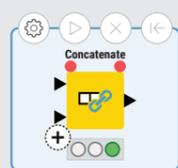


Guia rápido: Construindo um fluxo de trabalho no KNIME para iniciantes

Começando com a Plataforma KNIME Analytics

- Use o Guia de Introdução para dar seus primeiros passos com fluxos de trabalho visuais em: www.knime.com/getting-started-guide
- Saiba mais sobre os nós incluídos e explore exemplos de trabalho em: **KNIME Analytics Platform Version 5 Starter Perspective Collection on KNIME Community Hub.**



Barra de ações do nó: Interaja diretamente com o nó para configurar, executar ou redefinir um nó. **Configurar:** Abre a janela de configuração. **Executar:** Executa o nó. **Cancelar:** Cancela a execução do nó. **Redefinir:** Redefine o nó. **Rótulos do nó:** Clique duas vezes em "Add comment" abaixo do nó para adicionar um comentário/rótulo. **Portas dinâmicas:** Portas de entrada adicionais podem ser adicionadas clicando no sinal de mais no lado esquerdo do nó.

- Não configurado:** O nó ainda não foi configurado e não pode ser executado com as configurações atuais.
- Configurado:** O nó foi configurado corretamente e pode ser executado a qualquer momento.
- Executado:** Não foi executado com sucesso, os resultados podem ser visualizados e utilizados nos nós seguintes.
- Erro:** O nó encontrou um erro durante a execução.

Leitura

CSV Reader: Lê arquivos CSV. Possui uma função de detecção automática para interpretar a estrutura do arquivo. Como em outros nós de leitura, é possível adicionar uma porta de "Conexão ao sistema de arquivos" para conectar a diferentes fontes de dados.

Excel Reader: Lê o conteúdo de planilhas em arquivos Excel (XLS, XLSX). A planilha e as células a serem lidas podem ser definidas na janela de configuração.

Table Creator: Permite que os usuários criem manualmente uma tabela de dados em sua janela de configuração como uma planilha. Células de dados podem ser copiadas e coladas na planilha.

Model Reader: Lê modelos de aprendizado de máquina gerados com qualquer um dos nós de aprendizado. Os modelos geralmente são salvos após o treinamento e reutilizados na implementação.

Table Reader: Lê dados de um arquivo .table. Arquivos .table são organizados usando um formato proprietário do KNIME, incluindo toda a estrutura do arquivo, sendo otimizados para espaço e velocidade, com máximo desempenho!

Google Sheets Reader: Lê dados de um arquivo do Google Sheets. A autenticação ocorre no site do Google. As credenciais não são salvas dentro do fluxo de trabalho do KNIME.

Nos nós de leitura e escrita, o caminho do arquivo é expresso em relação a uma localização chave

Visualização

(Todas as visualizações são interativas)

Scatter Plot: Representa as linhas de dados de entrada como pontos em um gráfico bidimensional. As dimensões de entrada (colunas) no eixo x-y e as propriedades gráficas podem ser alteradas na janela de configuração ou interativamente na (Todas as visualizações são interativas) visualização do nó.

Sunburst Chart: Exibe colunas categóricas através de uma hierarquia de anéis. Cada anel é fatiado de acordo com os valores nominais na coluna correspondente e com a hierarquia selecionada. Este é um gráfico poderoso para análise multivariada.

Stacked Area Chart: Gráficos de dados numéricos em várias colunas sobrepostas, utilizando a linha anterior como referência base. As áreas entre linhas são coloridas para facilitar a comparação. Este tipo de gráfico é normalmente usado para visualizar tópicos em tendência.

Line Plot: Plota valores numéricos nas colunas de dados em relação aos valores de uma coluna de referência. Os pontos de dados são conectados por linhas coloridas. Se a coluna de referência no eixo x contiver valores de tempo ordenados, o gráfico de linhas representa graficamente a evolução de uma série temporal.

Color Manager: Atribui uma propriedade de cor a cada linha de entrada com base no valor da linha em uma coluna selecionada. Essa propriedade de cor afeta a representação gráfica nas visualizações subsequentes.

Pie Chart: Visualiza dados categóricos através de gráficos circulares, em que cada fatia representa uma categoria e o seu tamanho indica a respectiva proporção atribuída.

Data Explorer: Fornece uma visualização interativa para resumir as estatísticas dos dados de entrada por meio de medidas estatísticas e histogramas – tanto para colunas numéricas quanto para colunas nominais.

Box Plot: Visualiza colunas numéricas utilizando as estatísticas dos quartis. Fique atento aos pontos no final dos "bigodes" – eles podem indicar valores atípicos (outliers)!

Bar Chart: Visualiza uma ou mais métricas agregadas para diferentes partições de dados com barras retangulares, onde as alturas são proporcionais aos valores das métricas. As partições são definidas por uma coluna categórica.

Implantação

Report PDF Writer: Exporta um determinado relatório para um arquivo em formato PDF no local especificado. Em alternativa, o nó Report HTML Writer exporta o relatório para um arquivo em formato HTML.

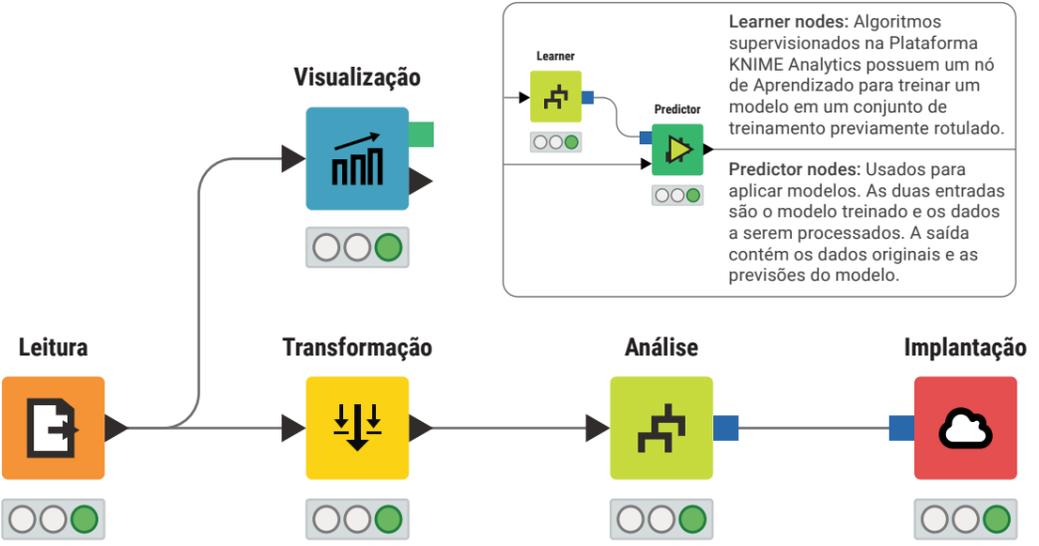
Excel Writer: Escreve a tabela de dados de entrada em uma planilha de um arquivo Excel (XLS ou XLSX).

Table Writer: Escreve a tabela de dados de entrada em um arquivo usando o formato proprietário .table do KNIME. Este formato inclui toda a estrutura do arquivo e é otimizado para espaço e velocidade. Incluir a estrutura da tabela no arquivo é uma grande vantagem – especialmente ao trocar arquivos de dados entre usuários.

CSV Writer: Escreve a tabela de dados de entrada em um arquivo CSV ou em um local remoto indicado por um URL.

Google Sheets Writer: Escreve a tabela de dados de entrada em um arquivo do Google Sheets. A autenticação ocorre no site do Google. As credenciais do Google não são salvas no fluxo de trabalho do KNIME.

Send to Tableau Server: Exporta a tabela de dados de entrada para um arquivo ou servidor Tableau para relatórios.



Transformação

GroupBy: Agrupa as linhas de uma tabela pelos valores únicos em colunas selecionadas e calcula medidas de agregação e estatísticas para os grupos definidos. Apesar de seu nome simples, oferece funcionalidades poderosas e tem muitos usos inesperados. Por exemplo, deduplicação de linhas.

Partitioning: Divide os dados em dois subconjuntos de acordo com uma estratégia de amostragem. Este nó é geralmente usado para produzir um conjunto de treinamento e um conjunto de teste para treinar e avaliar um modelo de aprendizado de máquina.

String to Date&Time: Converte valores em uma coluna de String em valores de Data&Hora. O formato de Data&Hora contido nos valores de String pode ser definido manualmente ou detectado automaticamente.

Column Renamer: Atribui novos nomes e tipos às colunas selecionadas, conforme configurado na janela de diálogo.

Concatenate: Mescla duas ou mais tabelas de dados verticalmente, empilhando células em colunas com o mesmo nome. Células em colunas que não se sobrepõem são preenchidas com valores ausentes.

Pivot: Expande a funcionalidade de agregação do nó GroupBy criando uma tabela de dados de saída com colunas e linhas para os valores únicos nas colunas de entrada selecionadas. Nota: os valores únicos da coluna de agrupamento tornam-se linhas.

Row Filter: Filtra linhas da tabela de dados de entrada de acordo com uma regra de filtragem. A regra de filtragem pode corresponder a um valor em uma coluna selecionada ou a números dentro de um intervalo numérico.

Cell Splitter: Divide os valores de uma coluna selecionada em duas ou mais substrings, conforme definido por um delimitador. O delimitador é um caractere definido, como vírgula, espaço ou qualquer outro caractere ou sequência de caracteres.

Joiner: Junta linhas de duas tabelas de dados com base em valores comuns em uma ou mais colunas-chave. A saída pode ser uma junção interna, junção externa à esquerda, junção externa à direita, junção externa completa ou os respectivos antijunções, podendo ser dividida em várias tabelas de saída.

Missing Value: Define uma estratégia para lidar com valores ausentes na tabela de dados de entrada – seja globalmente em todas as colunas ou individualmente para cada coluna.

Column Merger: Compara os valores de duas colunas com base numa coluna principal e numa coluna secundária definidas antemão. O nó gera uma nova coluna onde o valor de cada linha será o da coluna principal, se este não estiver em falta; caso contrário, será utilizado o valor da coluna secundária.

Column Filter: Filtra colunas da tabela de dados de entrada de acordo com uma regra de filtragem. As colunas a serem mantidas podem ser selecionadas manualmente ou de acordo com seu tipo, ou com uma expressão regex que corresponda ao nome da coluna.

Sorter: Ordena a tabela em ordem crescente ou decrescente com base nos valores de uma coluna escolhida. Além disso, é possível ordenar com base em várias colunas.

String Cleaner: Disponibiliza operações para limpar e pré-processar dados em formato de texto, como remover caracteres indesejados, eliminar espaços em branco, converter para minúsculas, entre outros. Para operações mais avançadas, recomenda-se o uso do nó Expression.

Análise

Decision Tree Learner: O nó de Aprendizado treina uma árvore de decisão C4.5 ou CART. A janela de configuração inclui opções para poda, parada precoce, medidas de informação, valores de divisão e mais. Tanto o nó de Aprendizado quanto o de Previsão fornecem uma visualização interativa onde a árvore de decisão é exibida junto com a propagação dos dados de entrada.

K-Means: O número de clusters deve ser definido antes da execução do nó. Este nó constrói os clusters. O nó de Cluster Assigner encontra o cluster mais próximo e o atribui à linha de dados de entrada. Sendo um algoritmo não supervisionado, este par de nós não segue o esquema clássico Learner - Predictor.

Logistic Regression Learner: O nó de Aprendizado treina um modelo de regressão logística para prever valores-alvo categóricos. A janela de configuração inclui opções para o solucionador, escolha de características de entrada, funções de regularização para evitar overfitting, e mais.

Scorer: Calcula várias medidas de desempenho, como acurácia, F1-score, ou Kappa de Cohen, para quantificar a qualidade de um classificador.

Numeric Scorer: Calcula várias medidas de erro numérico, como erro quadrático médio, erro absoluto médio, ou R², para quantificar a qualidade de um modelo preditivo numérico.

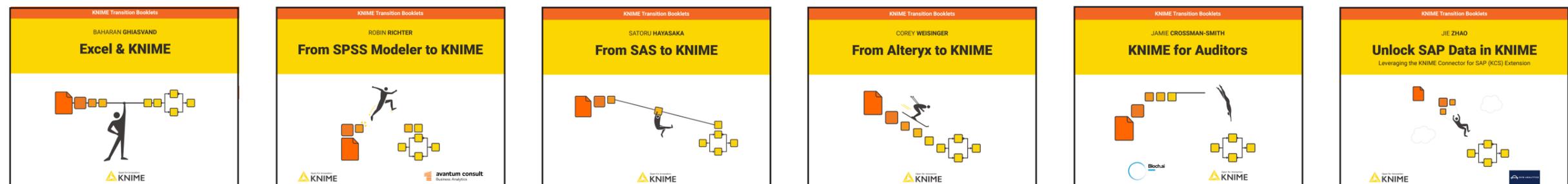
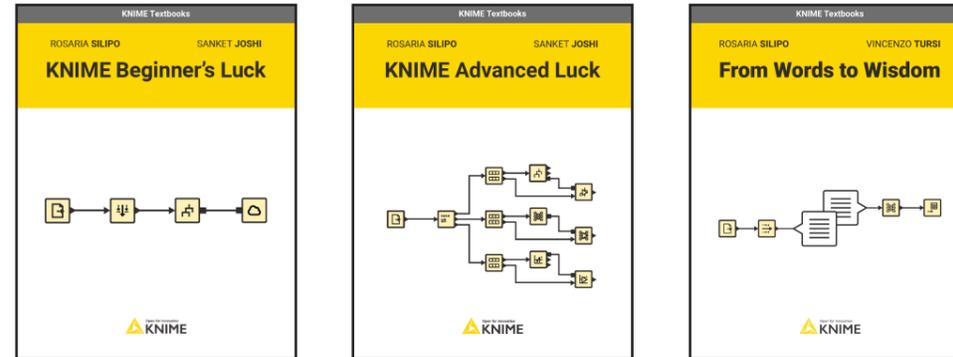
ROC Curve: Exibe a curva ROC (Receiver Operating Characteristic) de um classificador trabalhando em um problema de classe binária. Uma das duas classes é arbitrariamente escolhida como a classe positiva, e a curva ROC é construída com base nas probabilidades/escores produzidos para essa classe no conjunto de dados de entrada.

Integrações com muitas ferramentas de análise de dados open source também estão disponíveis. Algumas utilizam a interface GUI dos nós do KNIME (H2O, Weka, Keras, Spark MLlib). Outras oferecem nós com um ambiente de desenvolvimento para scripting e depuração (R, Python, Java).

Resources

- **KNIME Press:** Acesse diversos livros de ciência de dados e outros guias rápidos em knime.com/knimepress, incluindo tópicos para iniciantes e avançados.
- **KNIME blog:** Tópicos envolventes, desafios, notícias da indústria e nuggets de conhecimento em knime.com/blog.
- **Self-paced courses:** Faça nossos cursos online gratuitos e autoguiados para aprender sobre análise de dados, engenharia de dados ou ciência de dados com KNIME (com exercícios práticos) em knime.com/learning.
- **KNIME Community Hub:** Armazena, cria versões, automatiza e permite colaborar em fluxos de trabalho privados, ou explora e partilha fluxos de trabalho públicos com a KNIME Community em hub.knime.com.
- **KNIME Forum:** Junte-se à nossa comunidade global e participe de conversas em forum.knime.com.
- **KNIME Business Hub:** Para colaboração em equipe, automação, gerenciamento e implantação, confira o KNIME Business Hub em knime.com/knime-business-hub.

Amplie seus conhecimentos sobre o KNIME com nossa coleção de livros da KNIME Press. Para usuários iniciantes e avançados, até aqueles interessados em tópicos específicos como detecção de tópicos, mistura de dados e soluções clássicas para casos de uso comuns usando a Plataforma KNIME Analytics - há algo para todos. Disponível para download em www.knime.com/knimepress.



**Necessita de ajuda?
Entre em contato!**

