RapidMiner

Séries Temporais - Validação

*Rui Humberto Pereira*

*Rui.humberto.pereira@gmail.com*

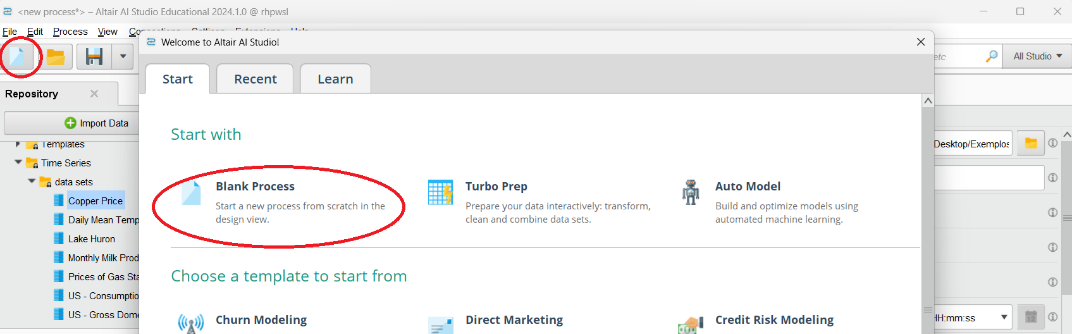
*Fevereiro de 2025*

**Validação do modelo**: Numa Série Temporal, o modelo faz uma previsão de evolução do valor da grandeza em análise ao longo do tempo. O erro da previsão é tanto maior quando a diferença entre os valores previstos e os valores reais observados.

Neste exercício vamos determinar o erro nos valores previstos pelo modelo de previsão no exercício anterior.

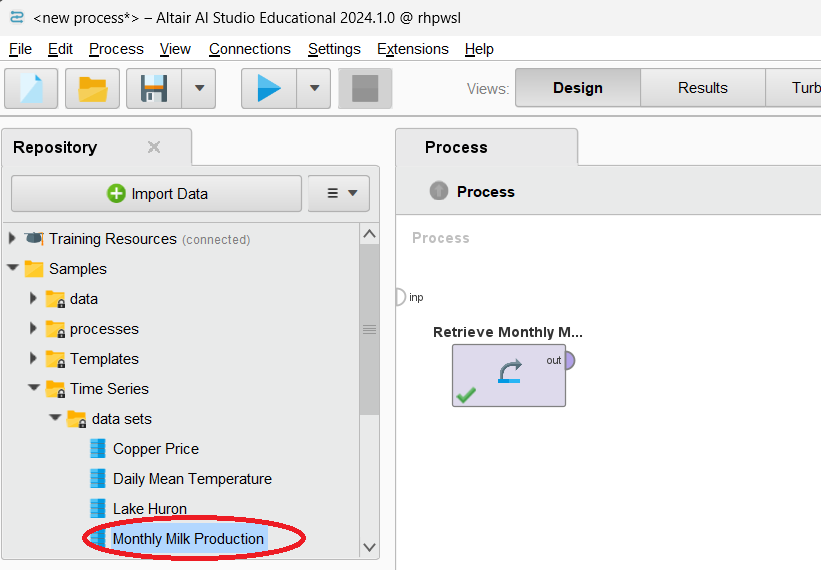
## Inicie um novo projeto no AI Studio

* Menu File -> New Process



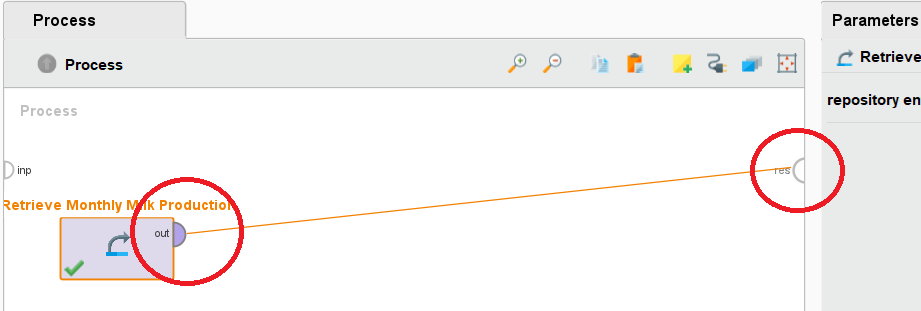
## Obtenção de dados

* O RapidMiner disponibiliza um conjunto de dados de exemplo que vamos utilizar neste exercício. No explorador do Repositório, selecione o dataset de demonstração “Monthly Milk Production” e arraste-o para a área de design do processo.

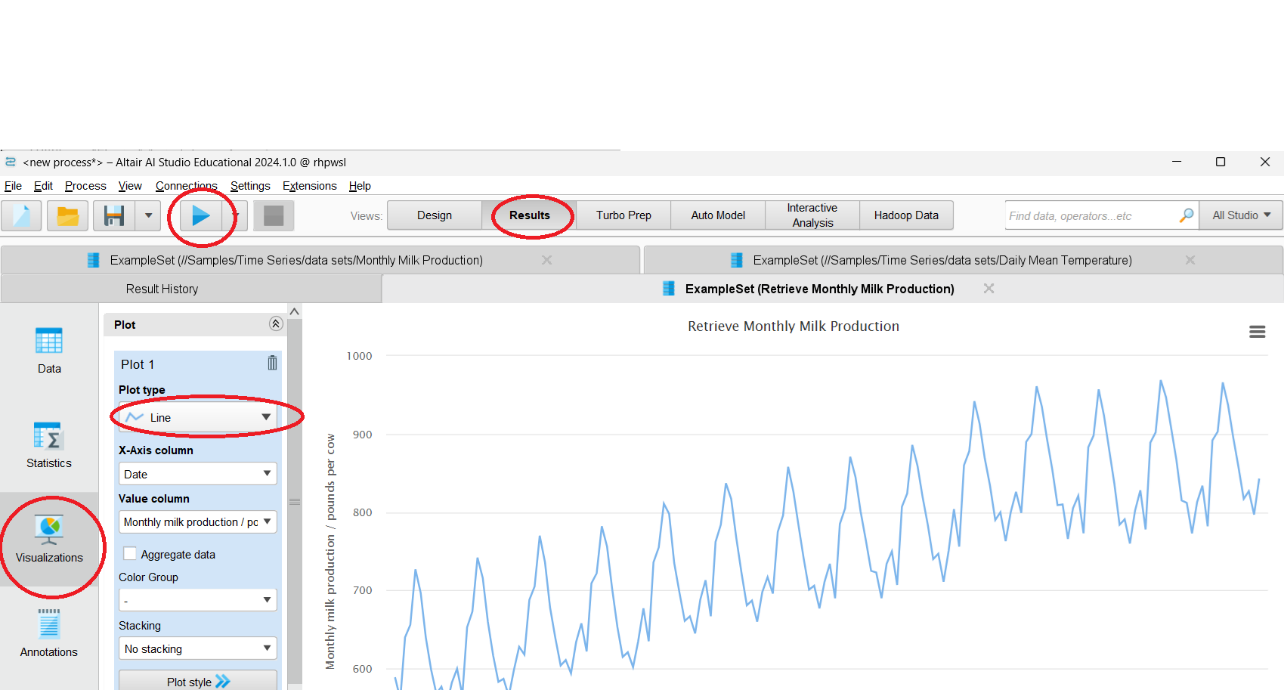


## Visualizar os dados

* Ligue o dataset à Saída de resultados premindo o botão do rato enquanto define a ligação.



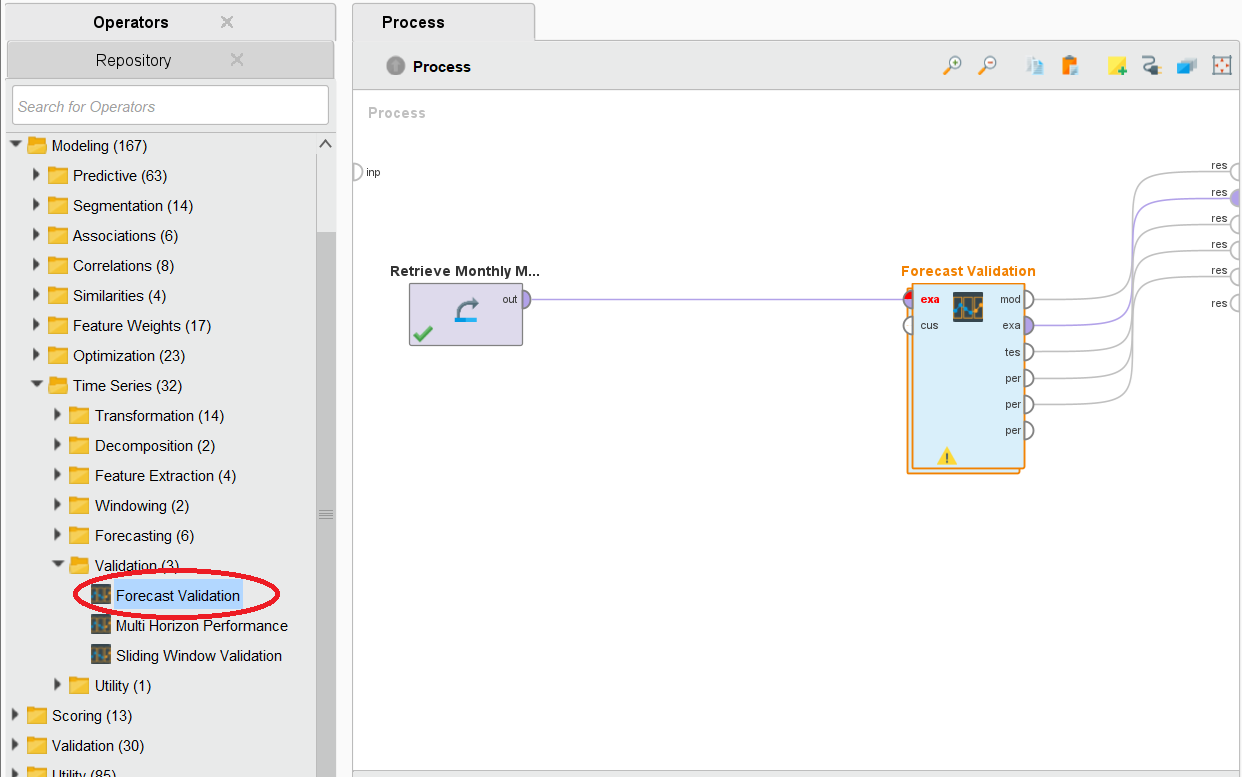
* Depois de estabelecida a ligação, faça “Run” do processo clicando no botão run . No separador de Resultados opte por visualizar os dados utilizando um gráfico de linhas.



* O gráfico que visualiza apresenta a produção de leite desde 1962 e 1976. Podemos verificar que existe uma clara tendência de crescimento na produção.

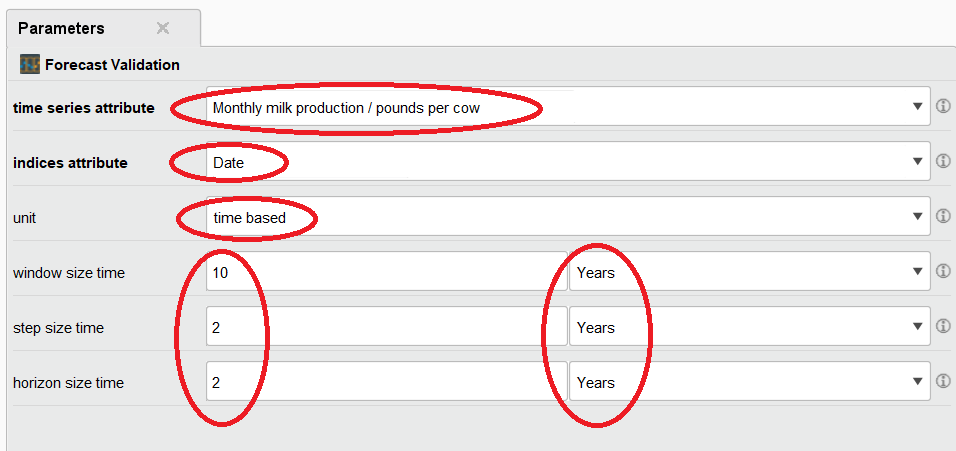
## Implementar e validar o modelo

* No exercício anterior desenvolvemos um modelo preditivo capaz de prever como a produção de leite evolui ao longo do tempo.
* Para validar esse modelo vamos dividir as nossas amostras de valores reais em duas partes, treino e teste, e determinar o erro ao tentar prever os dados de teste.
* Acrescente um bloco (Processo) de validação (Forecast Validation) que recebe os dados e valida o modelo. Estabeleça a ligações conforme a figura.

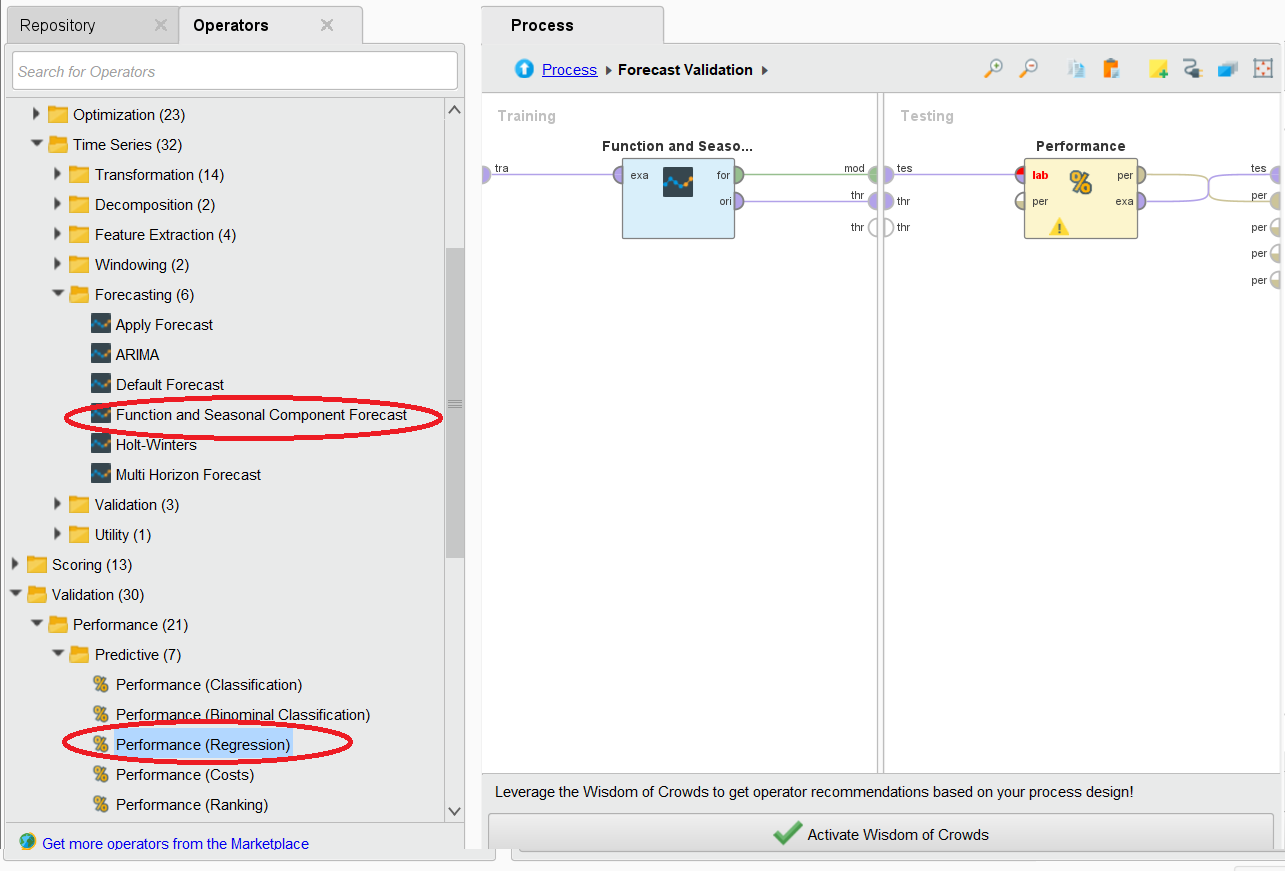


* Abra os parâmetros deste operador. Defina a unidades como “time based”, os dados dos primeiros 10 anos para treino (window size time), 2 anos para a validação (horizon size time) e no parâmetro “Step size time” coloque 2 anos.

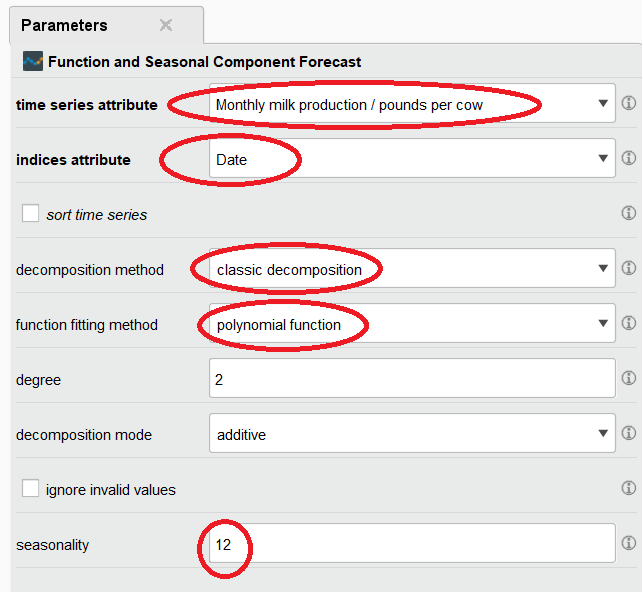
Define o atributo da série temporal que será utilizado para e o índice, que neste caso é a data.



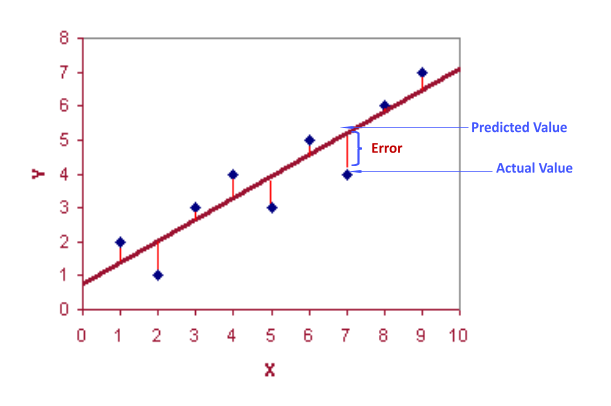
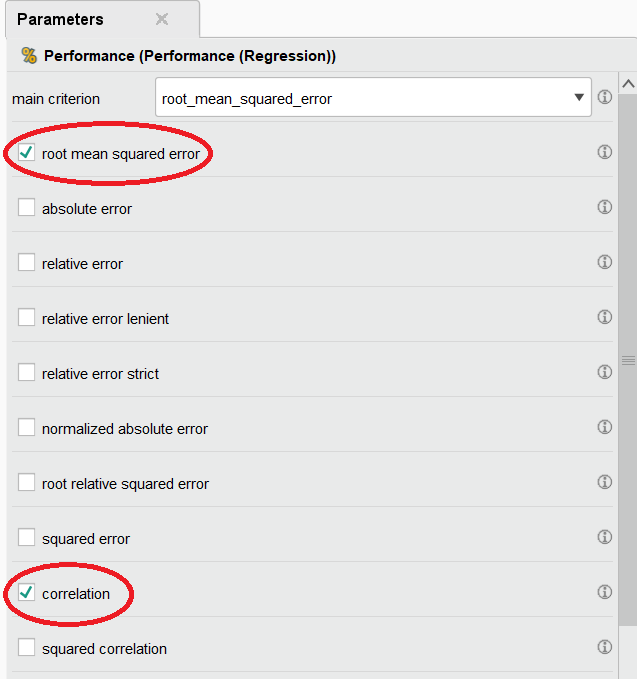
* Para completar a definição do processo de validação, teremos de definir o modelo de Forecast que será treinado com os dados de treino e como o desempenho deste modelo será avaliado. Dando um duplo clique no processo de validação abrem duas áreas: uma para definir o modelo e outra para o validar. Arraste para cada uma das áreas os blocos “Function And Seasonal Component Forecast” e “Performance (Regression)” e defina as ligações conforme a figura.



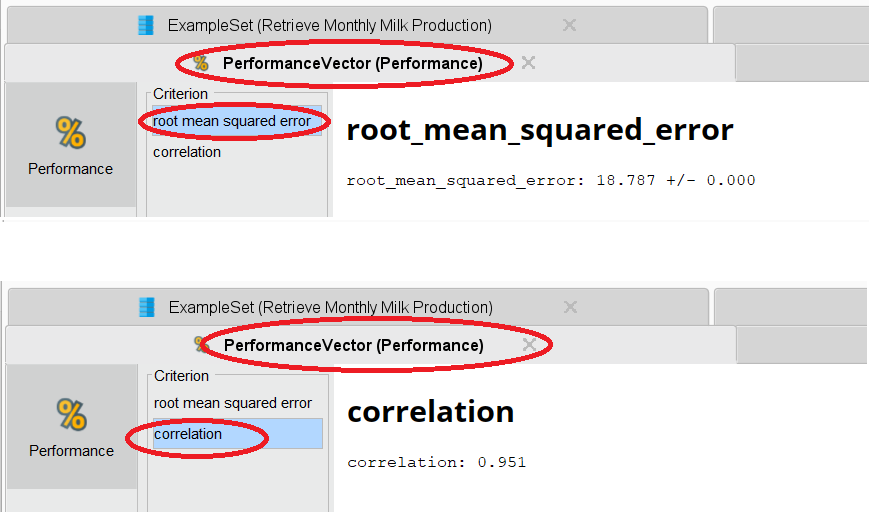
* Abra o separador com os parâmetros do modelo e defina os mesmos que utilizou no exercício anterior



* Depois os parâmetros do bloco que avalia o desempenho em termos de Root Mean Squared Error e de correlação.



* Execute o processo fazendo clicando no botão run .
* Visualize as duas métricas de desempenho



* Visualize graficamente a previsão para os dois anos



* Experimente outros parâmetros em termos das janelas de treino e verifique como as métricas de desempenho se comportam. Também poderá testar o desempenho do modelo aplicando a decomposição multiplicativa.
* Guarde o processo no repositório ou exporte-o para um ficheiro externo.

