features, these are the data types, it can be numeric, categorical and the roll. So, the roll is basically categorized into 3 classes, either it is a feature or it is metadata or it is the value to be predicted. And these are some of the values that these variables take.

(Refer Slide Time: 02:36)



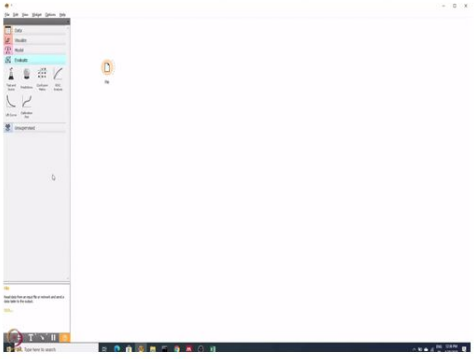
Now let us briefly describe the various categories. So, the first category of widgets is Data. Now, we are choosing this particular file. So, now that particular file has been uploaded. Widgets in this category allow importing the files and pre-processing of the data. The second category is Visualize. This has widgets that give a pictorial representation of the data.

(Refer Slide Time: 03:08)



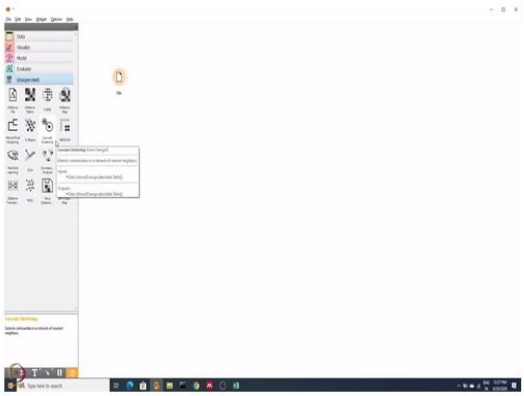
The third is the Model. In the model, the widgets provide us with access to various machine learning models that we can use for regression and classification task. We will touch upon some of these models in the course.

(Refer Slide Time: 03:23)



The next category is Evaluate category. So it helps us in the assessment of the models. That is it gives us an idea about how well our model is performing. This is demonstrated through the use of various metrics that we will talk about in the course in the coming weeks.

(Refer Slide Time: 03:44)



The last category is the widgets Unsupervised. So these provide us unsupervised models along with some widgets that help in feature transformation.