**SAP Predictive Analytics**

**Time Series Analysis**

**Student Name: Nupur Ray Student Id: nxr163130**

**Date: 18/11/2016**

**Note: You must use the following conventions to name objects/systems created in this exercise.**

**Replace SM with A - for Fall semester**

**B - for Spring semester**

**C - for Summer semester**

**Replace YY with the last 2 digits of the current year.**

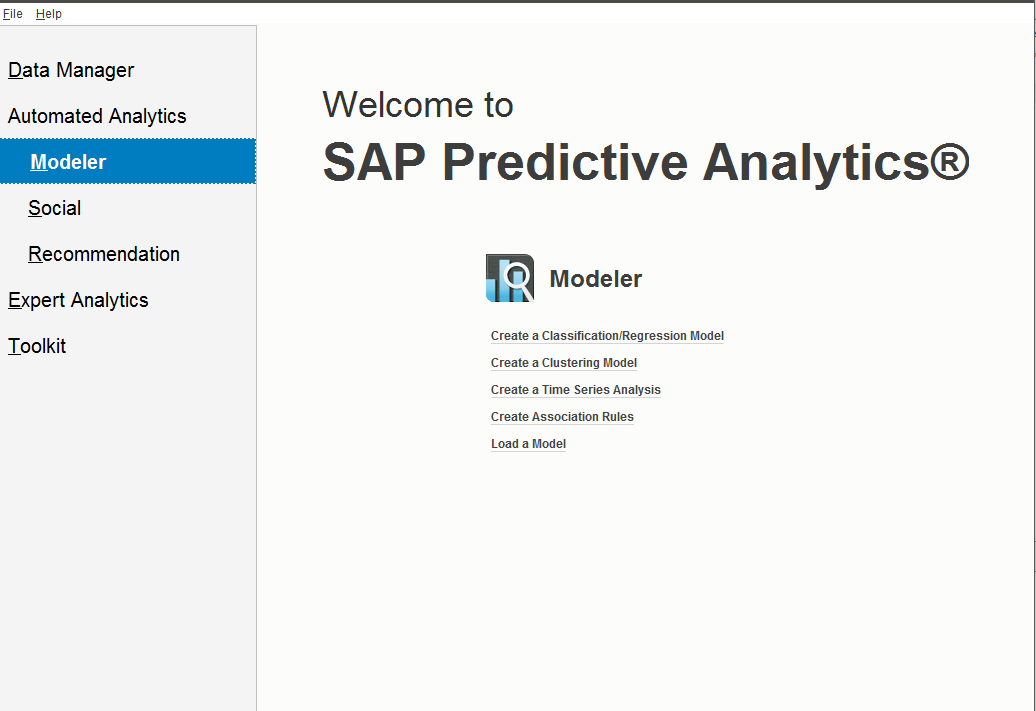
**Replace XXX with your SAP ID.**

**Objective**

Time series analysis is useful for predicting future results for a variable based on past results. In this example, we shall look at a dataset measuring cashflows for a business and we will create two different models. In the first model, we shall project results based only on the Cash and Date variables. In the second model, we shall also consider the contribution of additional variables to see as to how these affect the modelling results.

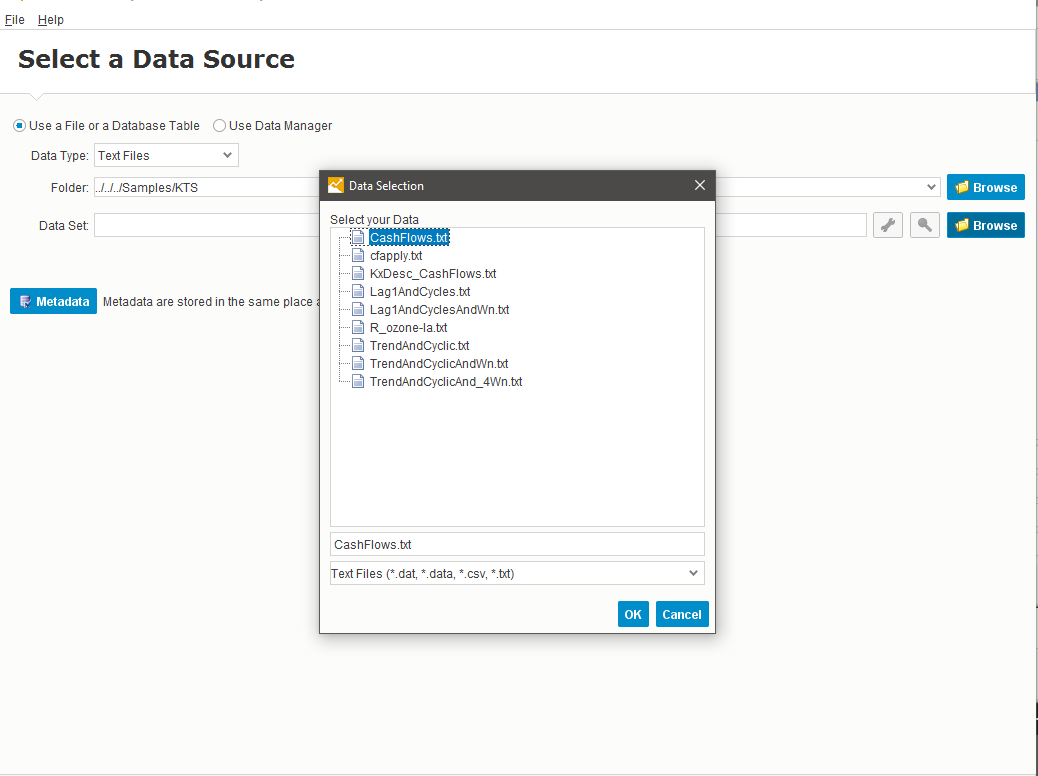
**Creating a Time Series Analysis**

Click on Create a Time Series Analysis.



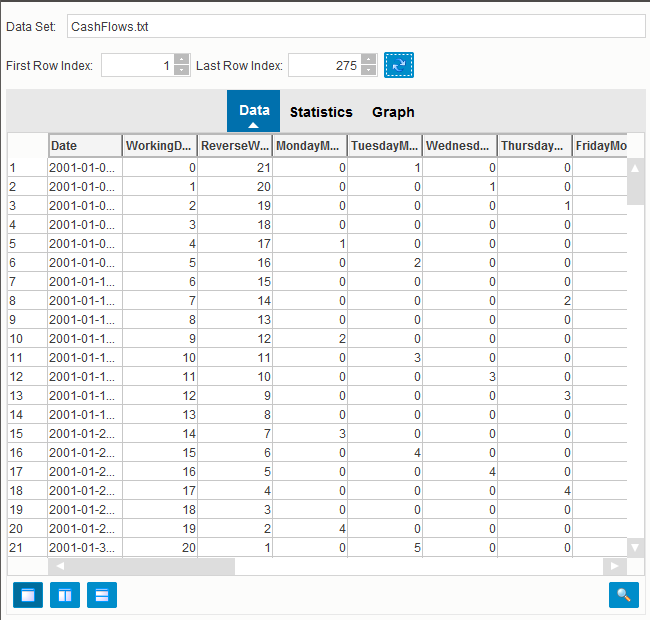
First, we need to select data source for our analysis. For this example, we shall use one of the sample files installed with InfiniteInsight.

Click Browse->Samples->KTS->**CashFlows.txt**.

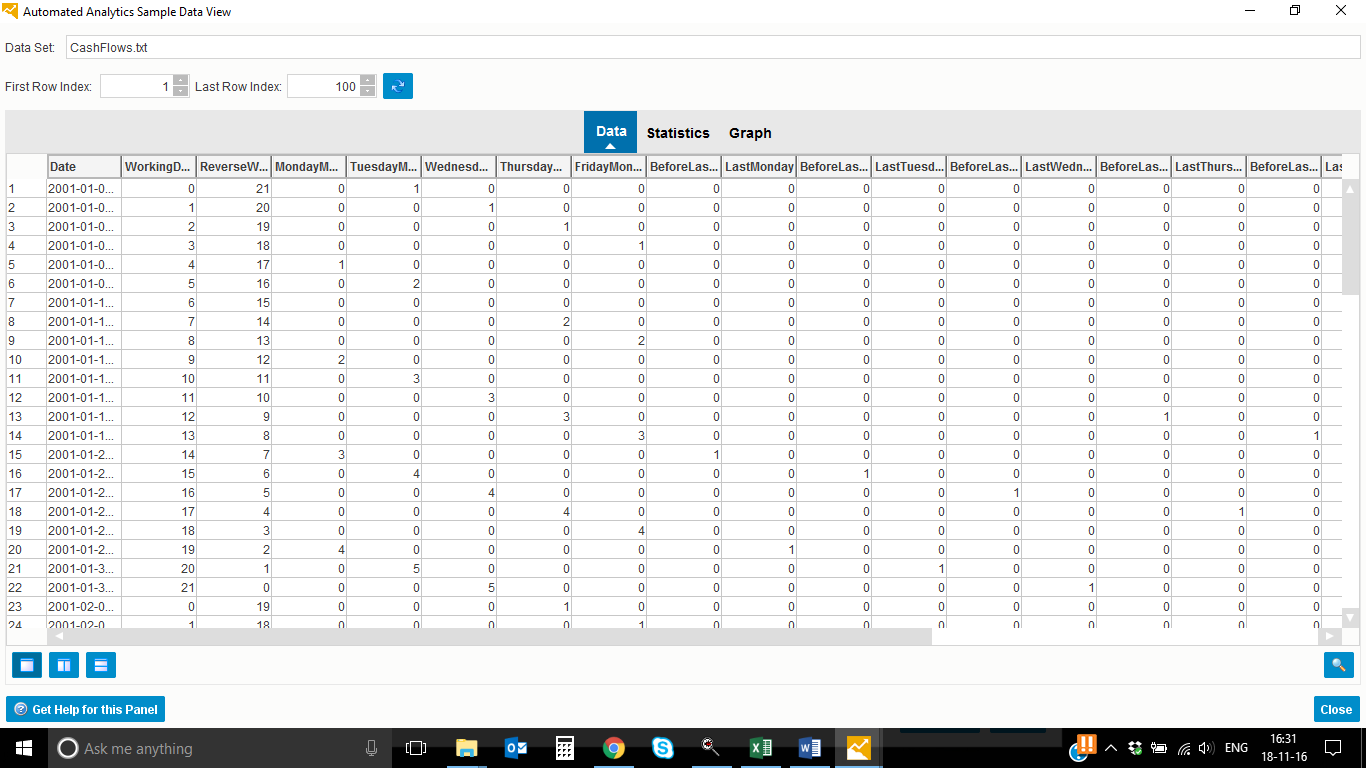


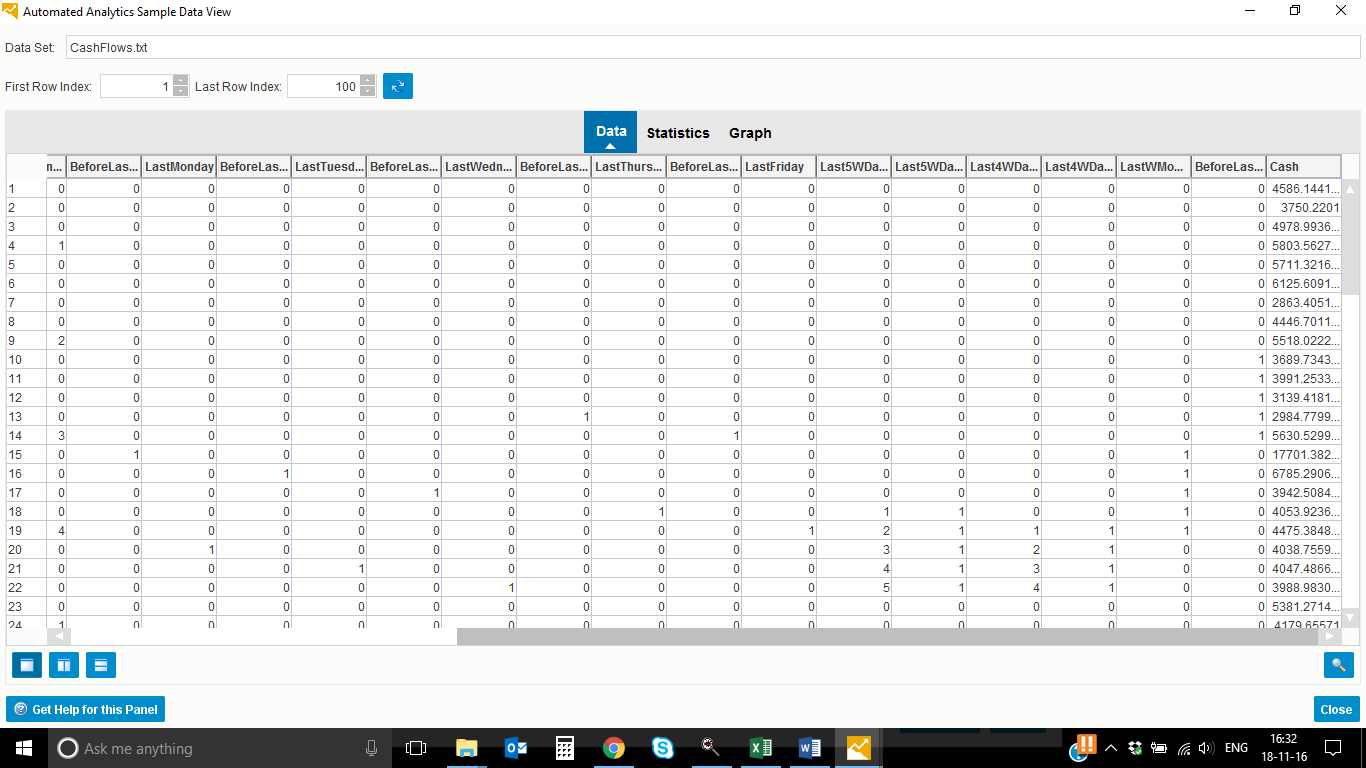
Click OK.

Next, we shall preview the columns to get a better understanding of the data source. Click  to preview the data.

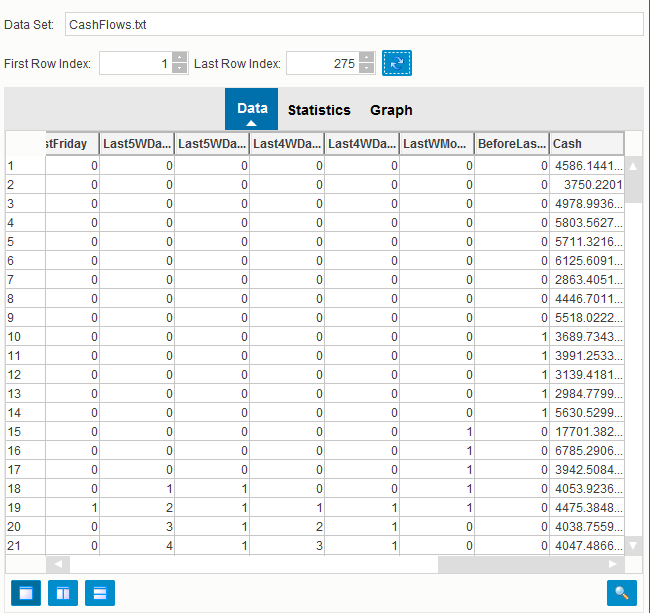


**Paste a screenshot here.**

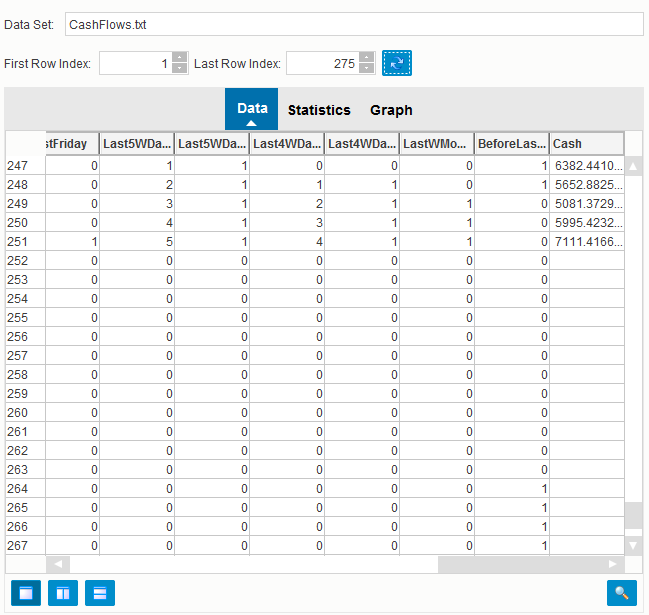




Note that records are arranged in chronological order and Cash column is at the right.



Here we can see that values in Cash column stop before that in all other columns. For our first model, we shall just consider Date and Cash variables to make predictions for the missing values in the Cash column. In our second model, we shall also include the other columns to see if they help the model to predict the missing Cash values better.

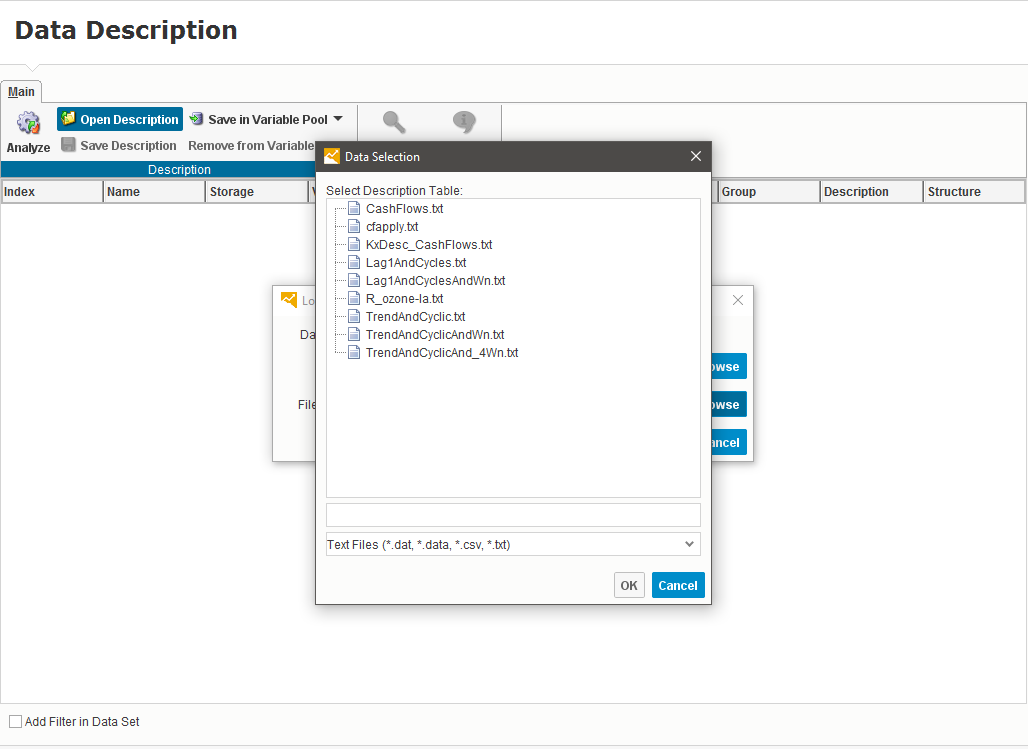


Click Close and then Click Next.

**Data Description**

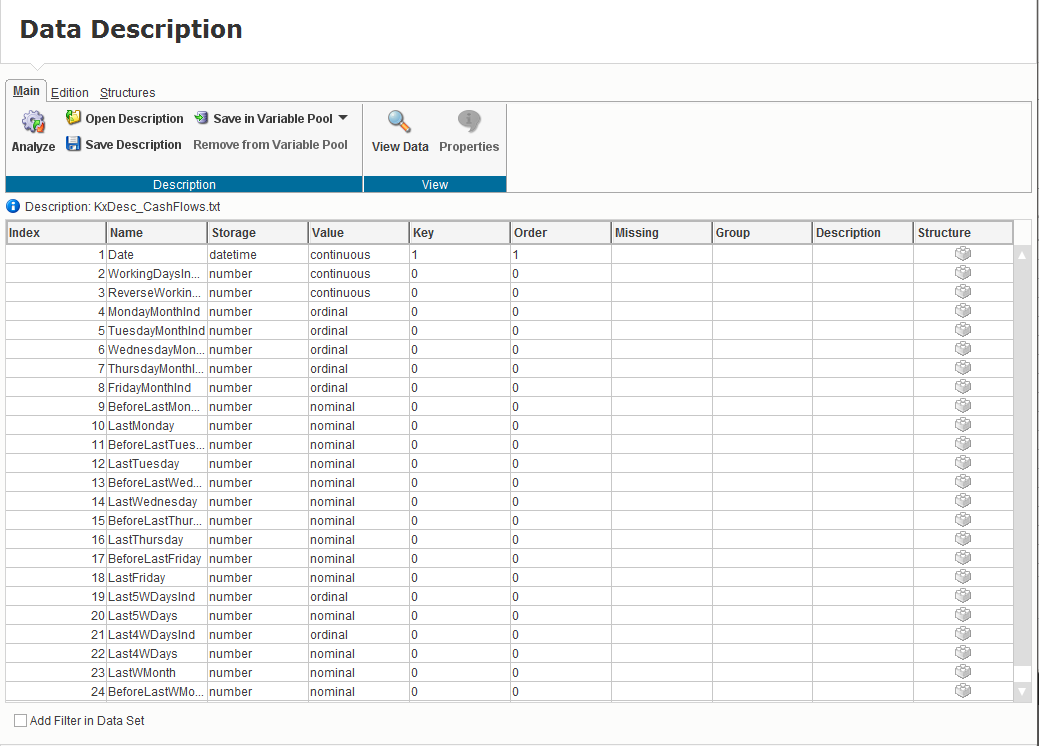
Next we need descriptions of the data to identify correctly the characteristics of the variables. We can do this by getting InfiniteInsight to analyze the data and then manually validate the results. We can also do this by loading an existing description file. For this example, we shall load a description file.

Click Open Description->Browse->KxDesc\_CashFlows.txt.

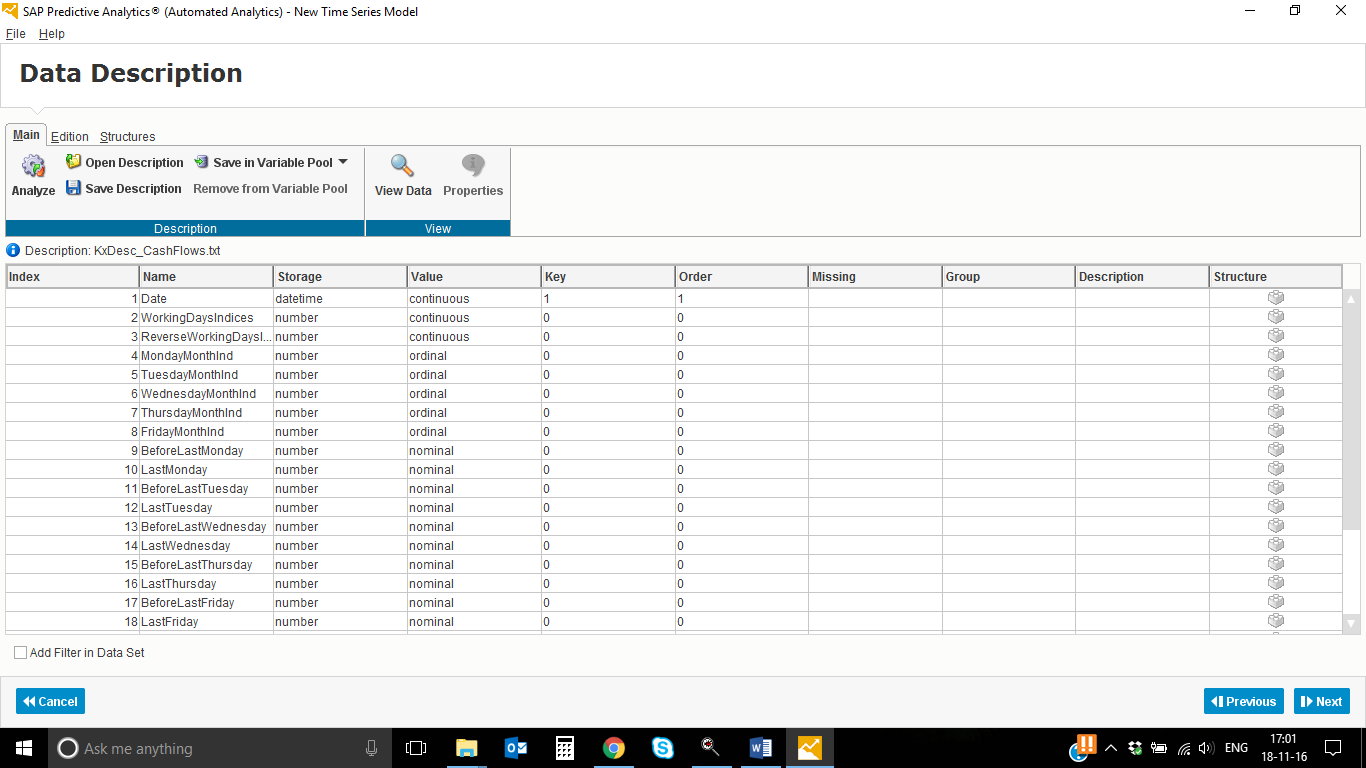


Click OK.

Now, we can see from the data source all the variables along with their associated data types.



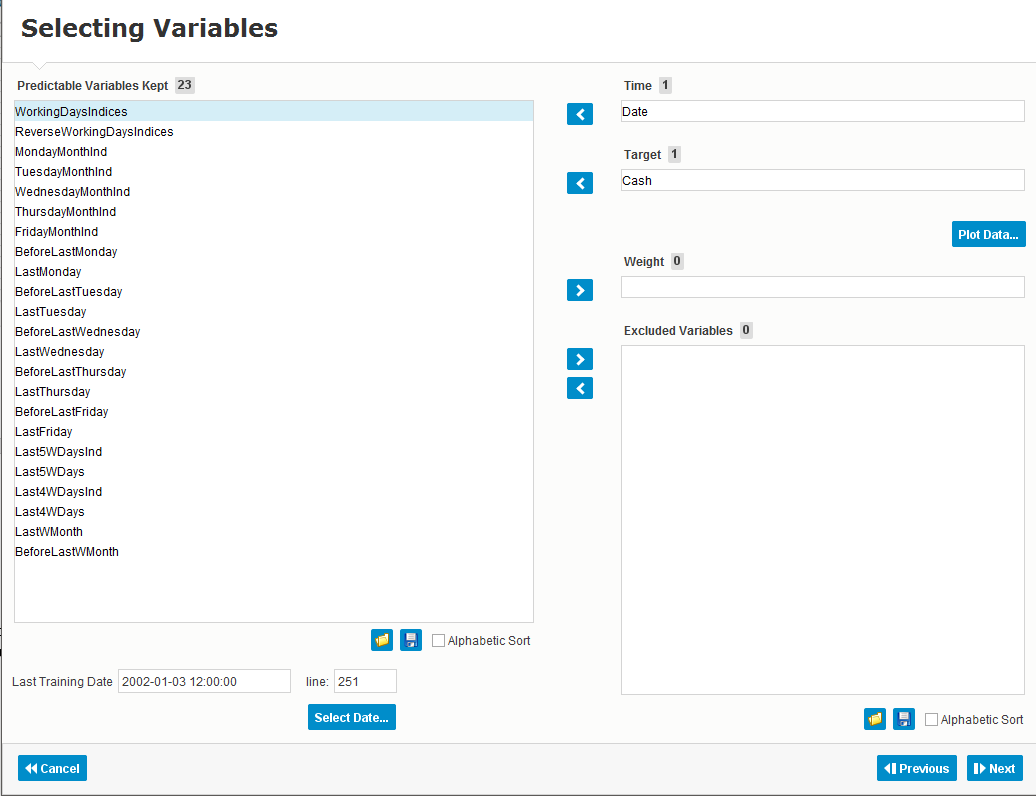
**Paste a screenshot here.**



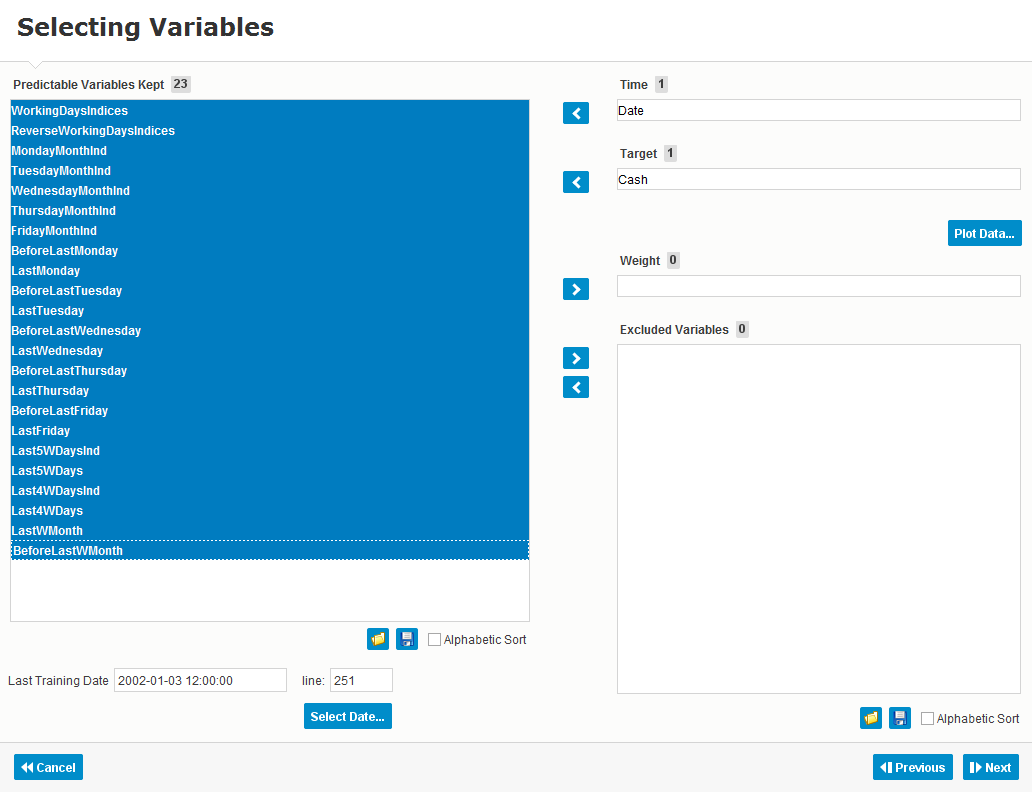
Click Next.

**Selecting Variables**

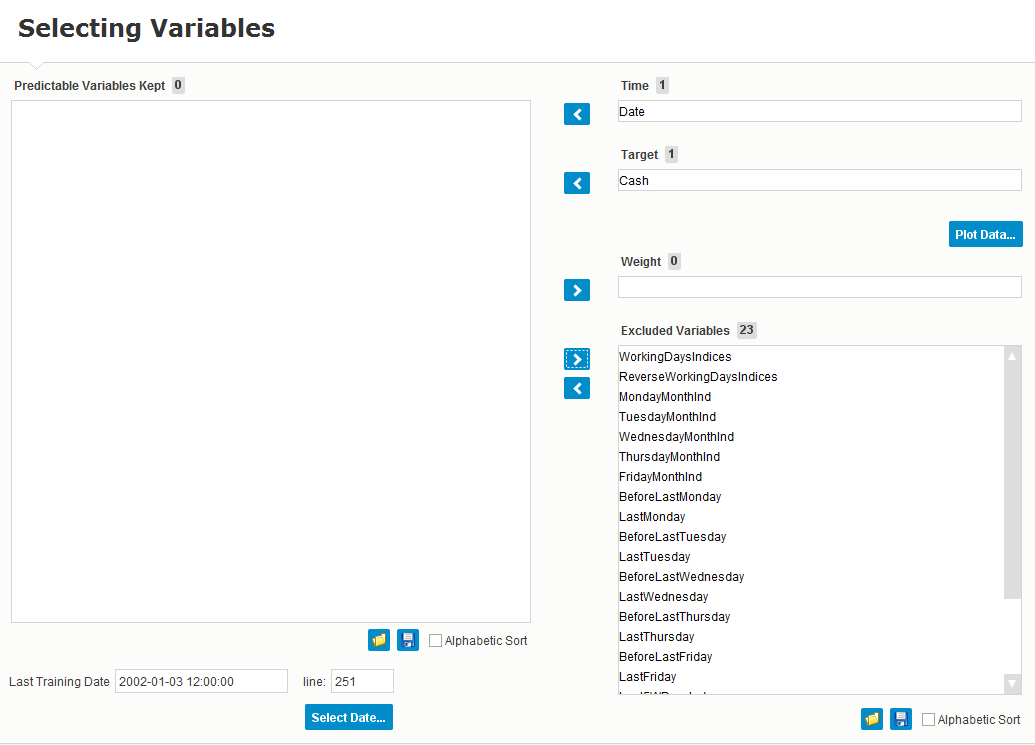
Time and Target variables have been correctly identified by default. On the left is the list of additional variables for analysis. For our first model, we shall exclude all of these additional variables.



Press Shift and select all the variables.



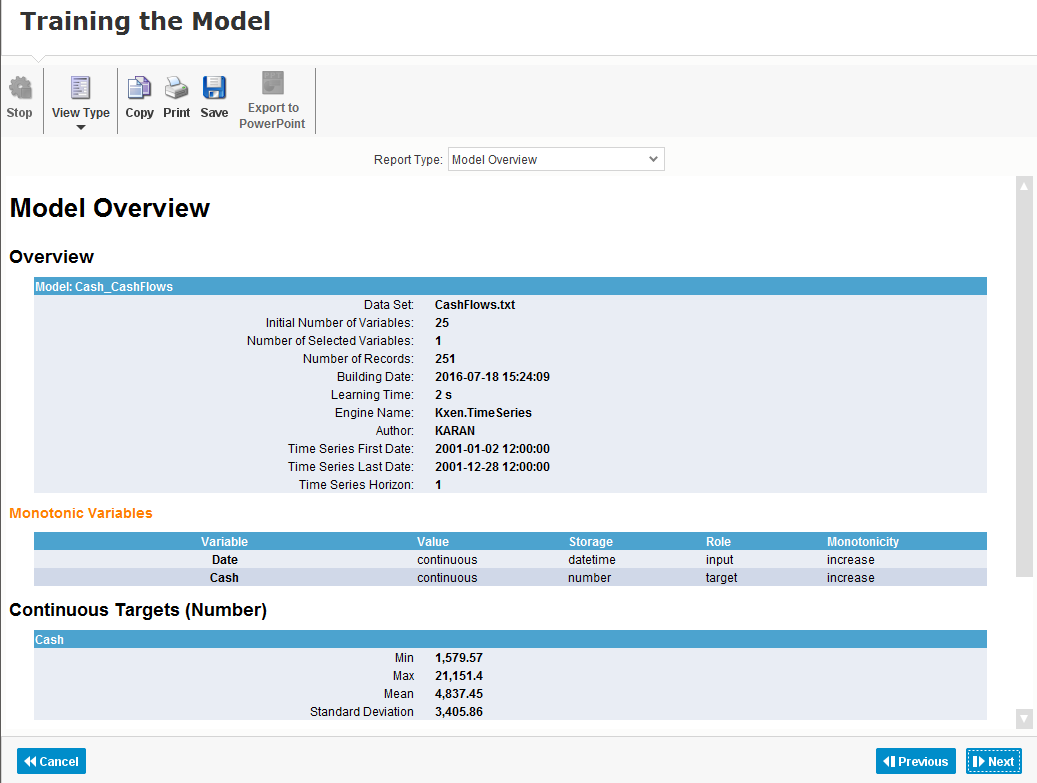
Click to add them to Excluded Variables list.



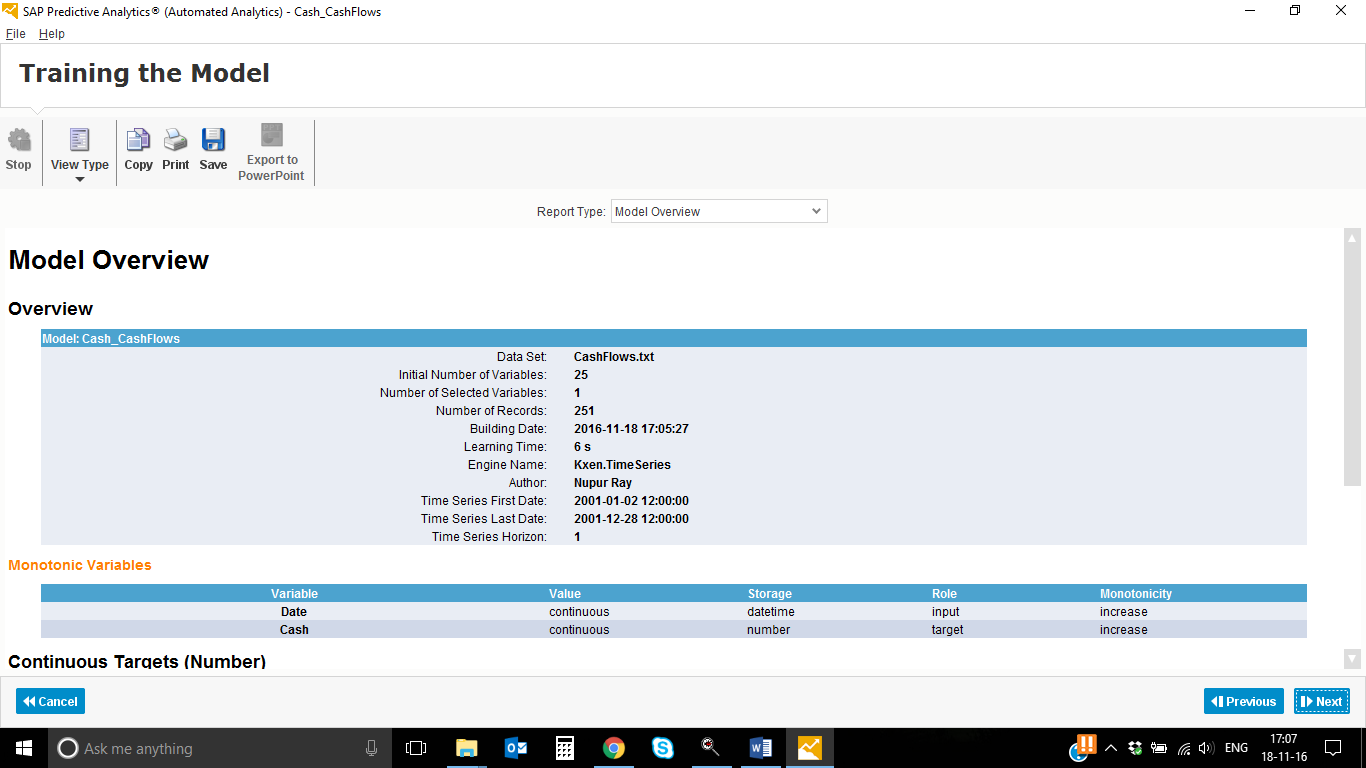
Click Next->Generate.

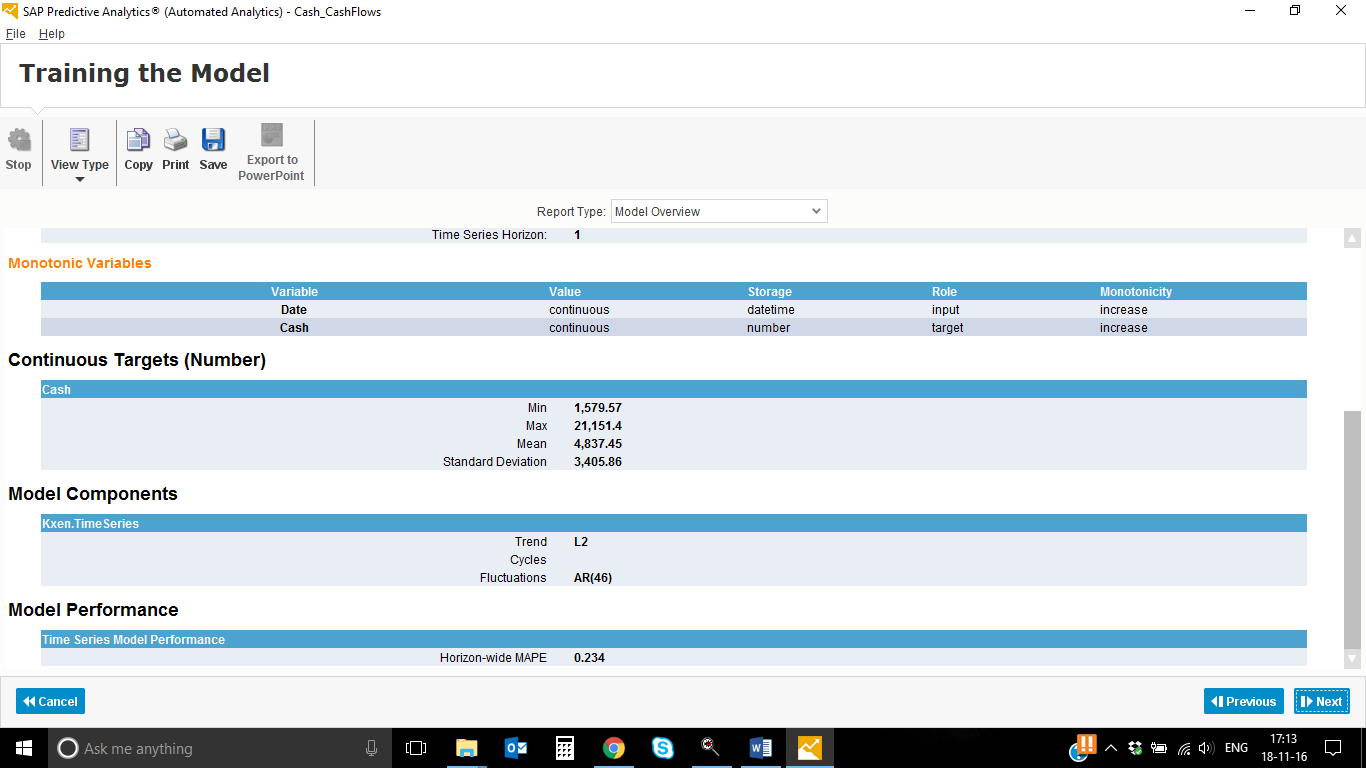
**Training the Model**

Here we can view a summary of model output.

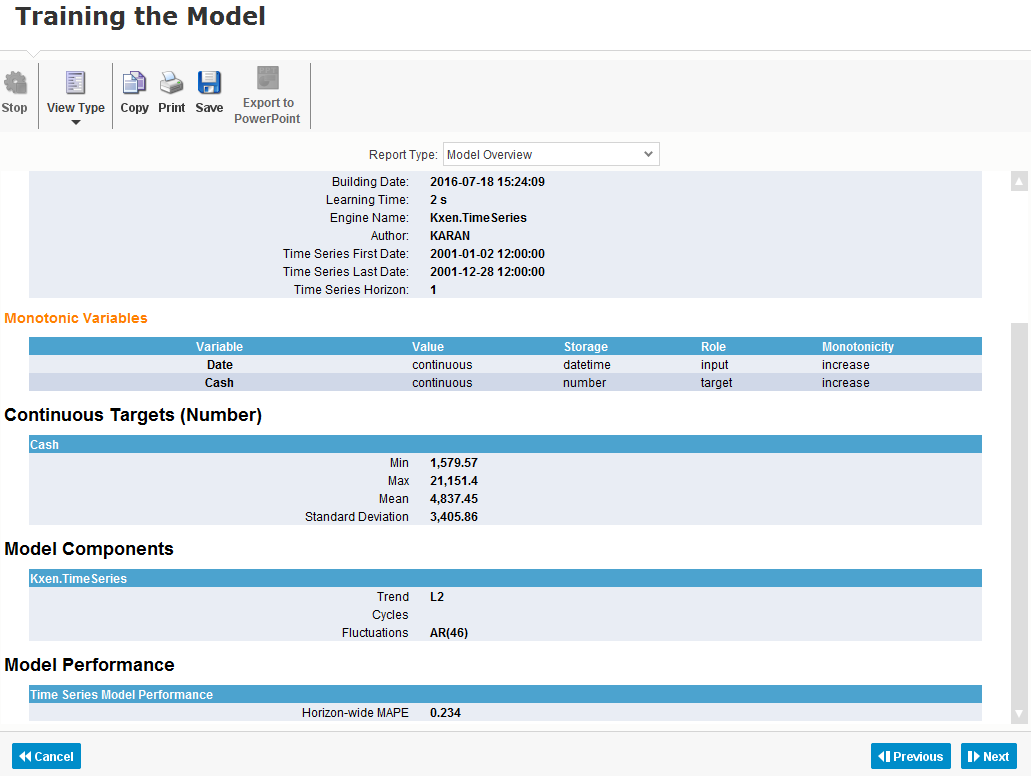


**Paste a screenshot here.**





Note that the Horizon-wide MAPE value for the model is 0.239. This means that the model has a forecasting error of around 24%.

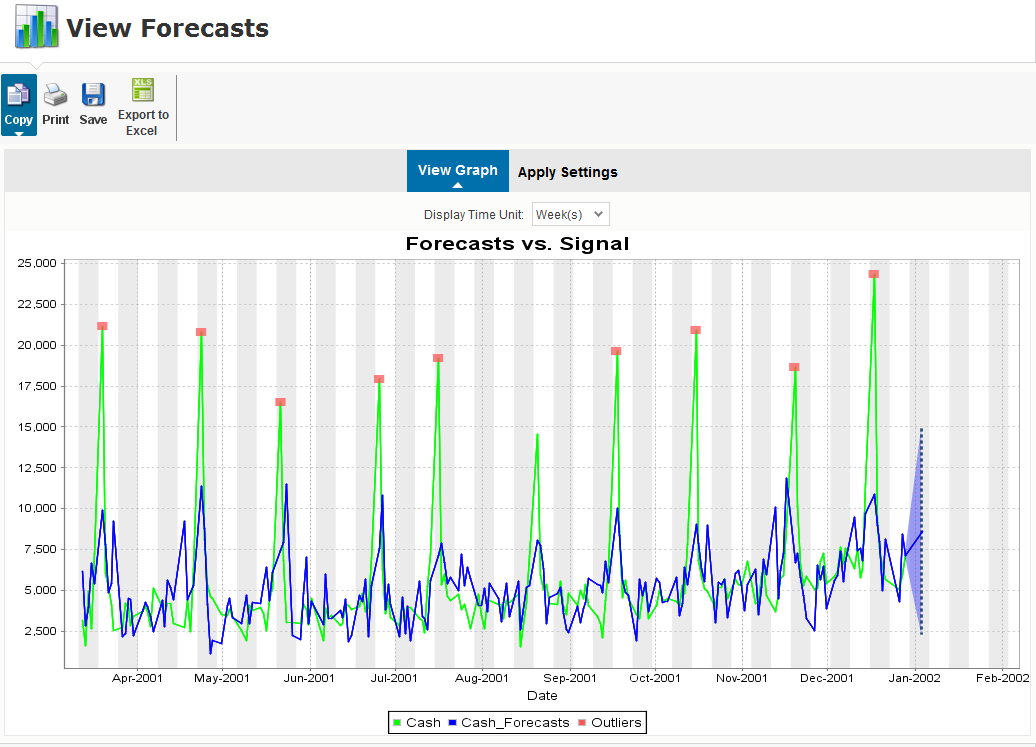


Click Next->View Forecasts.

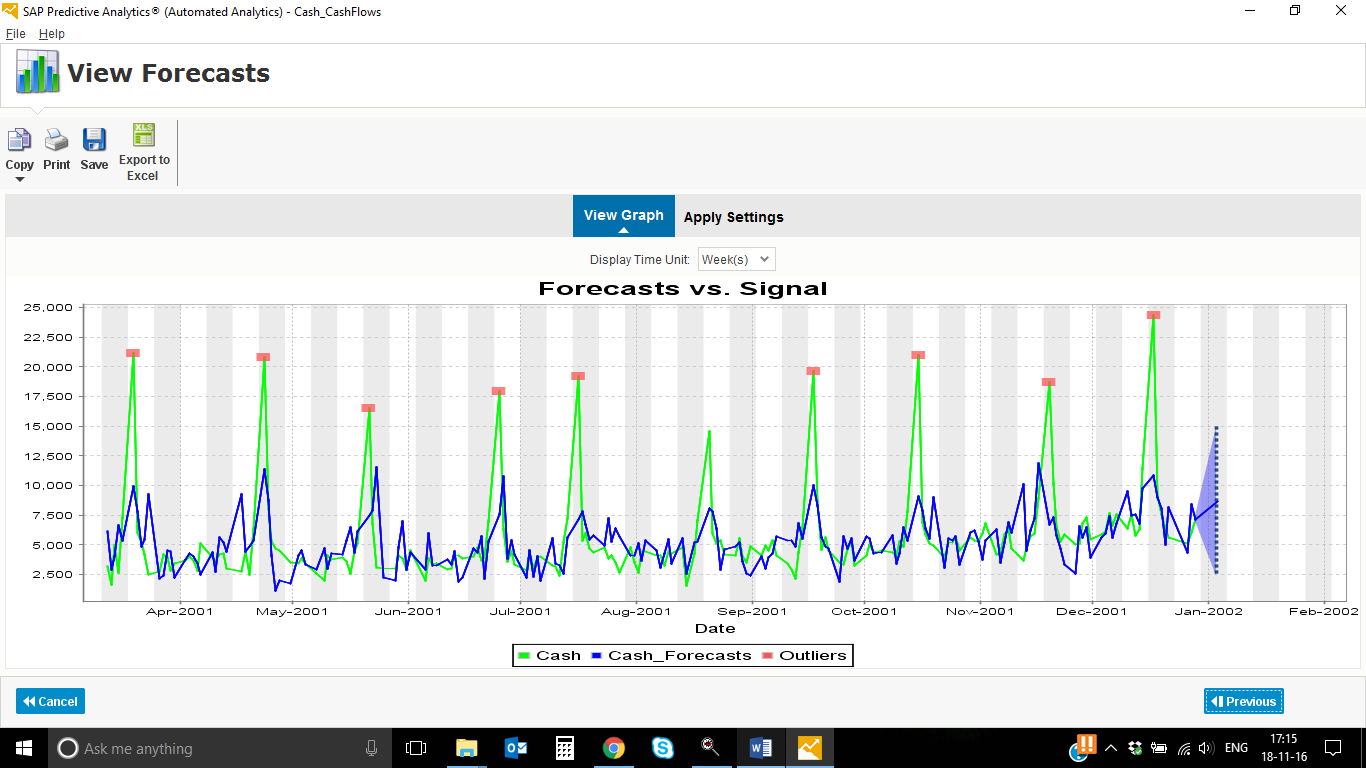
**Using the Model**

Here, we can see the actual Cash values in green against the projected values in blue. The orange blocks identify outliers, where the predicted and actual values differ significantly. At the end of green line, only predicted values in blue are available. Note that there is a blue triangle area that shows the minimum and maximum error range for the predicted values, besides that the range expands over time.

Now we shall re-generate the model and include all of the additional variables in our analysis.

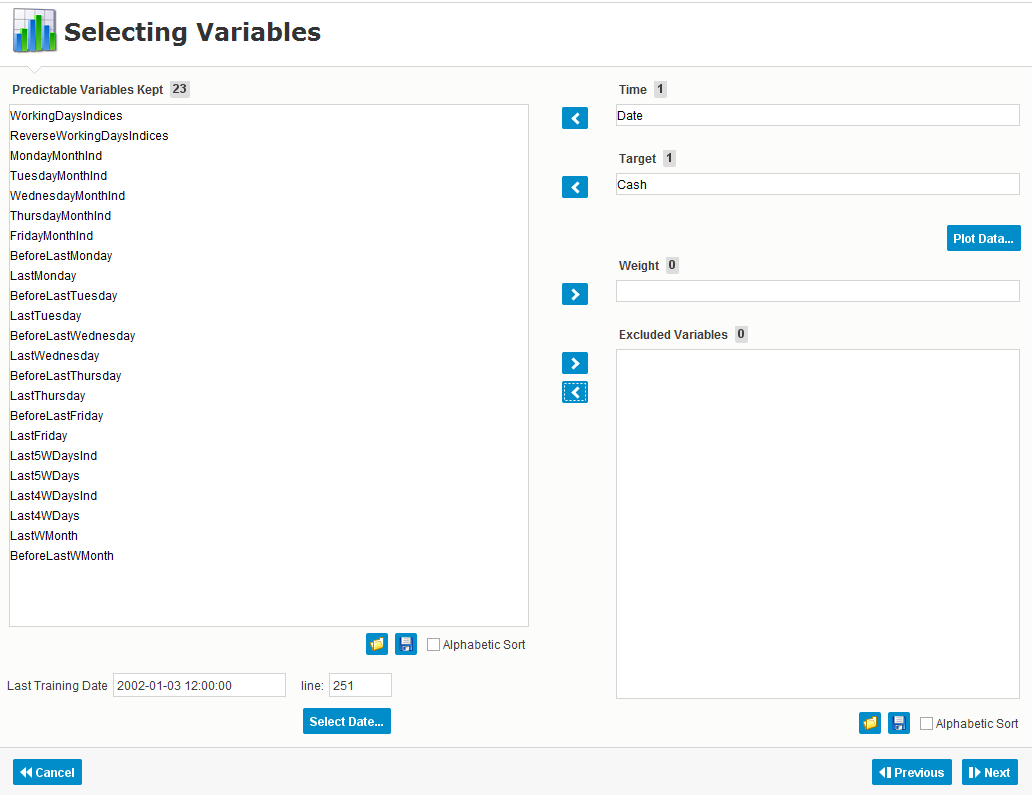


**Paste a screenshot here.**



Click Previous->Previous->Previous->Previous->Previous

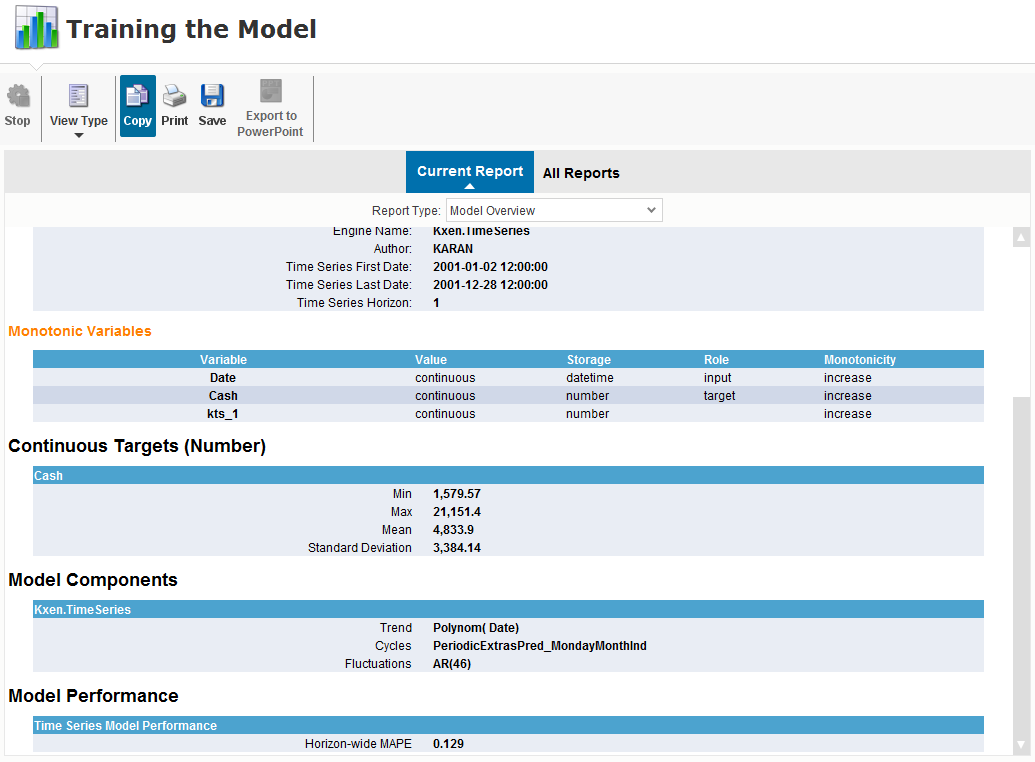
To include all excluded variables in the model, we shall now move them back to Predicted Variables Kept area. Note that these variables are useful, since we have results for them in that period for which we want to predict the missing Cash values.



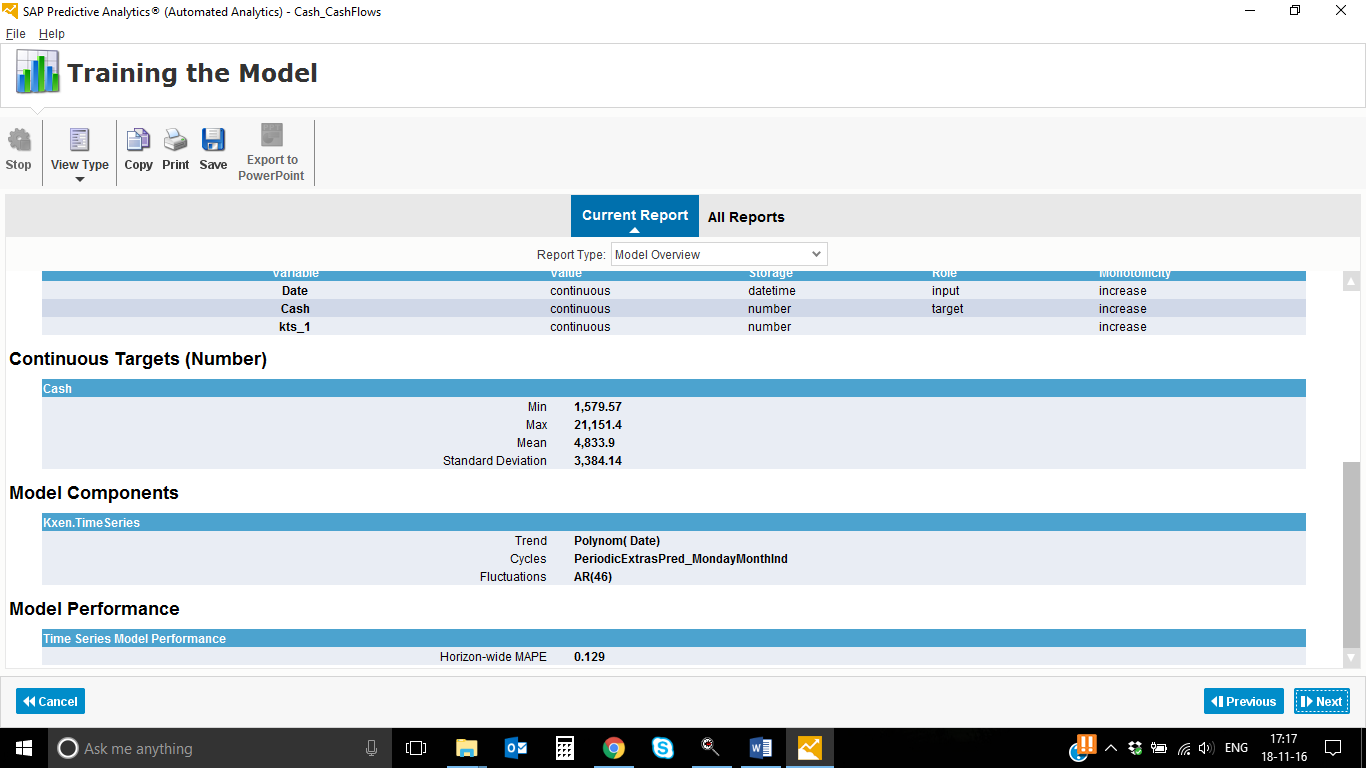
Click Next->Generate.

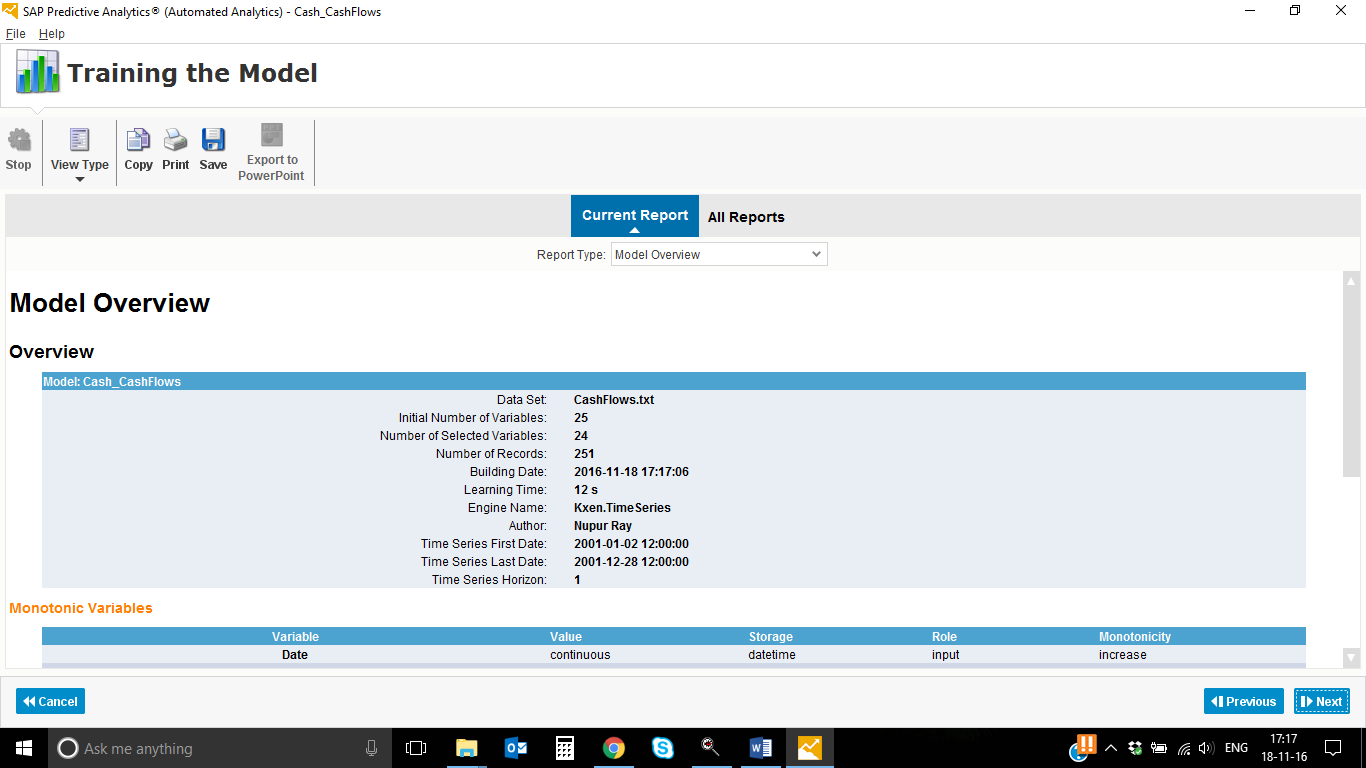
**Training the Model**

Note that Horizon-wide MAPE value for the model is now 0.129 meaning that the model has a forecasting error of around 13%. This is a big improvement over the first model, which had a forecasting error around 24%.



**Paste a screenshot here.**

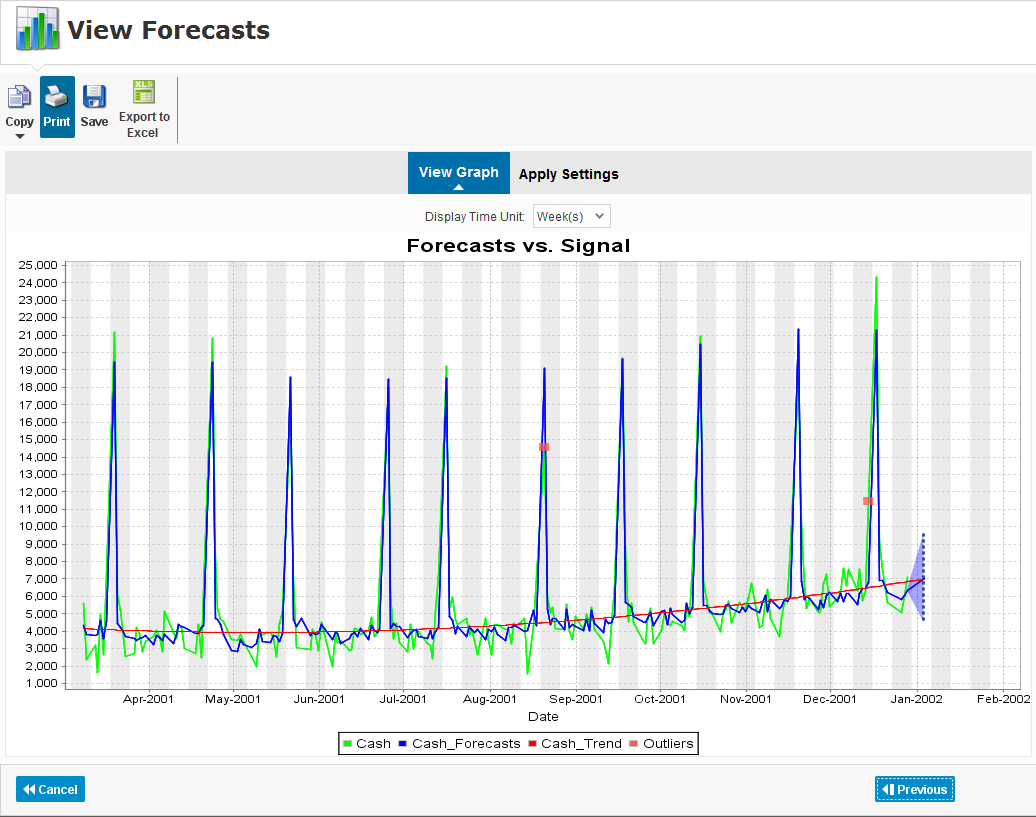




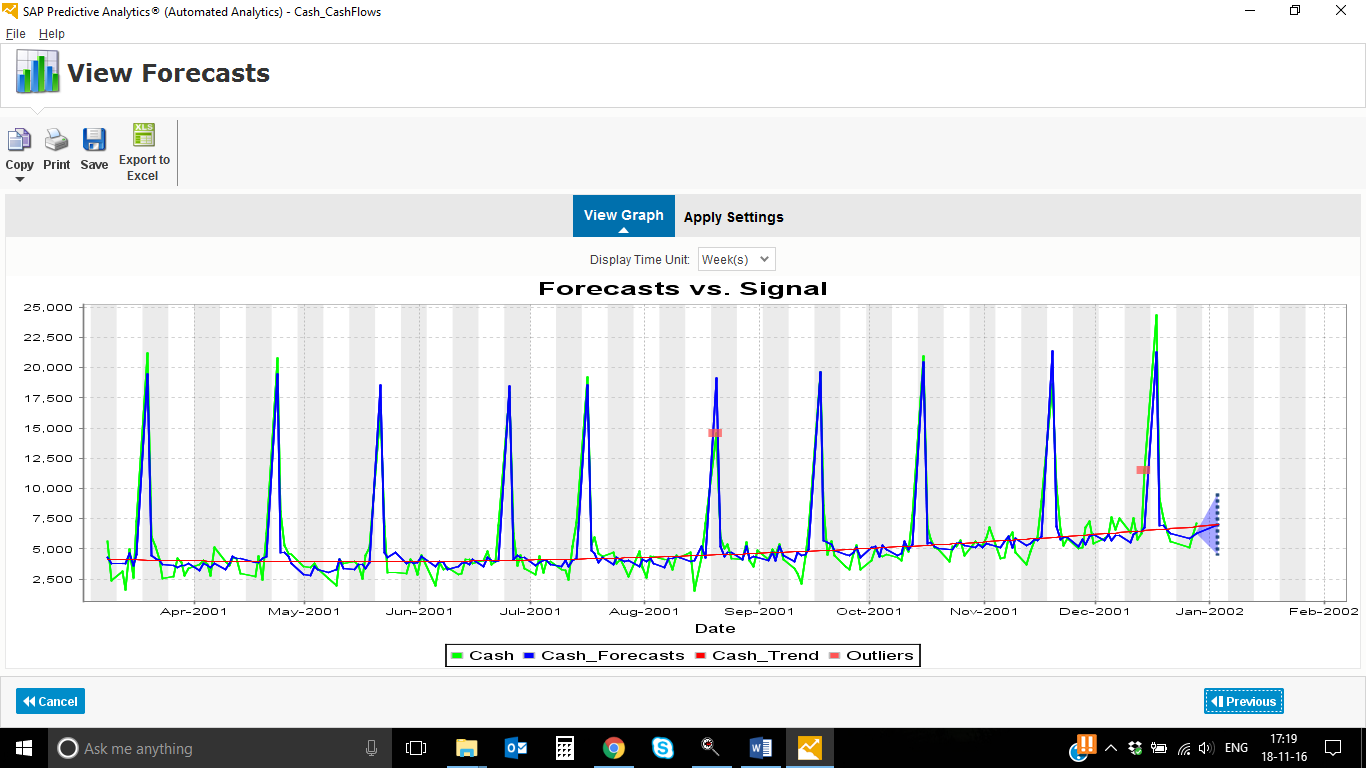
Click Next->View Forecasts.

**View Forecasts**

Here we can see that blue line for forecasts is much closer to the green line measuring the actual Cash values, as also now there are fewer outliers. The blue triangle area, showing the maximum and minimum projected error range, is much smaller, indicating that values can be predicted with greater accuracy.

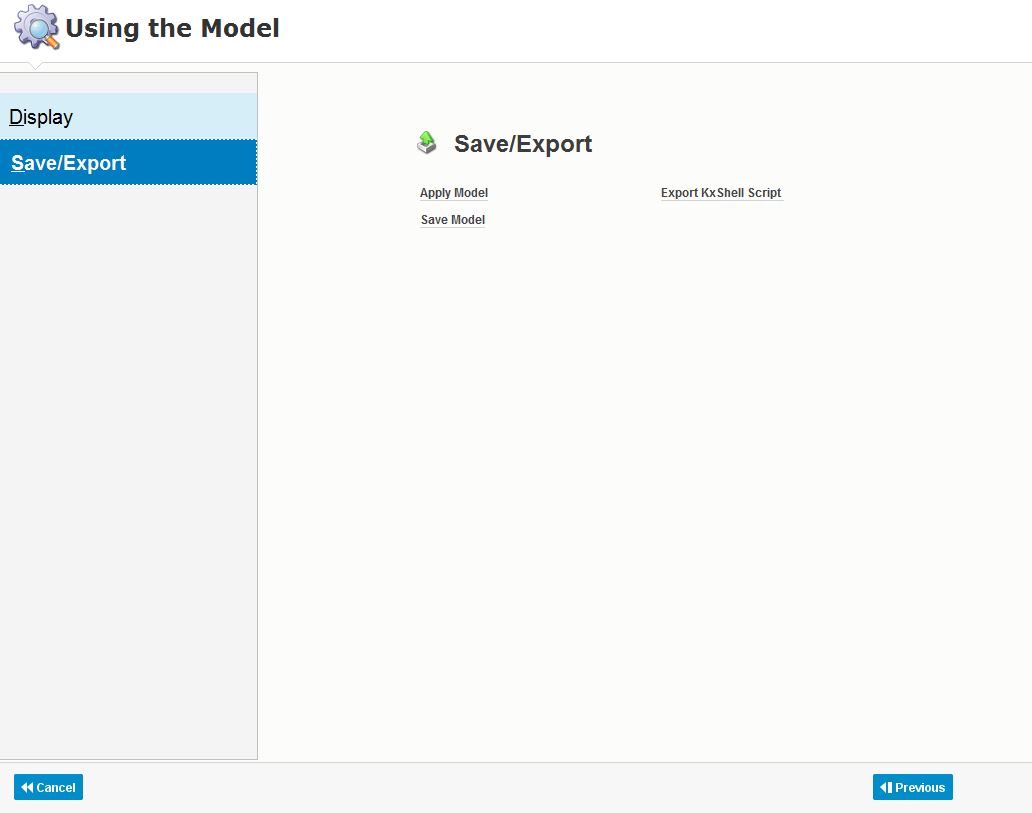


**Paste a screenshot here.**



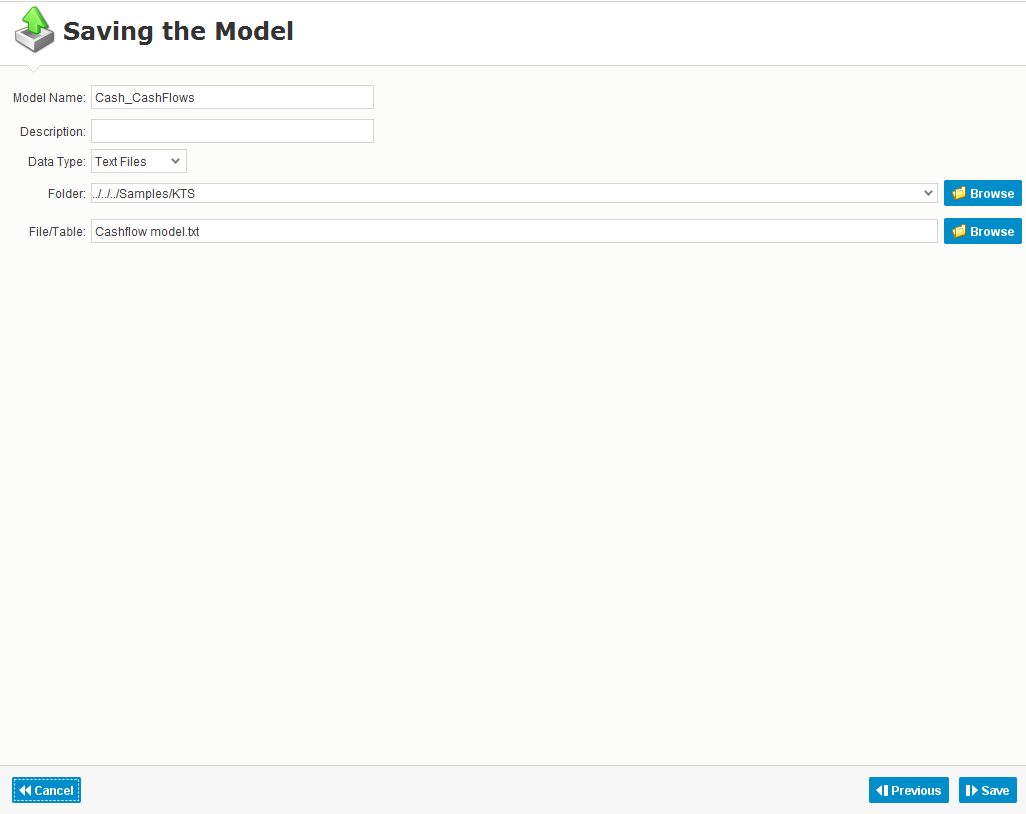
**Using the Model**

Click Previous->Save/Export->Save Model.



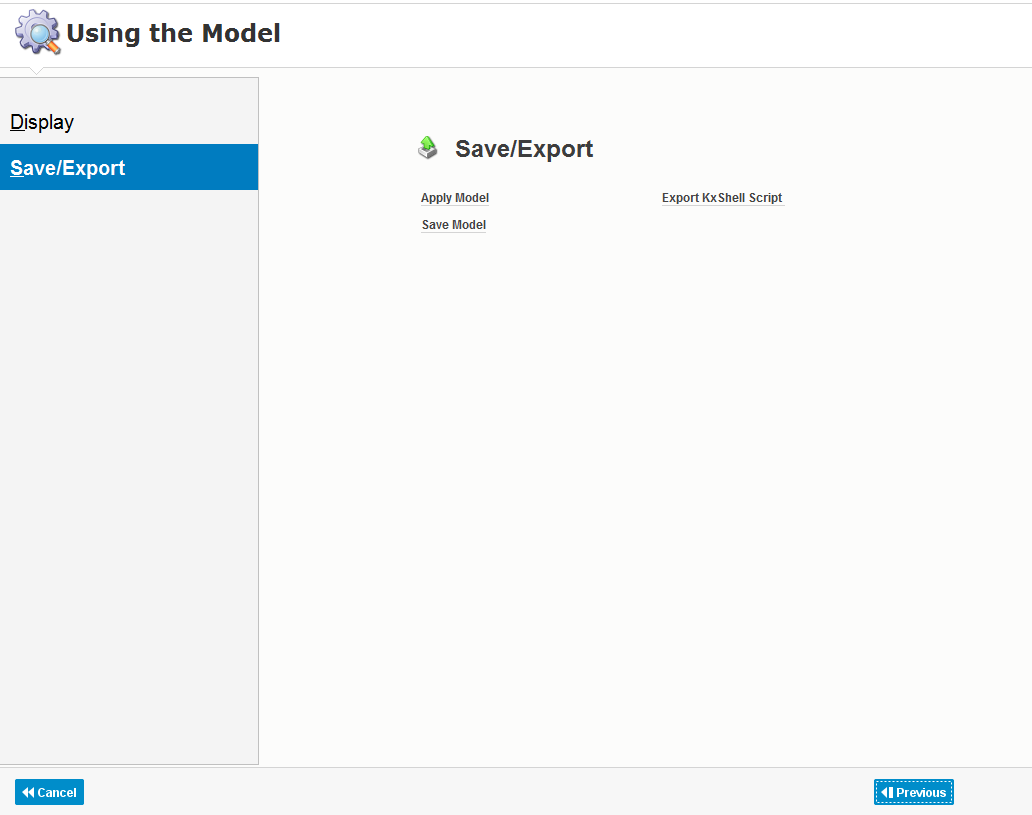
We shall save the model in the exact same location as the original data source. Enter the name as **CashFlow\_SMXXXYY.**

Click Save.



Now we shall apply the model against the same dataset that has been used to generate the model. This will generate an output that contains values predicted by the model for each record.

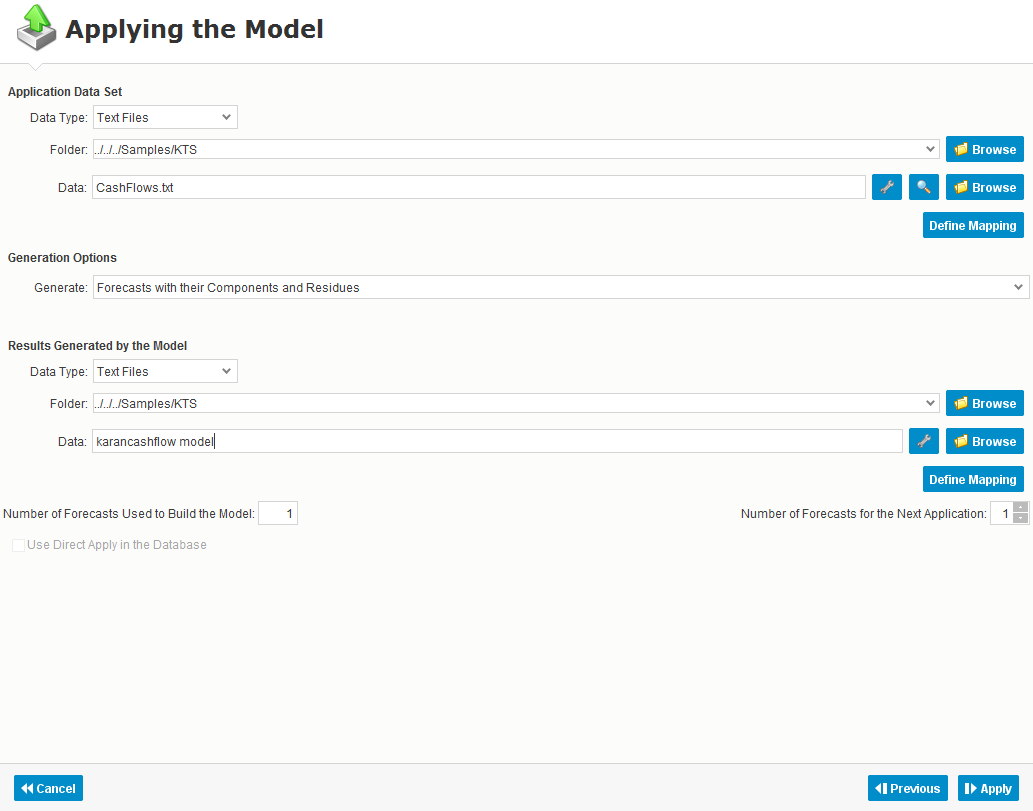
Click Apply Model



**Applying the Model**

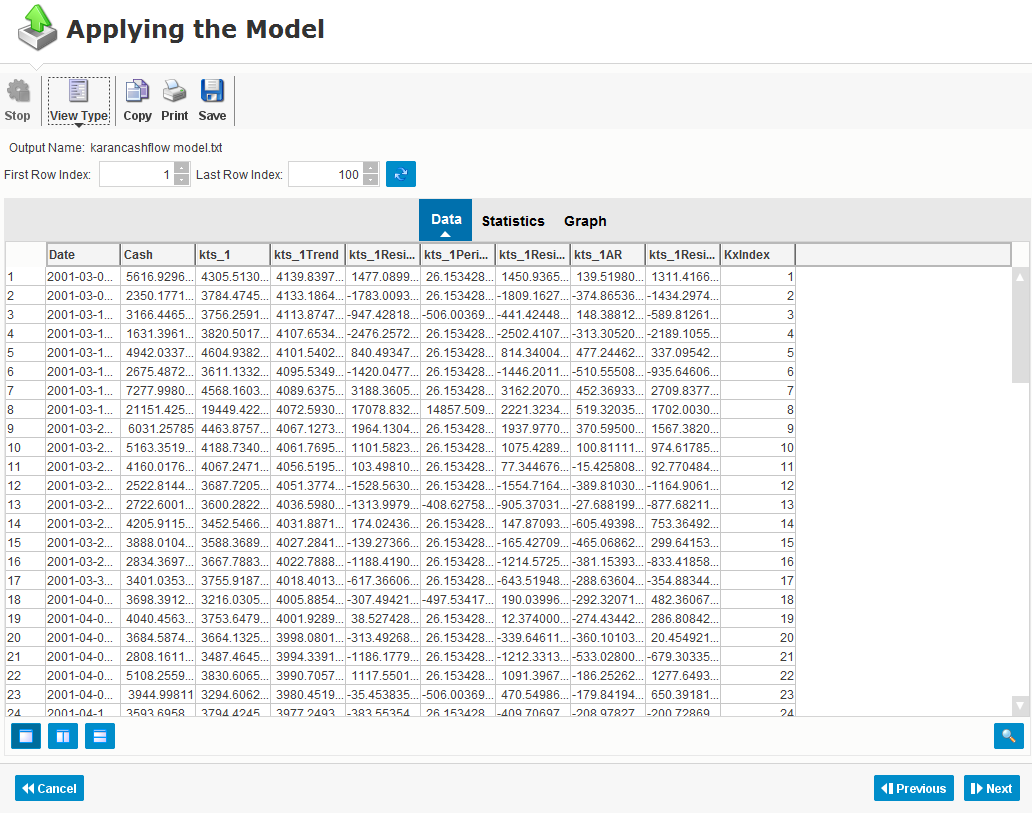
Dataset, used to create the model, is automatically selected as the model application dataset. For this example, we shall use the default generation option, i.e. Predicted Value Only. This output method will include Date and Cash variables from the original dataset, along with predicted Cash values.

Next, we need to enter a name for the output of the model. Enter the name as **CashFlowApply\_SMXXXYY**.



Click Apply->View Output.

Here we can see the Date and Cash columns from the original data source, along with kts\_1 column that holds predicted Cash values.



**Paste a screenshot here.**

