Select Columns Transform

Updated: July 18, 2015

Creates a transformation that selects the same subset of columns as in the given dataset

Category: Data Transformation / Manipulation (https://msdn.microsoft.com/en-us/library/azure/dn905863.aspx)

You can use the **Select Columns Transform** module to create a transformation that will ensure that the same set of columns is always used in downstream operations. This can be useful if you want to apply an operation that always requires specific columns and you make sure that column selections do not change, as this might break the experiment or change the results.

Note that you can also use Project Columns (https://msdn.microsoft.com/en-us/library/azure/dn905883.aspx) to choose a subset of columns to use in a downstream module. However, Project Columns (https://msdn.microsoft.com/en-us/library/azure/dn905883.aspx) is not always the easiest to configure if there are many columns. Moreover, there are times when the selection of columns in an input dataset might change depending on feature selection or other operations.

For example, suppose you use Filter Based Feature Selection (https://msdn.microsoft.com/en-us/library/azure/dn913071.aspx) with a dataset to automatically find the best features in a dataset, and then use the dataset created by Filter Based Feature Selection (https://msdn.microsoft.com/en-us/library/azure/dn913071.aspx) as an input to Train Model (https://msdn.microsoft.com/en-us/library/azure/dn906044.aspx).

Because Filter Based Feature Selection (https://msdn.microsoft.com/en-us/library/azure/dn913071.aspx) evaluates the feature importance based on the values in the column, it is impossible to know beforehand which columns to use when scoring. Moreover, if you apply Filter Based Feature Selection (https://msdn.microsoft.com/en-us/library/azure/dn913071.aspx) to the scoring dataset, it might choose a different set of columns, which would cause the scoring operation to fail.

In this scenario, you can use the **Select Columns Transform** module to generate a transformation (which you can save as an ITransform interface (https://msdn.microsoft.com/en-us/library/azure/dn905982.aspx)) to ensure that the same set of columns is used for scoring that is used for training.

Expected Inputs

Name	Type	Description
Dataset with desired columns	Data Table (https://msdn.microsoft.com/en-us/library/azure/dn905851.aspx)	Dataset containing desired set of columns

Outputs

Name	Type	Description
Columns selection transformation	ITransform interface (https://msdn.microsoft.com/en-us/library/azure/dn905982.aspx)	Transformation that selects the same subset of columns as in the given dataset.

Exceptions

Exception	Description
Error 0003 (https://msdn.microsoft.com/en-us/library/azure/dn906003.aspx)	Exception occurs if one or more of inputs are null or empty.

See Also

Data Transformation / Manipulation (https://msdn.microsoft.com/en-us/library/azure/dn905863.aspx)
Project Columns (https://msdn.microsoft.com/en-us/library/azure/dn905883.aspx)
A-Z List of Machine Learning Studio Modules (https://msdn.microsoft.com/en-us/library/azure/dn906033.aspx)