Forecast Time Series Experiment Assistant workflow

Experiments manage the data source, algorithm, and the parameters to configure that algorithm within one framework. An Experiment is an exclusive knowledge object in the Splunk platform that keeps track of its settings and history, as well as its affiliated alerts and scheduled trainings. Experiment Assistants enable machine learning through a guided user interface.

The following visualization shows a forecast of sales numbers using the Kalman Filter algorithm.

Available algorithms

The Forecast Time Series Experiment Assistant uses the following algorithms:

- State-space method using the Kalman filter
- Autoregressive Integrated Moving Average (ARIMA)

Create an Experiment to forecast a time series

The Forecast Time Series Assistant predicts the next value in a sequence of time series data. The result includes both the predicted value and a measure of the uncertainty of that prediction. Forecasting is one type of prediction. Forecasting refers to the use of past time series data to make predictions.

To build a forecast, input the data and select the field by which you want to forecast.

Assistant workflow

Follow these steps to create a Forecast Time Series Experiment.
1. From the MLTK navigation bar, click **Experiments**.
   - If this is the first Experiment in the MLTK, you will land on a display screen of all six Assistants. Select the **Forecast Time Series** block.
   - If you have at least one Experiment in the MLTK, you will land on a list view of Experiments. Click the **Create New Experiment** button.
2. Fill in an Experiment Title, and add a description. Both the name and description can be edited later if needed.
3. Click **Create**.
4. Run a search, and be sure to select a date range.
5. (Optional) Click **Add a step** to add preprocessing steps
6. Select an algorithm from the **Algorithm** drop-down menu.
   - If you are not sure which algorithm to choose, start with the default Kalman filter algorithm.
7. From the **Field to Forecast** list, select the field you want to forecast.
   - The **Field to Forecast** drop-down menu populates with fields from the search.
8. Select your parameters. Use the following table to guide your choices.

<table>
<thead>
<tr>
<th>If</th>
<th>Then</th>
</tr>
</thead>
<tbody>
<tr>
<td>You chose the <strong>Kalman</strong> filter algorithm</td>
<td>These methods consider subsets of features such as local level (an average of recent values), trend (a slope of line that fits through recent values), and seasonality (repeating patterns). Specify the values for: AR (autoregressive); p, I (integrated); d, MA (moving average); q parameters.</td>
</tr>
<tr>
<td>You chose the <strong>ARIMA</strong> algorithm</td>
<td>For example, AR(1) means you would forecast future values by looking at 1 past value. I(1) means it took 1 difference, where each data point was subtracted from the one that follows it, to make the time series stationary. MA(1) means you would forecast future values using 1 previous prediction error.</td>
</tr>
</tbody>
</table>
9. Specify the **Future Timespan**, which indicates how far beyond the data you want to forecast. The size of the confidence interval is used to gauge how confident the algorithm is in its forecast.
10. In the **Holdback** field, specify the number of values to withhold. Decide how many search results to use for validating the quality of the forecast. The larger the withholding, the fewer values available to train your model.
11. Select the **Confidence Interval**, which is the percentage of the future data you expect to fall inside of the confidence envelope.
12. For the Kalman algorithm, select the **Period**, which indicates the period of any known repeating patterns in the data to assist the algorithm.
   - For example, if your data includes monthly sales data that follows annual
patterns, specify 12 for the period.

13. (Optional) Add notes to this Experiment. Use this free form block of text to track the selections made in the Experiment parameter fields. Refer back to notes to review which parameter combinations yield the best results.

14. Click **Forecast**. The Experiment is now in a Draft state.

Draft versions allow you to alter settings without committing or overwriting a saved Experiment. An Experiment is not stored to Splunk until it is saved.

The following table explains the differences between a draft and saved Experiment:

<table>
<thead>
<tr>
<th>Action</th>
<th>Draft Experiment</th>
<th>Saved Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create new record in Experiment history</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Run Experiment search jobs</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>(As applicable) Save and update Experiment model</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>(As applicable) Update all Experiment alerts</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>(As applicable) Update Experiment scheduled trainings</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

You cannot save a model if you use the DBSCAN or Spectral Clustering algorithm.

**Interpret and validate results**

After you forecast a time series, review your results in the following tables and visualizations.

<table>
<thead>
<tr>
<th>Result</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Data Preview</td>
<td>This displays the raw data from the search.</td>
</tr>
<tr>
<td>Forecast</td>
<td>This graphs displays the actual value as a solid line and the predicted value as a dotted line, surrounded by a confidence envelope.</td>
</tr>
</tbody>
</table>

Values that fall outside the confidence envelope are outliers. A vertical line indicates where training data
<table>
<thead>
<tr>
<th>Result</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>stops and test data</td>
<td>stops and test data begins. When the real data ends, forecasted values are displayed in shades of green. The larger the envelope, the less confidence we have about forecasts around that time. The size of the envelope is directly related to the specified confidence interval percentage.</td>
</tr>
<tr>
<td>R2 Statistic</td>
<td>This statistic explains how well the model explains the variability of the result. 100% (a value of 1) means the model fits perfectly. The closer the value is to 1 (100%), the better the result.</td>
</tr>
<tr>
<td>Root Mean Squared Error</td>
<td>This statistic explains the variability of the result, essentially the standard deviation of the residual. The formula takes the difference between actual and predicted values, squares this value, takes an average, and then takes a square root. The result is an absolute measure of fit, the smaller the number the better the fit. These values only apply to one dataset and are not comparable to values outside of it.</td>
</tr>
<tr>
<td>Prediction Outliers</td>
<td>This result shows the total number of outliers detected.</td>
</tr>
</tbody>
</table>

**Predicting with the ARIMA algorithm**

When predicting using the ARIMA algorithm, additional autocorrelation panels are present. Autocorrelation charts can be used to estimate and identify the three main parameters for the model:

- the autoregressive component $p$
- the integrated component or order of differencing $d$
- the moving average component $q$

**ACF: Autocorrelation function chart**

The autocorrelation function chart shows the predicted field's autocorrelations at various lags, surrounded by confidence interval lines. For example, the column at lag 1 shows the amount of correlation between the time series and a lagged version of itself.
**PACF: Partial autocorrelation function chart**

The partial autocorrelation function chart shows the predicted field's autocorrelations at various lags while controlling for the amount of correlation contributed by earlier lag points. This chart is also surrounded by confidence interval lines. For example, the column at lag 2 shows the amount of correlation between the time series and a lagged version of itself, while removing the correlation contributed by the lag 1 data points.
**ACF Residual: Autocorrelation function residual chart**

The autocorrelation function residual chart shows prediction errors. The errors are the difference between the series and the predictions. The ACF of the residuals should be close to zero. If the errors are highly correlated, the model might be poorly parameterized, or the series might not be stationary.

![ACF Residual Chart](image)

**PACF Residual: partial autocorrelation function residual chart**

The partial autocorrelation function chart shows prediction errors. The errors are the difference between the series and the predictions. The PACF of the residuals should be close to zero. If the errors are highly correlated, the model might be poorly parameterized or the series might not be stationary.

![PACF Residual Chart](image)
Refine the Experiment

After you create a forecast, you can select an alternate algorithm option to see whether a different choice yields better results.

Be advised that the quality of the forecast primarily depends on the predictability of the data.

Save the Experiment

Once you are getting valuable results from your Experiment, save it. Saving your Experiment results in the following actions:

1. Assistant settings saved as an Experiment knowledge object.
2. The Draft version saves to the Experiment Listings page.
3. Any affiliated scheduled trainings and alerts update to synchronize with the search SPL and trigger conditions.

You can load a saved Experiment by clicking the Experiment name.

Deploy the Experiment

Saved forecast time series Experiments include options to manage, but not to publish.

In Experiments built using the Forecast Time Series Assistant a model is not persisted, meaning you will not see an option to publish. However, you can achieve the same results as publishing the Experiment through the steps below for Outside the Experiment framework.

**Within the Experiment framework**

To manage your Experiment, perform the following steps:

1. From the MLTK navigation bar, choose Experiments. A list of your saved Experiments populates.
2. Click the Manage button available under the Actions column.
The toolkit supports the following Experiment management options:

- Create and manage Experiment-level alerts. Choose from both Splunk platform standard trigger conditions, as well as from Machine Learning Conditions related to the Experiment.
- Edit the title and description of the Experiment.
- Delete an Experiment.

Updating a saved Experiment can affect affiliates alerts. Re-validate your alerts once you complete the changes. For more information about alerts, see Getting started with alerts in the Splunk Enterprise Alerting Manual.

Experiments are always stored under the user's namespace, meaning that changing sharing settings and permissions on Experiments is not supported.

**Outside the Experiment framework**

1. Within the Experiment click **Open in Search** to generate a New Search tab for this same dataset. This new search will open in a new browser tab, away from the Assistant.
   This search query uses all data, not just the training set. You can adjust the SPL directly and see results immediately. You can also save the query as a Report, Dashboard Panel or Alert.
2. Click **Show SPL** to generate a new modal window/ overlay showing the search query you used to forecast. Copy the SPL to use in other aspects of your Splunk instance.

**Learn more**

To learn about implementing analytics and data science projects using Splunk's statistics, machine learning, and built-in custom visualization capabilities, see the Splunk Education course of Splunk for Analytics and Data Science.