Predictive Model Markup Language (PMML) FAQs

Smarten

Document Information	
Document ID	Smarten-PMML-FAQs
Document Version	1.0
Product Version	5.3
Date	10-November-2022
Recipient	NA
Author	EMTPL

Smarten

Contents

1   What is a PMML file?

PMML stands for "Predictive Model Markup Language." It is an XML-based model interchange format designed to provide a way through which predictive solutions can be transferred from one application to another.

Smarten supports PMML models in its workflow to enable interoperability of the predictive models created in other platforms (such as R and Python) and utilize those PMML models to make predictions on the operational data.

You can read more about PMML on Wikipedia here: link

2   Why PMML? What are the benefits?

It is the de facto standard to represent predictive solutions supported by many platforms. It supports other aspects, such as data dictionary, data transformations, and missing value treatment, and that makes PMML as an *"all in one"* format that supports interoperability of models among different tools and applications without a need for custom coding.

PMML also supports transparency; thus, models are no more a "black box." Anyone can actually *read* the PMML model by simply opening the PMML file!

3   How do you create a PMML file?

A PMML file contains a *pretrained model*. Thus, before generating a PMML file, you will need to train your own predictive model, such as "Simple Linear Regression," with your training data on your preferred platform. Since various platforms support the PMML as an export option, you will need to refer to the steps to "Export as PMML" in respective platforms.

Here are some links on how to generate PMML from popular platforms:

- KNIME
- IBM SPSS
- Python
- R

4   What is the PMML version? Which versions are supported on Smarten?

PMML has been around for a long time, and there have been various improvements on the file structure. Smarten supports all PMML versions 3.0 and above. Versions before 3.0 are not supported, and we would advise you not to use them, as those were before year 2004!

5   What is the difference between the SmartenInsight model vs. the PMML model?

The models generated in SmartenInsight are the natively supported models on Smarten, and they are highly flexible. They can be retrained on different sets of parameters, and all functionality that the Assistive Predictive Modeling offers is accessible *(e.g., model summary, interpretation, and key influencer analytics)*. Smarten provides complete workflow from creating a model, auto recommendation, training the model, testing and validating the model, changing the model, and even setting model parameters, generating, for example, simple natural language interpretation of the model and visualizations.

The PMML is a static file that contains the pretrained model generated from any platform. You will only have access to the information available in the PMML file, and hence the functionalities on Assistive Predictive Modeling will be limited. PMML models can't be retrained on different sets of parameters. They can only be used for predictions on the new data.

6   How do you make predictions using the PMML model using Smarten?

Using PMML models for prediction using Smarten is simple and takes just a few steps.

- Import valid PMML file.
- Model will be loaded in Smarten and show important elements, such as Interpretation, Model Summary, and Model Information.

> Note:
> Smarten will be able to show such information for the regression, classification, and clustering types of techniques. For models created using other techniques, Smarten may not be able to offer such information, but you will still be able to make predictions.

- You can use a single record instance of a test record and predict results using the **Apply** option in Smarten.
- You can use multiple records of test records and predict results using the **Mass Apply** option in Smarten. These test records can be uploaded using an XLS or a CSV file or from the Smarten Dataset.
- Web services, such as REST-API, can also be used from third-party applications to predict results.

7   Can I retrain the PMML model with new data using Smarten?

No, you can't. Because PMML is not a model, it is a file that stores the pretrained model in a simple text that can only be used to predict on the new data. To update a PMML file with a new model and on the new data, you will have to create another PMML file.

8   What if I want to update an existing PMML file?

PMML should not be altered with manual edits. It is only supposed to make predictions on the new data, and if you try to change some portion of the file manually, it is highly likely that the file will become corrupt and not be useful.

To change anything as per your needs, you will have to regenerate the file from the platform/tool that you are using along with the changes you want to have.

9   My PMML file is not supported on Smarten. What should I do?

There are a number of potential reasons for this. Mostly, this should be related with the PMML version. Kindly check if the PMML version is higher than **3.0**. Smarten should give you the appropriate error related to the PMML version.

10 What types of PMML models are supported on Smarten?

Smarten is able to run most PMML models. Smarten will be able to show the model interpretation and model summary for the regression, classification, and clustering types of techniques.

For other models created using other techniques, Smarten may not be able to offer custom visualizations, interpretation, or a model summary.

Regardless of this aspect, you will be able to test the model (with a single instance record or multiple records) for most of the PMML files using Smarten.

11 Can I get a validation summary for my testing data from the PMML model?

To validate the predictions, the testing data should have the actual target values, which means for the validation summary if the actual values for the predictions are given in the testing data, the validation summary will be shown; otherwise, you will only get the predictions.

12 Will I have to perform data transformations for test data explicitly before making predictions?

Smarten automatically performs datatype conversion if the datatype provided in test data is different from that in the PMML file.

The following table explains how Smarten performs datatype transformation on the test data before passing into the PMML model:

Datatype in PMML file	Testing data / user input	Converted value after Smarten data autotransformation	Explanation
Numeric	"ABC"	NULL	String value such as "ABC" can't be interpreted as a number.
Numeric	"12"	12	String value that contains a number is convertible.
Integer	12.345	12	A decimal value will be converted into an integer. Possible loss of precision.
Decimal	12	12.0	Given integer converted as a decimal number. No loss of precision.
String	12	"12"	A number converted into a String. Taken as a categorical value.
Number	NULL	NULL	Null will be considered as Null, and the prediction will not be made for that record unless the PMML has missing value treatment.

A PMML file often contains the required transformation rules that can be applied on the test data, such as *Standard-Scaler.* So, when you upload the data, you no longer have to worry about transforming data manually.

For an example:

Transformer in the PMML file	Testing data / user input	Automatic conversion done by Smarten PMML engine	Explanation
Standard-Scaler	-100, 50, 100	-3, 0, 3	This transformer shrinks the range usually ranging from -3 to 3.
Ordinal Encoder	"Apple," "Banana," "Potato"	0, 1, 2	This transformer converts String values into a numerical value.

13 Product and Support Information

Find more information about Smarten and its features at www.smarten.com
Support: support@smarten.com
Sales: sales@smarten.com
Feedback & Suggestions: support@smarten.com
Support & Knowledgebase Portal: support.smarten.com