PROJETO TÉCNICO

Modelo automatizado para gestão de indicadores de estoque com uso do paradigma ETL e inteligência artificial

Júlio César Bastos de Figueiredo

Ago/2021

Resumo

O objetivo desta pesquisa foi constituir um modelo baseado no paradigma ETL (Extract, Transform & Load) para permitir aos gestores da área de operações terem um melhor entendimento de como automatizar o processo decisório de gestão de estoques dentro do conceito RTBI (Real Time Business intelligence). Para isso, foram analisadas informações bases de dados públicas de dados transacionais de estoque e utilizadas técnicas de Mineração de Dados (Data Mining) para descoberta de estruturas latentes nos dados. Também foram aplicadas técnicas de detecção de tendências e de regras ou padrões que pudessem ser utilizados como suporte dos processos de gerenciamento de estoques. A metodologia utilizada para implantação dos modelos a CRISP-DM (CRoss-Industry Standard Process for Data Mining), que contempla os passos para o planejamento, organização, implementação e documentação de um projeto de mineração de dados. Para a construção do modelo, foi utilizada a plataforma KNIME (www.knime.com). Como resultado, mostra-se que sistemas automatizados para tratamento em tempo real de dados de estoque com suporte a métodos de inteligência artificial possuem potencial para melhoria da assertividade e da eficiência nos processos de gerenciamento de estoques.

1. Introdução

Empresas que buscam boas margens de lucros e visam conquistar a confiança de seus clientes devem gerenciar de forma eficiente suas cadeias. Dentro deste contexto, a questão dos estoques tem destaque. Evitar a escassez de produtos é um dos princípios básicos de uma boa gestão da cadeia, ou seja, o nível de serviços tem de ser mantido. Contudo, uma gestão eficiente de estoques também considera custos de estocagem. Neste sentido as empresas de varejo precisam otimizar seu gerenciamento, uma vez que podem lidar com grandes volumes

de produtos, fazendo com que seus custos de estocagem representem uma parcela significativa dos seus custos de operação.

Com o crescimento do setor de varejo no ano de 2020, principalmente pela expansão de vendas via e-commerce durante a pandemia do COVID-19, foi possível observar que um processo simplificado de gerenciamento de estoques, otimizado, confiável e de rápida resposta, pode tornar-se uma vantagem competitiva para a empresa. Isso pode ser obtido através das tecnologias hoje existentes.

Nesta pesquisa busca-se evidenciar os benefícios de uma maior utilização de mineração de dados e sistemas de classificação supervisionada baseados em inteligência artificial. A ideia geral é permitir que se possa classificar de maneira rápida as melhores estratégias e ações a serem adotadas nas abordagens de gerenciamento rápido de estoques, ao mesmo tempo em que se desenvolvem novos insights a respeito de sua utilização.

1.1 A transformação Digital e a Busca pelas Capacidades Analíticas

A transformação digital nas empresas não começou com a crise gerada pela COVID-19. Muitas empresas viram na tecnologia uma oportunidade para melhorar seu desempenho, tornar-se mais competitiva, ou ainda garantir melhores resultados. Para outras empresas, a tecnologia foi uma porta de entrada para facilitar seus ingressos em mercados que antes eram muito consolidados, como é o caso de startups.

Na área de cadeia de suprimentos, a introdução das tecnologias digitais resultou em uma mudança na forma como os profissionais da área encaram as operações da empresa. Em algumas áreas, a tecnologia passou a assumir algumas das funções exclusivas de profissionais de SCM, em outras, ela complementou estas funções e em outras ainda, persistem funções-chave diretamente controladas por pessoas A questão é que, de modo geral, vê-se de uma transformação na forma de gerar entrega de valor, que abrange pilares técnicos, e humanos.

O principal benefício do uso de técnicas analíticas em SCM foi o a otimização. Tanto da gestão como dos processos de tomada de decisão. Soluções analíticas fornecem vantagens como confiabilidade, automatização e simplicidade.

O custo anual de manter uma única unidade de estoque pode variar de 15% a 35% do valor do produto (Timme e Williams-Timme 2003). Portanto, o sucesso da empresa em um mercado

competitivo depende da sua capacidade de controlar e planejar o estoque a um custo mínimo (Hockey 2009) mantendo satisfatório o nível de serviço ao cliente.

Esta conciliação pode ser atingida por meio de um melhor uso das informações em tempo real, tanto de demandas esperadas como produção e vendas efetivas. A combinação destas informações permite a identificação de padrões de resposta eficientes, que geram novas informações e insights sobre como melhorar a gestão interna dos estoques.

É neste contexto de identificação de padrões de resposta, que permitem uma maior eficiência na tomada de decisão que as tecnologias de Inteligência Artificial ganham destaque. Por exemplo, um exemplo do uso de Inteligência Artificial para previsões precisas e dinâmicas é um estudo realizado para prever o número de casos de COVID-19 na China, utilizando dados do período de 11 de janeiro de 2020 até 27 de fevereiro do mesmo ano. Por meio de uma análise com uso de redes neurais foi possível agrupar províncias para investigar a estrutura de transmissão. Com isto obtive-se uma previsão de trajetória da doença na China com baixos índices de erro, o que permitiu auxiliar a logística de saúde e as políticas públicas.

Em resumo, o uso de sistemas analíticos e de inteligência artificial para gestão de estoque pode simplificar os processos aumentando a precisão, auxiliando na tomada de decisão de curto, médio e longo prazos. Com é possível direcionar ações da empresa acerca de novas oportunidades de negócio minimizar custos relacionados à gestão de estoque, mas mantendo em mente a manutenção nível de serviço ao cliente.

2. Metodologia

A metodologia utilizada para desenvolvimento técnico do protótipo de extração, transformação e entrega de informações analíticas de estoque feito nesta pesquisa foi a CRISP-DM (*CRoss-Industry Standard Process for Data Mining*), que contempla seis passos que determinam o planejamento, organização, implementação e documentação de um projeto de mineração de dados. O CRISP-DM foi oficialmente lançado em 2000, por um consórcio de quatro empresas, sendo elas: SPSS (um provedor de soluções comerciais de DM), NCR (um provedor de banco de dados), Daimler Chrysler (fabricante de automóveis) e OHRA (uma seguradora). O framework pode ser aplicado independente da indústria ou das tecnologias utilizadas no projeto (Wirth & Hipp, 2000). Na Figura 1, vemos as 6 etapas da metodologia CRISP DM, publicada por Wirth & Hipp (2000).

PROJETO TÉCNICO

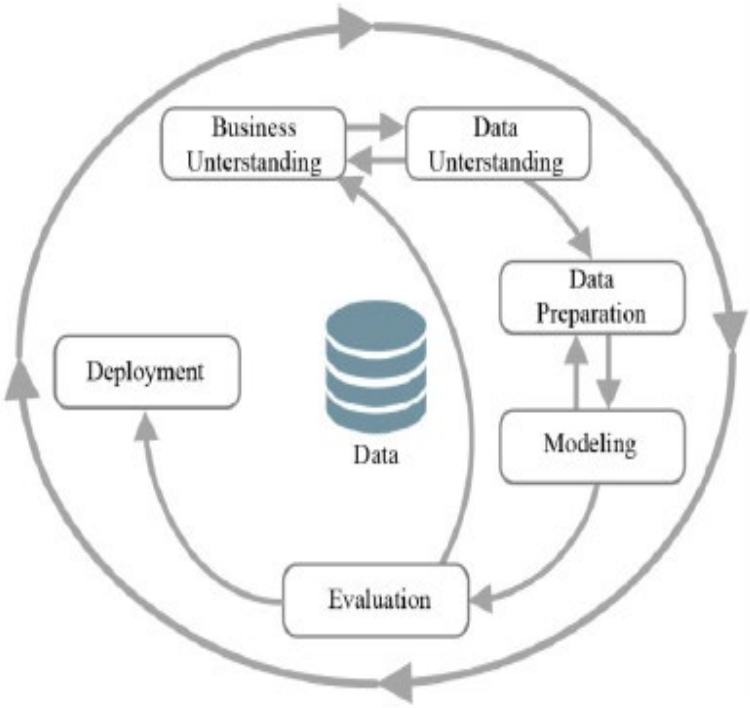


Figura 1: Fases do modelo CRISP-DM. Fonte: Wirth & Hipp (2000)

A metodologia CRISP-DM tem como objetivo tornar os projetos de mineração e tratamento de dados reproduzíveis, gerenciáveis e rápidos. Ela possibilita uma visualização abrangente do processo de descoberta de conhecimento, que vai além de aplicações estatísticas e algoritmos (Huber et al., 2019).

O processo especializado que prescreve as etapas a serem seguidas, oferece direcionamentos para cada tarefa e auxilia na estruturação dos projetos. Além de servirem como guia e garantir que nada seja esquecido, essas definições trazem maior clareza sobre os resultados esperados ao término do projeto (Wirth & Hipp, 2000).

2.1 Passos Metodológicos

A partir do framework CRISP-DM proposto por Wirth & Hipp (2000), esta pesquisa seguiu os passos metodológicos descritos abaixo.

2.1.1 Compreensão do Problema

O primeiro passo trata-se da compreensão do problema. Nesta fase inicial, são identificados os objetivos e requisitos do projeto a partir da perspectiva dos negócios. Em seguida, esse conhecimento é convertido em um problema de mineração de dados. No caso específico desta pesquisa, o problema é como garantir em tempo real a obtenção de respostas de ação quanto ao gerenciamento de estoques.

2.1.2 Exploração dos Dados

No segundo passo será realizada uma exploração sobre os dados disponíveis. A partir de uma coleta inicial, são realizadas atividades que possibilitam um maior entendimento sobre os dados que serão utilizados. Nesta etapa, a qualidade dos dados será avaliada e serão gerados os primeiros *insights,* onde serão detectados eventuais subconjuntos para a formação de hipóteses.

Nesta pesquisa foram utilizados os dados de uma base anonimizada de transações de compra disponível no hub de exemplos do KNIME (api.hub.knime.com). Esta base de dados não determina o tipo de produto registrado nas transações. Os campos de dados serão descritos na seção sobre o desenvolvemos. O acesso ao arquivo de dados utilizado (**transacional.xlsx**) pode ser encontrado, juntamente com os fluxos de dados, no link:

http://technical.paper.nom.br/ETL_estoques/

2.1.3 Organização dos Dados

O terceiro passo trata da organização dos dados. Durante a fase de organização são realizadas as atividades para a construção do conjunto final de dados que serão utilizados na etapa seguinte de modelagem. Essas atividades incluem a montagem das tabelas, seleção dos atributos, limpeza e transformação dos dados. Elas podem ser realizadas diversas vezes e sem uma ordem prescrita. As atividades desta etapa neste projeto estão detalhadas na seção de implementação.

2.1.4 Modelagem

No quarto passo compete a definição das técnicas de modelagem que serão utilizadas, visto que várias técnicas e algoritmos podem ser aplicados para o mesmo tipo de problema de mineração de dados. Esta etapa possui uma ligação estreita com a etapa de preparação dos dados, pois algumas técnicas requerem formatos de dados específicos e frequentemente são percebidos problemas ou ideias para a construção de novos conjuntos.

Em particular, neste projeto, serão usadas especificamente duas técnicas, um método de projeção autoagressivo conhecido por ARIMA (Autoregressive Integrated Moving Average) e uma técnica de classificação baseada em redes neuras de multicamadas – MLNP (Multilayer Neural Network in Parallel-configuration)

2.1.5 Avaliação do Modelo

No quinto passo, é realizada uma avaliação sobre a qualidade das informações obtidas quanto à sua pertinência em relação aos objetivos. Nesta fase, o modelo como um todo será avaliado e revisado para garantir que os objetivos de negócio são atendidos e se alguma questão importante não foi considerada. Ao final desta etapa, será estabelecido a forma de uso dos resultados da mineração dos dados.

2.1.6 Disseminação e Uso

Por fim, no sexto passo será definido como o conhecimento adquirido será organizado e disseminado, assim como a forma como o modelo poderá ser utilizado pelos usuários finais.

Embora a descrição das fases acima represente uma sequência de eventos, na prática, as tarefas podem ser realizadas na ordem necessária e eventualmente será preciso retomar tarefas anteriores e repetir determinadas ações em ciclos contínuos de melhoria dos modelos.

3. Construção do Modelo

A figura 2 abaixo apresenta o modelo KNIME desenvolvido neste projeto. O KNIME Analytics Platform (www.knime.com) é o software gratuito de código aberto para trabalhar fluxos de dados por meio de programação visual. A programação visual é frequentemente confundida com como sendo uma programação "sem código", onde o objetivo é permitir que pessoas que nunca tiveram acesso à aprendizado de programação possam codificar.

Na verdade, os sistemas de programação visual foram criados para abstrair todas as coisas com as quais um cientista de dados não deveria se preocupar durante o processo de modelagem (formação, diagramação, sintaxe de parametrização, etc.) para que possa se preocupar especificamente em fazer ciência de dados impactante.

Além disto, criar fluxos de trabalho visuais fornece uma maneira altamente eficiente de expressar a lógica de um programa de maneira mais compreensível do que técnicas tradicionais.

Conforme já mencionado, todos os dados e os fluxos de programação desenvolvidos neste projeto, estão disponibilizados no endereço http://technical.paper.nom.br/ETL_estoques/.

PROJETO TÉCNICO

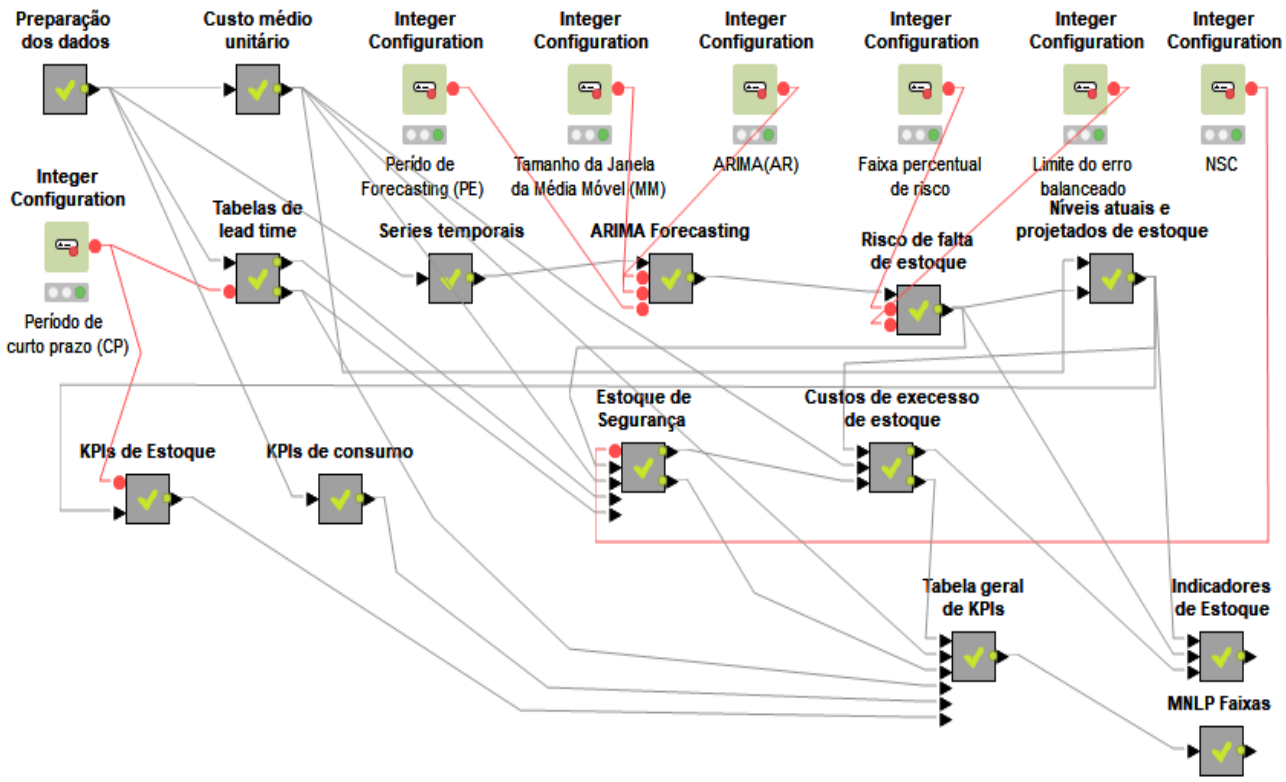


Figura 2: Fluxo KNIME do modelo desenvolvido no projeto

A seguir serão apresentadas e detalhadas as etapas de construção do modelo. Serão discutidas as escolhas feitas durante o processo de modelagem e os resultados obtidos em cada etapa.

3.1 Descrição dos Dados

Conforme o que já foi apresentado na seção de metodologia, nesta pesquisa foram utilizados os dados de uma base anonimizada de transações de compra disponível no hub de exemplos do KNIME (api.hub.knime.com).

Optou-se por fazer um acesso de leitura a uma base estática representada por um arquivo Excel, todavia, o sistema permite a substituição deste acesso estético por um acesso dinâmico a bases de dados online, por exemplo, de um ERP. O KNIME possui bibliotecas específicas para isto. Bastaria substituto bloco de acesso ao arquivo Excel por um bloco de acesso a um drive de banco de dados.

A tabela 1 apresenta a descrição dos dados do arquivo **transacional.xlsx**.

Tabela 1: Descrição dos dados do arquivo transacional.xls

PartID : Código de identificação do produto

Cost : Custo unitário do produto

tr_effdate : Data em que a transação (de compra ou venda) foi efetivada

po_ord_date : Data de emissão de ordem de compra (purchase order)

wo_ord_date : Data de emissão de ordem de produção (work order)

Consuption : volume vendido

Supply : Volume comprado

	Column Type
PartId	String
Cost	Number (double)
tr_effdate	Local Date
po_ord_date	Local Date
wo_ord_date	Local Date
Consumption	Number (integer)
Supply	Number (integer)

O arquivo **transacional.xlsx** contém 20.132 transações de 64 produtos conforme apresentado na tabela 2:

Tabela 2: Estrutura dos dados do arquivo transacional.xls

1	PartId	Cost	tr_effdate	po_ord_date	wo_ord_date	Consumption	Supply
2	4683-001-209	17,8514	2009-01-05			2	
3	4683-019-219	38,9683	2009-01-05			2	
4	4683-001-209	17,8514	2009-01-05			2	
5	4683-019-219	38,9683	2009-01-05			2	
⋮							
20129	C24547-135-002	447,529311	2016-08-05	2016-05-16			5
20130	C24547-135-002	447,529311	2016-08-05	2016-06-08			5
20131	A56377-065-001	337,8103701	2016-08-08	2016-07-08			5
20132	C24886-002ae6937	1660,756607	2016-08-08		2016-07-18		4

3.2 Leitura dos Dados

Nesta etapa de construção do fluxo os dados são lidos da tabela de dados. Em seguida são criadas duas colunas adicionais na tabela de dado transacionais para facilidade posterior do processamento. É criada a coluna TrDate, que contém a data da transação, que pode ser uma data de venda, produção ou compra, e uma coluna YearMonth que registra o ano e mês da transação para facilitar a agregação mensal dos dados em outras etapas do fluxo, pois a

maioria das agregações será mensal. O fluxo de leitura de dados com estas etapas é então encapsulado em um meta-bloco definido como “Preparação de dados”. A figura 3 ilustra estas etapas.

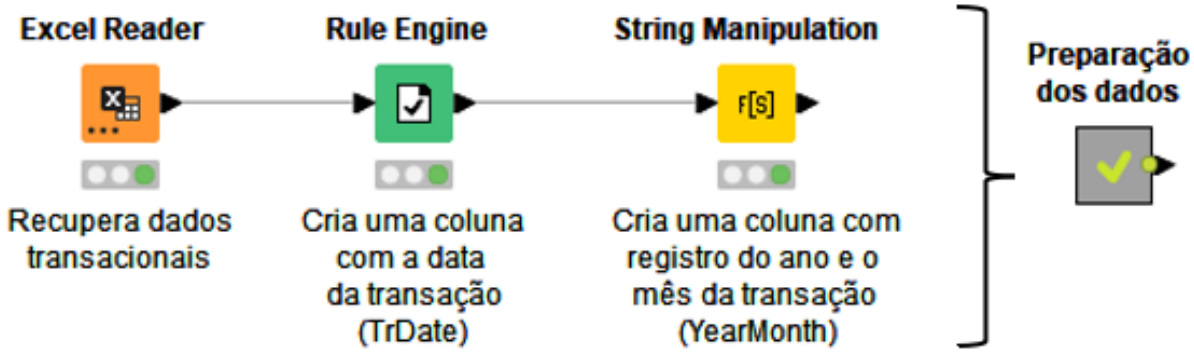


Figura 3: Fluxo KNIME de preparação de dados.

A tabela 3 mostra a saída do meta-bloco de preparação de dados. Note que os dados correspondem aos dados da tabela transacional acrescidos das colunas adicionais. A ideia é que o fluxo de processamento vai modificando a tabela transformando as informações de acordo com as necessidades. Este tipo de processamento pode ser feito em tempo real, bastando para isto definir um tempo de atualização para o fluxo.

Tabela 3: Estrutura dos dados da saída do meta-bloco “Preparação de dados”

Row ID	PartId	Cost	tr_effd...	po_ord...	wo_ord...	Consu...	Supply	TrDate	YearMo...
Row0	4683-001-209	17.851	2009-01-05	?	?	2	?	2009-01-05	2009-01
Row1	4683-019-219	38.968	2009-01-05	?	?	2	?	2009-01-05	2009-01
Row2	4683-001-209	17.851	2009-01-05	?	?	2	?	2009-01-05	2009-01
Row3	4683-019-219	38.968	2009-01-05	?	?	2	?	2009-01-05	2009-01
Row4	4683-001-209	17.851	2009-01-05	?	?	2	?	2009-01-05	2009-01
Row5	4683-019-219	38.968	2009-01-05	?	?	2	?	2009-01-05	2009-01
Row6	-60878s102	6.044	2009-01-05	?	?	3	?	2009-01-05	2009-01

3.3 Cálculo dos Lead Times

3.3.1 Lead time histórico

A saída do fluxo de preparação dos dados é utilizada para calcular os lead times das ordens de produção e compra. Para isto é necessário fazer uma agregação mensal dos dados de venda para cada produto. A data de transação que será associada a um determinado mês é a máxima de transação realizada naquele mês. O lead time estimado em um determinado mês para um

produto (PartID) será dado pela média dos lead times (diferença entre a emissão e recebimento da ordem) das transações no mês. O fluxo de dados, representado na figura 4, é então encapsulado no meta-bloco “Calcula lead time histórico”. Este nome se deve ao fato que estão sendo calculados os lead times para todos os períodos da tabela. Em certas situações os cálculos de médias considerarão apenas períodos mais recentes.

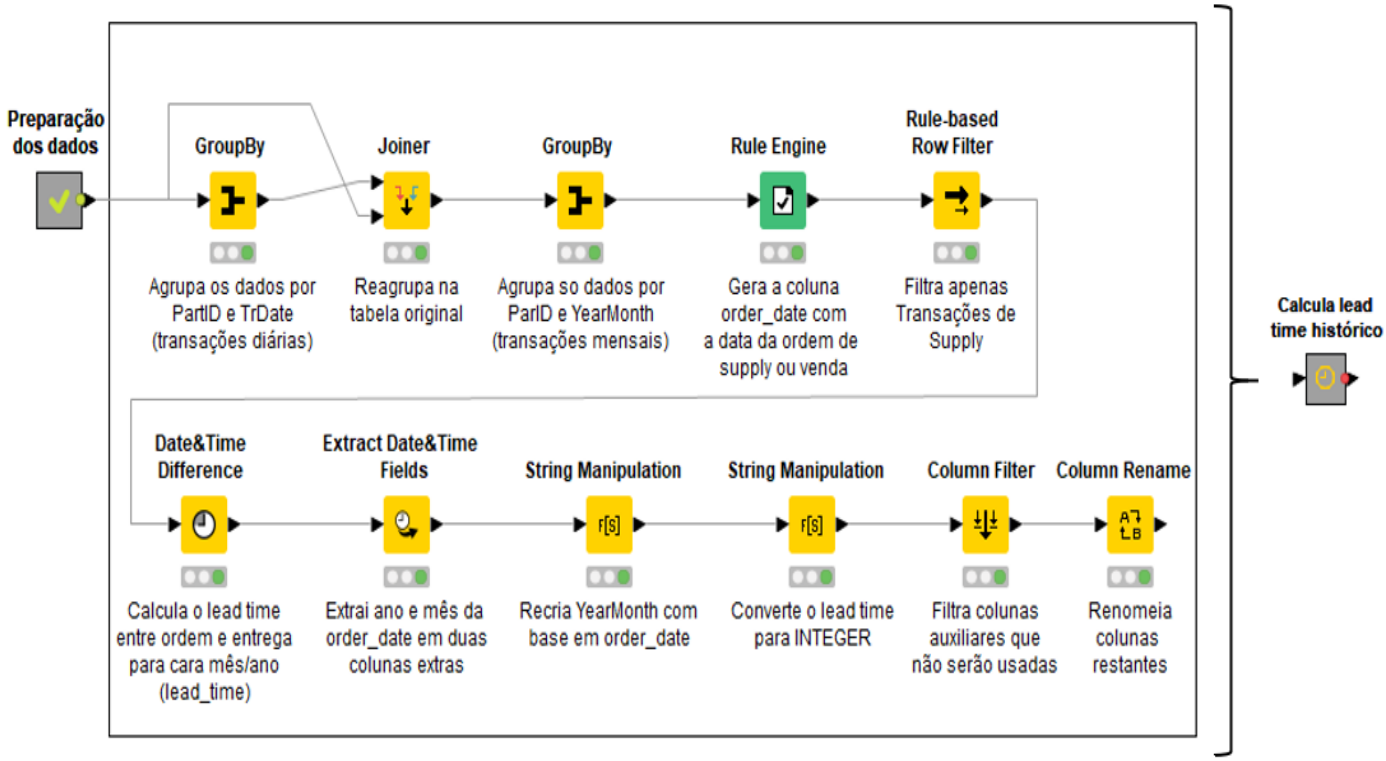


Figura 4: Fluxo KNIME para cálculo dos lead times históricos.

A tabela 4 apresenta a estrutura dos dados da saída do meta-bloco “Calcula lead time histórico” para o produto PartID=-06766-011. Note que para cada mês e ano é fornecido o lead time em dias. Há uma coluna com o cálculo do lead time médio histórico do produto e seu desvio padrão. Tais informações serão posteriormente utilizadas para as estimativas dos estoques de segurança.

Tabela 4: Estrutura dos dados da saída do meta-bloco “Calcula lead time histórico”

S PartId	D Mean(l...	D Standa...	I lead_time	S YearMo...
-06766-011	48.5	73.553	24	2009-1
-06766-011	48.5	73.553	2	2009-3
-06766-011	48.5	73.553	15	2010-2
-06766-011	48.5	73.553	11	2010-5
-06766-011	48.5	73.553	2	2011-6
-06766-011	48.5	73.553	1	2012-1
-06766-011	48.5	73.553	179	2013-5
-06766-011	48.5	73.553	154	2014-11

3.3.2 Lead time short (curto prazo)

A partir do fluxo da figura 4 é construído um outro fluxo onde são calculados os lead times apenas dos meses mais recentes, que são importantes pra capturar riscos de curto prazo. Neste fluxo, o parâmetro "Período de curto prazo (CP)" indica o total de meses passados que serão considerados dados de curto prazo. Este valor, nas simulações feitas, foi configurado para 36 meses e pode ser ajustado para outras simulações. O fluxo de dados construído é encapsulado no meta-bloco "Calcula lead time short" (figura 5).

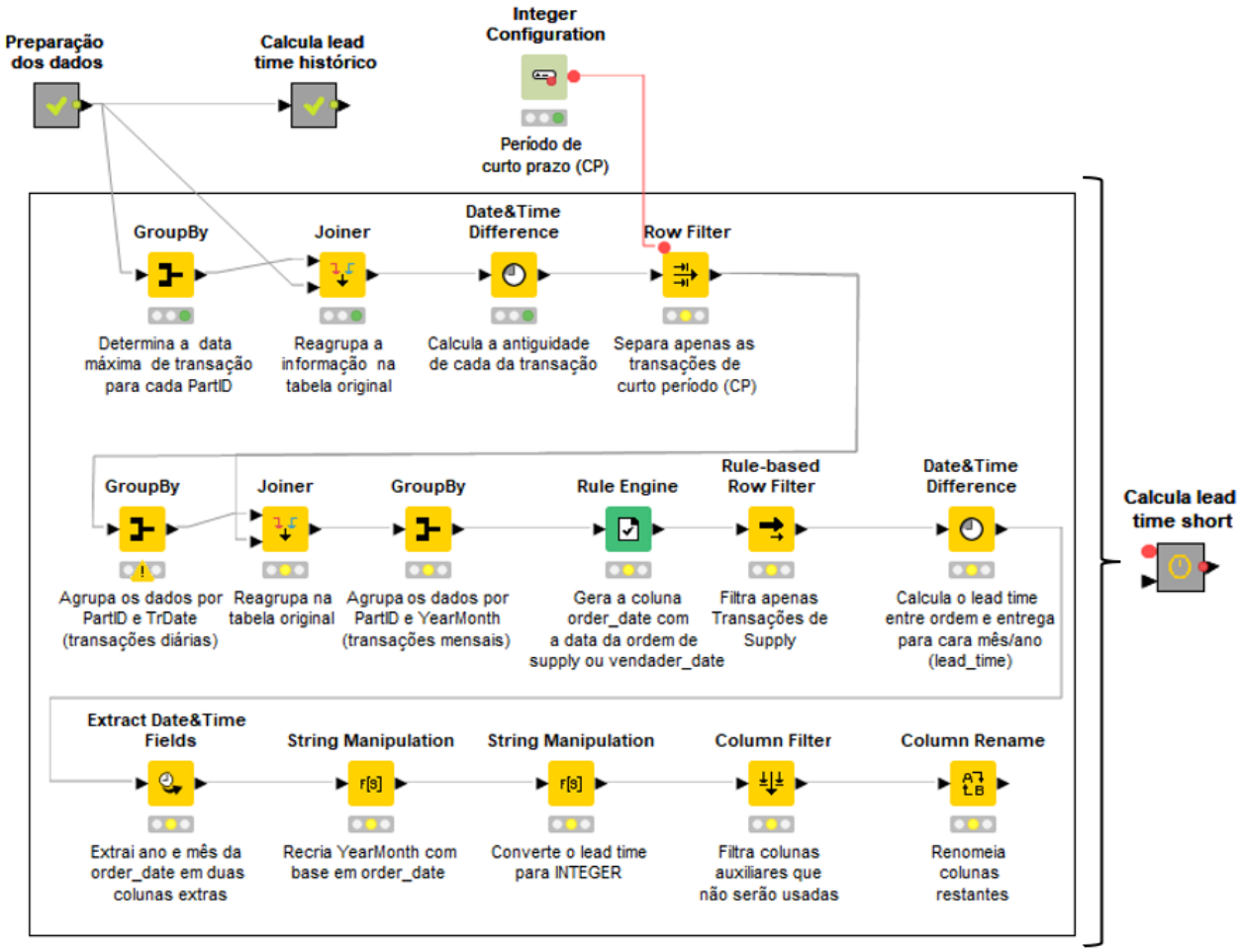


Figura 5: Fluxo KNIME para cálculo dos lead times de curto prazo.

Por fim, os lead times médio de Supply, histórico e de curto prazo, são agregados em duas tabelas, uma contendo os lead times mensais e outra contendo apenas o lead time médio (histórico e de curto praz) de cada produto. Cada uma destas tabelas será utilizada em um contexto diferente. A figura 6 representa o fluxo de criação das duas tabelas de lead time. O fluxo de dados é encapsulado no meta-bloco "Tabelas de lead time" que possui duas saídas, uma com os dados mensais de lead time d cada produto, e outra com os dados médios.

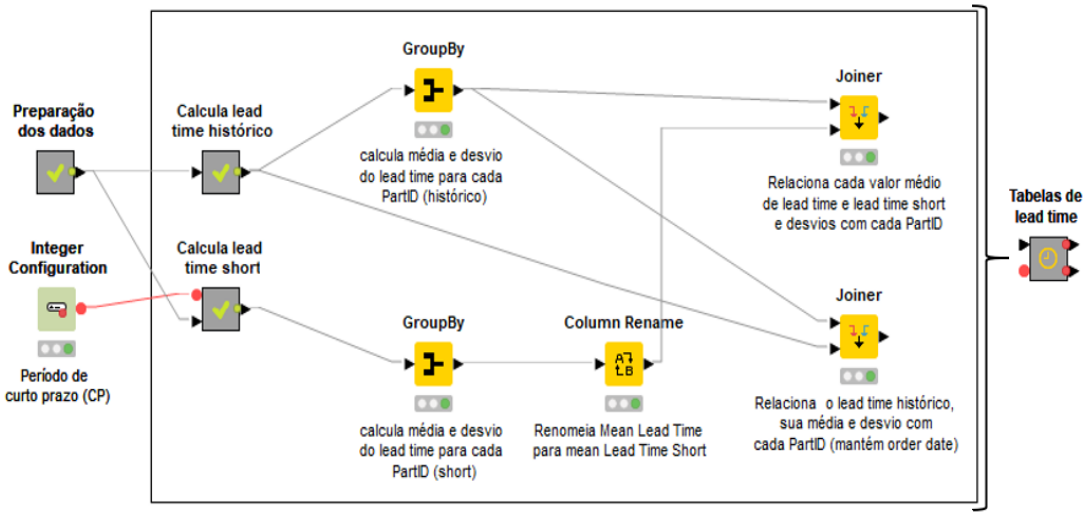


Figura 6: Fluxo KNIME para agregação das tabelas de lead time.

3.4 Cálculo dos Custos Médios

Para que possam ser feitas estimativas dos custos de estoque, é necessário criar um fluxo para cálculo do custo médio unitário de cada produto. Isto é feito de acordo com o fluxo apresentado na figura 7. É feita uma agregação de dados para cada PartID e em cada agregação é calculado o custo médio considerando todos as transações de compra.

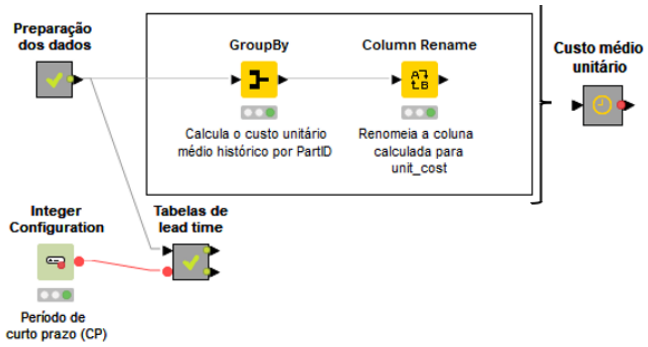


Figura 7: Fluxo KNIME para cálculo dos custo médio unitário de cada PartID.

3.5 Séries Temporais

Um aspecto importante do gerenciamento de estoque é a capacidade de estimar os riscos de quebra e excesso de estoque. Para isto é necessário ter capacidade de prever em certa medida a demanda para que, com base nesta previsão, estimativas de falta e quebra possam ser estimadas. No fluxo deste projeto é necessário construir, para cada PartID uma série temporal mensal do consumo. Como os dados de vendas de um determinado produto podem ser incompletos, os períodos faltantes deve ser interpolados para criação da série temporal que

servirá de base para o modelo predito. O fluxo que realiza esta tarefa está representado na figura 8. Os dados do fluxo são então encapsulados no meta-bloco “Séries Temporais”.

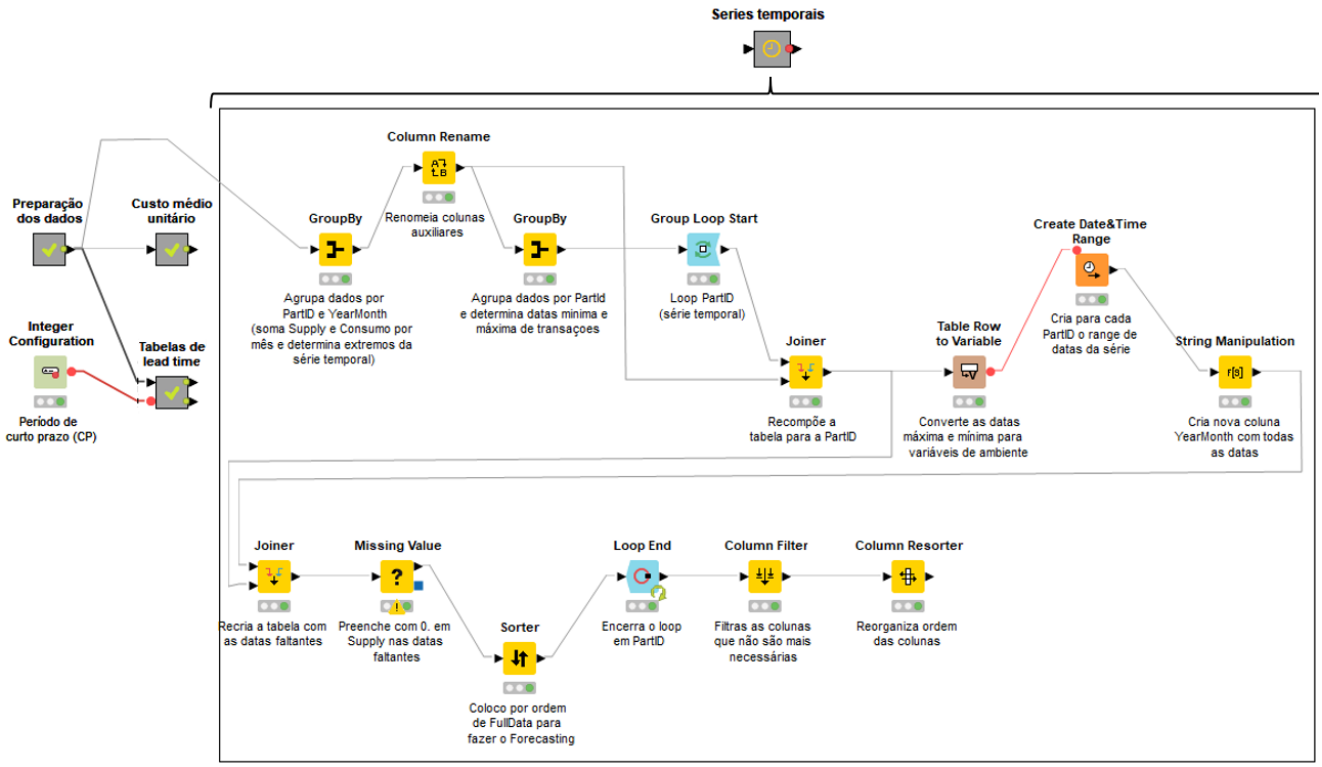


Figura 8: Fluxo KNIME para construção das séries temporais históricas de cada produto.

A tabela 5 mostra a saída do meta-bloco “Séries Temporais”. São criadas duas colunas com o total de vendas (Consumption) e o tal de compras (Supply):

Tabela 5: Estrutura dos dados da saída do meta-bloco “Séries Temporais”

S PartId	S YearMo...	Full...	D Sum(Consumption...	D Sum(Supply) ...
-06766-011	2009-01	2009-01-06	34	40
-06766-011	2009-02	2009-02-06	8	0
-06766-011	2009-03	2009-03-06	12	54
-06766-011	2009-04	2009-04-06	8	0
-06766-011	2009-05	2009-05-06	4	0
-06766-011	2009-06	2009-06-06	16	0
-06766-011	2009-07	2009-07-06	83	0
-06766-011	2009-08	2009-08-06	4	0

3.6 Forecasting (ARIMA)

As séries temporais históricas são em seguida processadas por um fluxo de previsão (Forecasting). Como dito antes, esta previsão é necessário para que se possam fazer estimativas dos riscos de desabastecimento e excesso de estoque, bem como para avaliar a

capacidade do estoque atual em atender a demanda. Neste projeto em particular, optou-se por utilizar um modelo autorregressivo de médias móveis (ARIMA - Autoregressive Integrated Moving Average). Os parâmetros do modelo ARIMA são configurados por meio de variáveis de ambiente: MM=3, AR=6 e PE=12 (Obs: PE corresponde ao período de forecasting).

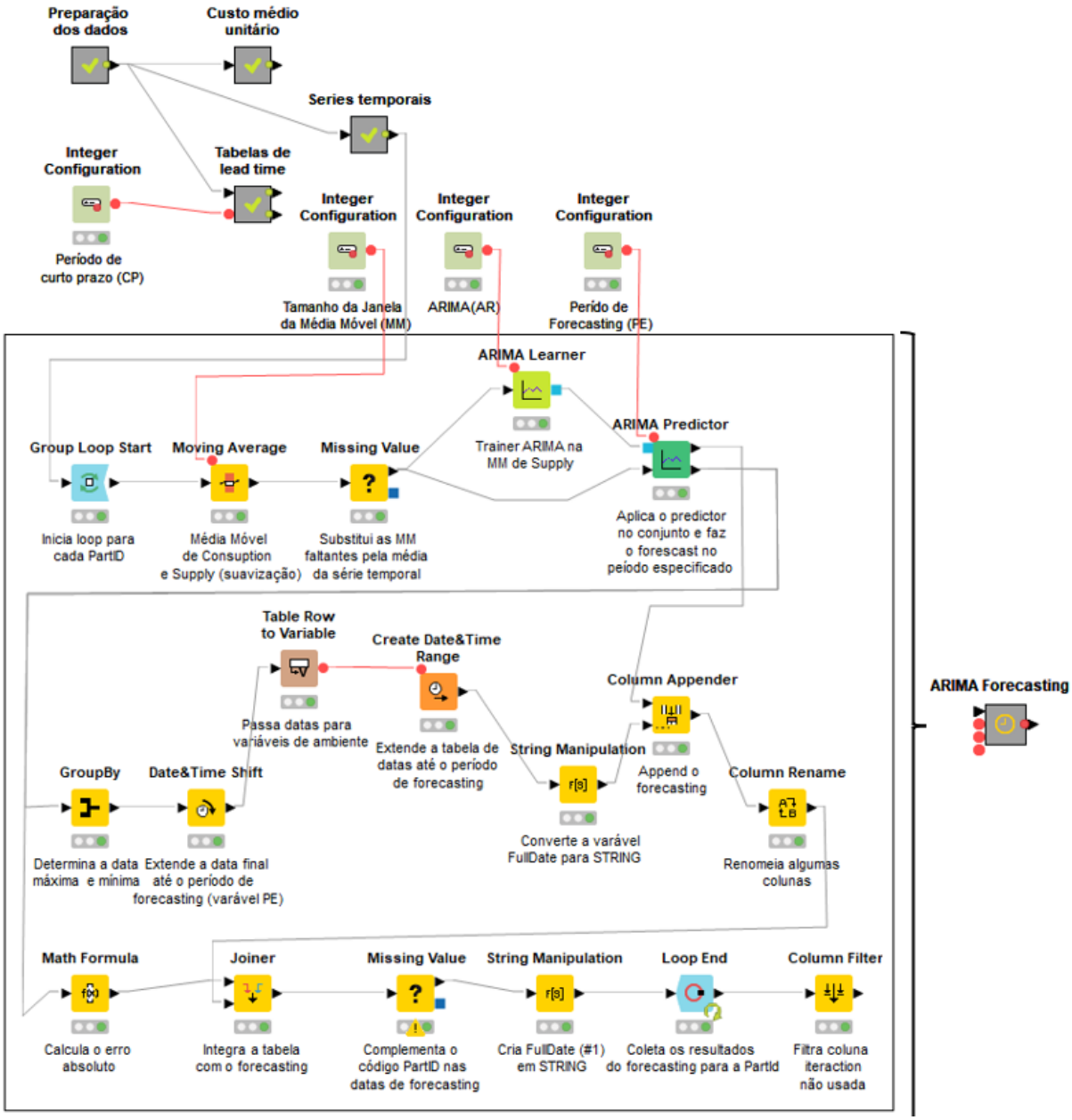


Figura 9: Fluxo KNIME para construção das séries temporais históricas de cada produto.

O fluxo da figura 9 é encapsulado no meta-bloco “ARIMA Forecasting”. A figura 10 apresenta uma das séries temporais de forecasing gerada pelo fuxo ARIMA. O período de forecasting foi PE=12 meses.

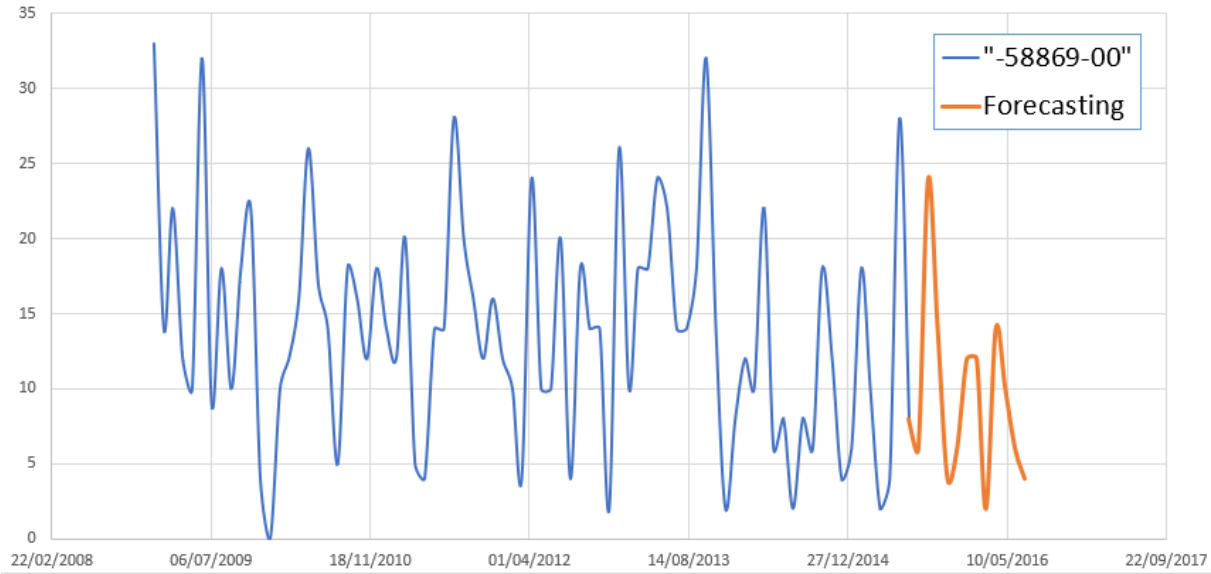


Figura 10: Fluxo KNIME para construção das séries temporais históricas de cada produto.

3.7 Risco de Falta de Estoque

Com base no fluxo de forecasting de cada produto e os dados médios de consumo é possível estimar o risco de falta de estoque de cada um dos produtos. Calculam-se então os erros absolutos de previsão e define-se uma faixa de tolerância para estes erros, dada pelo parâmetro “Faixa Percentual de Risco. Valores de forecasting abaixo desta faixa representam um "Over-stock risk" e valores de forecasting abaixo desta faixa um "Low stock risk". Valores de forecasting dentro da faixa representam uma situação “Ótima”. Ou outro parâmetro (“Limite de erro balanceado”) representa o limite da razão entre a soma dos erros absolutos de previsão de cada produto e o volume de previsão. Ele é usado como corte para que não se classifique o risco de produtos com alto índice de erro na previsão. Neste trabalho não se classificou o risco de produtos que apresentassem mais de 10% de erro balanceado na previsão ARIMA.

A figura 11 corresponde ao fluxo de cálculo e classificação dos riscos de estoque. O resultado corresponde a uma tabela que classifica pra cada período de previsão e para cada produto o risco de estoque como "Low stock risk", "Over-stock risk" ou “Optmal”. Esta classificação pode servir como indicador de alerta nos painéis de monitoramento. O fluxo é encapsulado dentro do meta-bloco “Risco de Falta de Estoque”.

PROJETO TÉCNICO

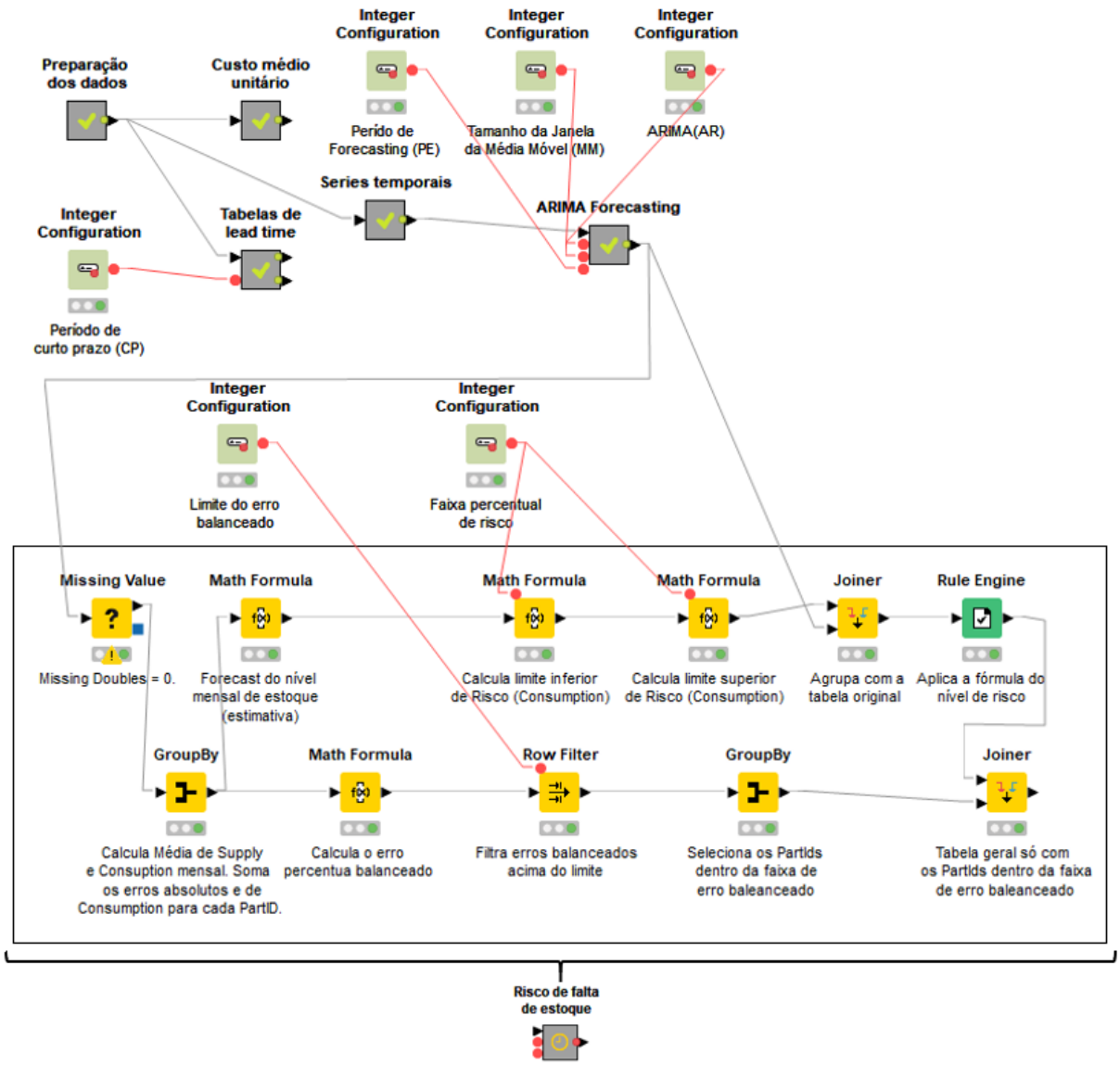


Figura 11: Fluxo KNIME para construção das séries temporais históricas de cada produto.

A tabela 5 apresenta saída do meta-bloco "Risco de Falta de Estoque". Pode-se perceber que há uma coluna (risk_level) que, em tempo real, classifica o risco de falta de estoques. Isto pode servir de indicador para um sistema de alerta.

Tabela 5: Estrutura dos dados da saída do meta-bloco "Séries Temporais"

S PartId	D INF_Co...	D SUP_C...	D Foresc...	FullDat...	S risk_level
-8002-7338	1.075	1.612	1.482	2017-06-09	Out-of-stock risk
-8002-7338	1.075	1.612	1.462	2017-07-09	Out-of-stock risk
-8002-7356	1.114	1.671	1.53	2009-02-09	Low stock risk
-8002-7356	1.114	1.671	1.53	2009-03-09	Low stock risk
-8002-7356	1.114	1.671	2.333	2009-04-09	Low stock risk
-8002-7356	1.114	1.671	0.667	2009-05-09	Low stock risk

3.8 Projeção dos Níveis de Estoque

Uma vez que o risco de falta de estoque é estimado, usando o forecasting, é necessário projetar os níveis necessários, atuais e futuros. Isto é feito estimando-se, com base no forecasting, os níveis necessários para atender a demanda.

A figura 12 representa o fluxo construído neste projeto para estimativa dos níveis de estoque. Um aspecto importante neste fluxo é que para estimar os níveis futuros de estoque é necessário conhecer os níveis atuais. Todavia, a base de dados utilizada não fornece os níveis atuais de estoque. Desta forma foi necessário assumir uma premissa de que os estoques iniciais correspondiam ao valor mínimo dos saldo negativo das transações históricas, ou seja, se considera aqui que no mês inicial haveria estoque para atender o mês de maior demanda.

Claramente esta premissa não seria necessária se na base de dados estivesse presentes os dados de estoque inicial. Na figura 13 o bloco grafado em amarelo indica onde este cálculo para o estoque inicial é feito. O fluxo de dados é encapsulado no meta-bloco "Níveis atuais e projetados de estoque". A figura 12 representa para o produto A34813-001 os níveis históricos e futuros projetados de estoque. Note que há uma projeção de queda do estoque baseados nos dados de forecasting da demanda. Esta informação será utilizada para estimativa dos níveis de estoque de segurança assim como para determinação dos níveis de pedido futuros.

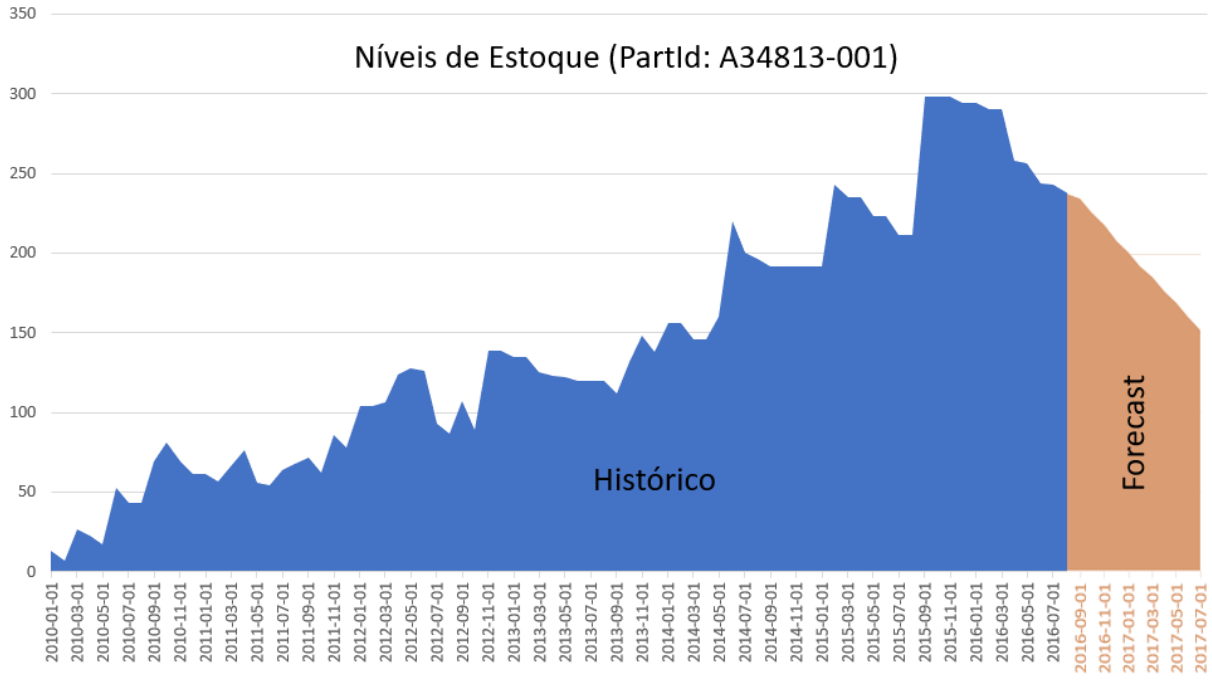


Figura 12: Projeção dos níveis futuros de estoque.

PROJETO TÉCNICO

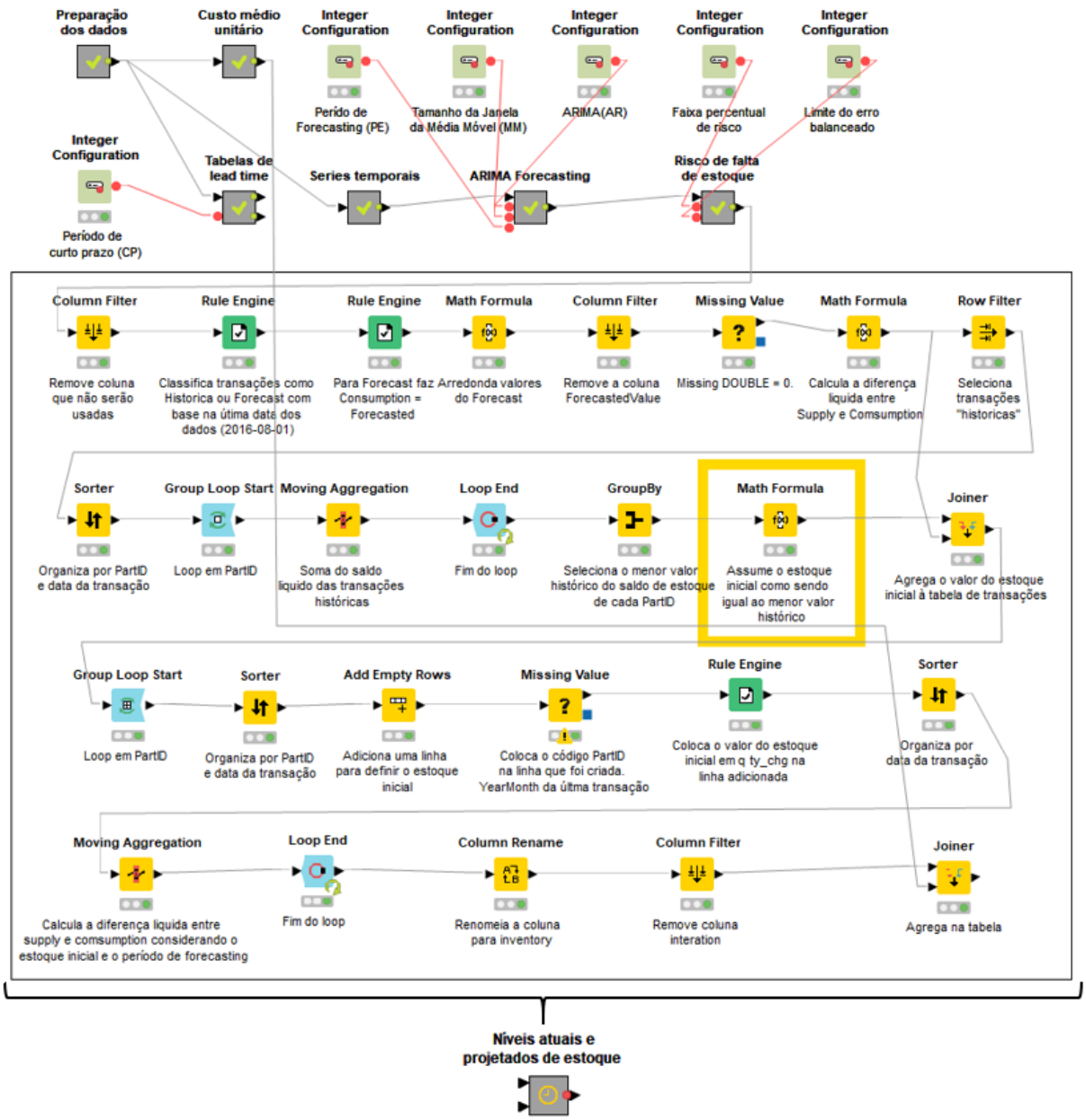


Figura 13: Fluxo KNIME de projeção dos níveis necessários (atuais e futuros) de estoque.

3.9 KPIs de Estoque

Para que se tenha um processo de tomada de decisão eficaz e eficiente no gerenciamento de estoques é importante ter-se à mão indicadores de performance de seu gerenciamento. Tais indicares podem ser utilizados como preditores de diversos problemas tais como falta de abastecimento ou falta de níveis para atendimento.

Para isto, é necessário combinar indicadores de níveis e de fluxo de consumo. Este indicadores devem ter abrangência histórica ,mas também de curto prazo. Isso é necessário para que se possa acompanhar mudanças na estrutura recente de consumo e abastecimento e compará-las com os valores históricos para a identificar pontos de mudança e ação.

A tabela 5 apresenta a lista de indicadores calculados nesta etapa. O período de curso prazo, como foi dito anteriormente, é dado por um parâmetro do fluxo geral de dados. Nas simulações deste projeto o valor escolhido foi de 36 meses.

Tabela 5: Estrutura dos dados da saída do meta-bloco "Séries Temporais"

PartId
last_inventory
unit_cost
mean_inventory_all_periods
mean_inventory_for_last_short_periods
sold_for_all_periods
sold_for_last_short_periods
inventory_turn_all_periods
inventory_turn_for_short_periods
avg_days_to_sell_all_periods
avg_days_to_sell_short_periods
inventory_value
Sum(inventory_value)

A figura 13 representa a saída do fluxo de KPIs de estoque especificamente para os indicadores de "average days to sell" para a família de produtos "-8002". Note que é possível identificar variações de padrões do comportamento de curtos prazo em relação ao comportamento históricos. Desta forma podem ser definidos limiares de variação a partir dos quais ações específicas podem ser disparadas.

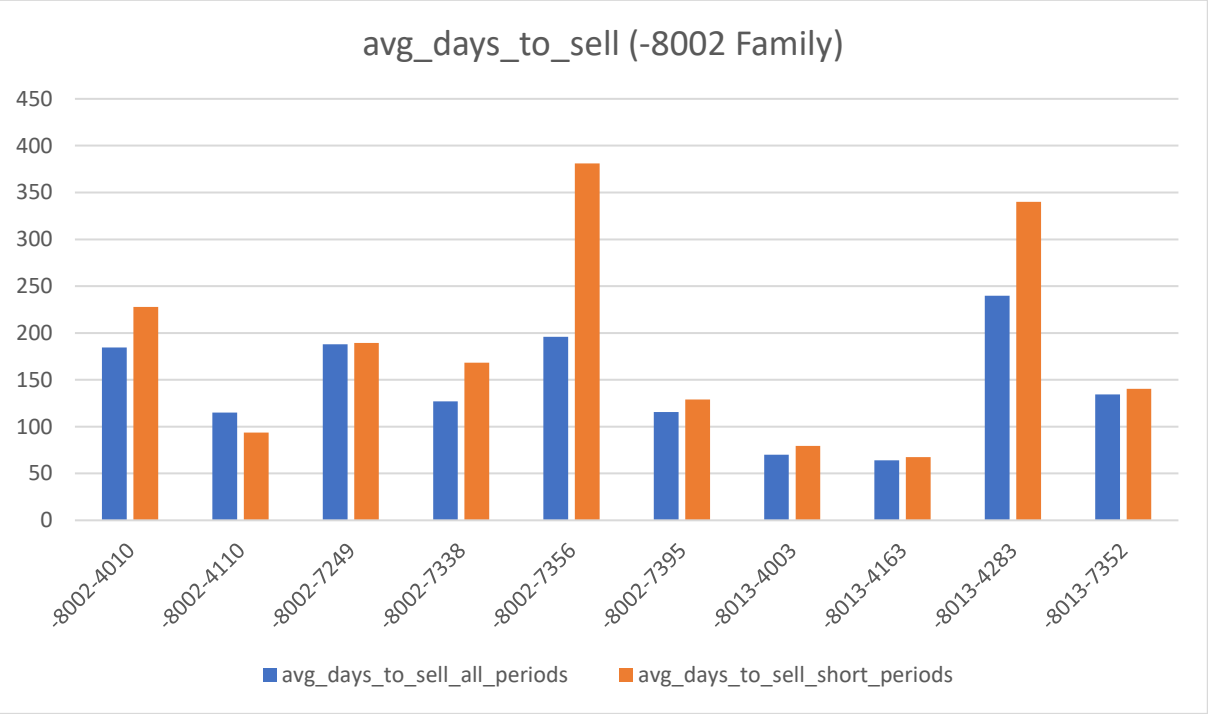


Figura 14: Distribuição de avg_days_to_sell (-8002 PartID Product Family).

PROJETO TÉCNICO

O fluxo de cálculo dos indicadores está representado n figura 14. Importante ressaltar que nesta etapa da modelagem os dados do fluxo de custos são utilizados para estimar o custo imobilizado de estoque de cada parte assim como o custo total do estoque.

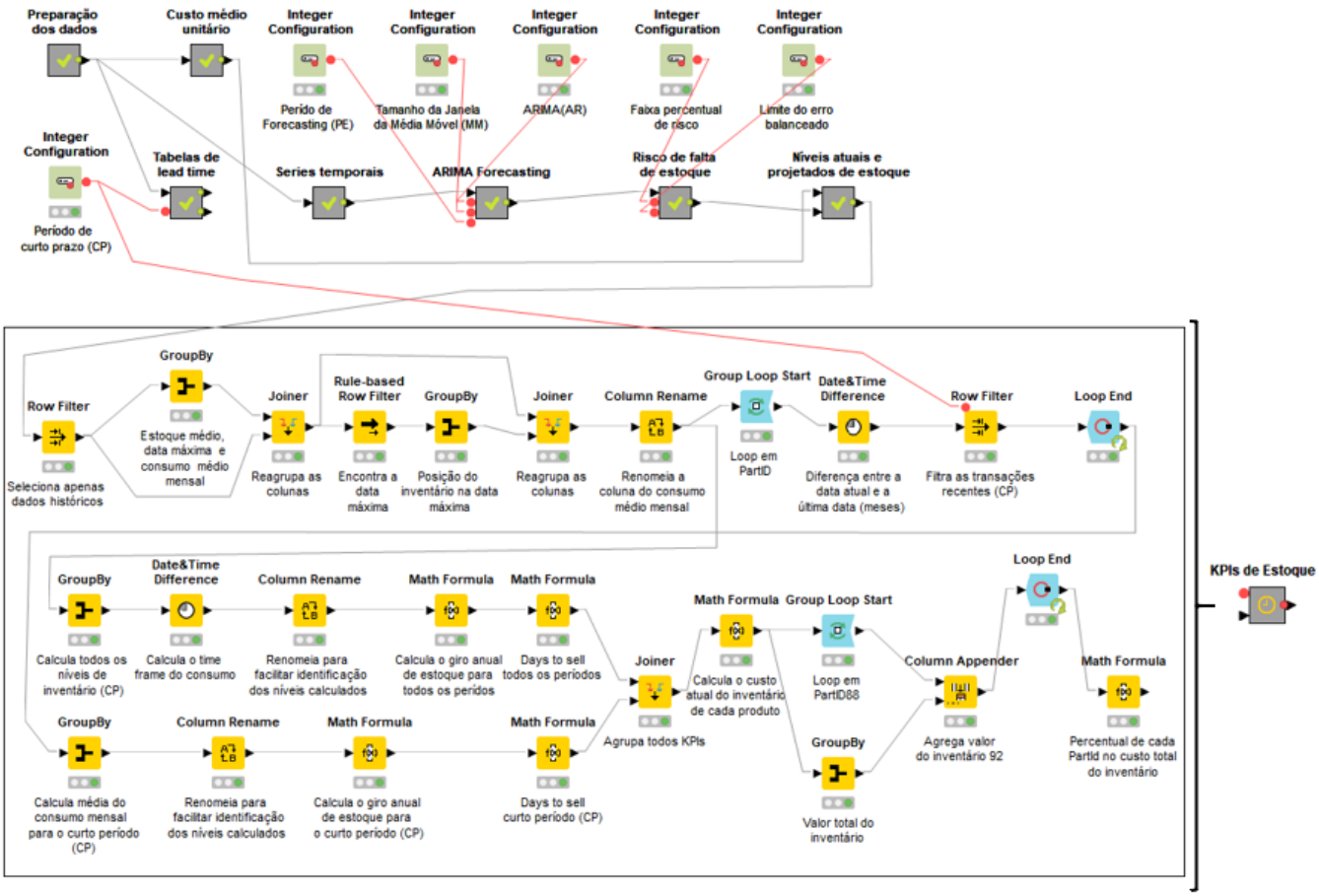


Figura 15: Fluxo KNIME para cálculo dos KPIs de estoque.

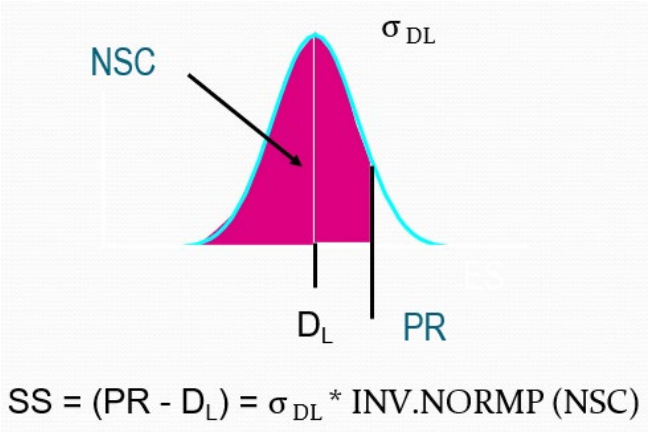
3.10 Estoque de Segurança

Os estoques de segurança (SS) são essenciais para garantia do nível de serviço de atendimento em logística (NSC). O estoque de segurança é todo o estoque adicional mantido para a empresa para se proteger contra eventuais incertezas de demanda, suprimentos ou internas.

Seu cálculo é baseado em um processo de revisão contínua dos níveis, é utilizado basicamente quando há a necessidade de manter estoques para substituir produtos vendidos em situações de alto giro. Neste caso, o estoque é continuamente acompanhado e um pedido d feito quando o estoque cai abaixo de um determinado o ponto de reposição (PR).

PROJETO TÉCNICO

Quando o estoque atinge o ponto de reposição (PR), deve-se colocar um pedido que só irá ser recebido após o período de leadtime (L). Quando este pedido chegar, o estoque idealmente deve ser igual ao estoque de segurança (SS). Estas premissas são utilizadas para calcular o nível de estoque de segurança que ao mesmo minimize o risco de quebra de estoque e ao mesmo tempo permita atingir os níveis de serviço desejados. Estima-se no momento do pedido a demanda média acumulada esperada durante o período de leadtime seja DL e seu desvio σ_{DL}. Desta forma, pode-se estimar o valor do estoque de segurança para um dado nível de serviço pela expressão abaixo:



A figura 16 ilustra algumas das saídas do fluxo de cálculo dos estoques de segurança. É possível criar-se cenários de custos dos estoques de segurança a partir de valores almejados de nível de serviço. Isto permite uma melhor forma de definir de forma tática os nível de estoque de segurança em momentos específicos.

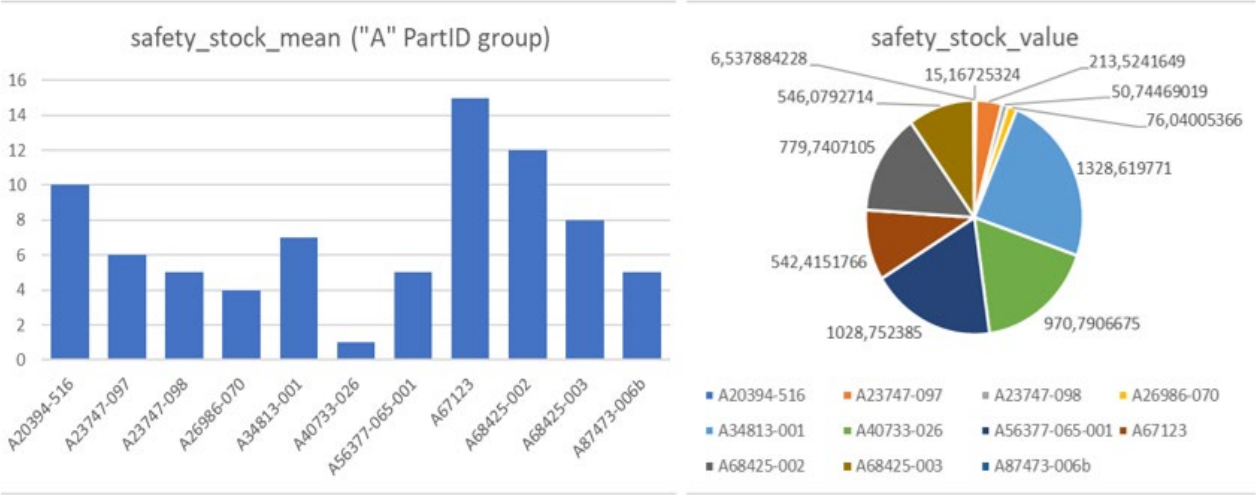


Figura 16: Média do estoque de segurança e custos do estoque de segurança ("A" PartID group)

PROJETO TÉCNICO

O fluxo de cálculo do estoque de segurança está representado n figura 14. Nesta etapa do fluxo de dados foi necessário incluir blocos externos de processamento utilizando o software estatístico R-CRAN. O KNIME permite o acesso a bibliotecas estatistas externas e o retorno de resultados processados. O fluxo foi encapsulado dentro do meta-bloco "Estoque de Segurança".

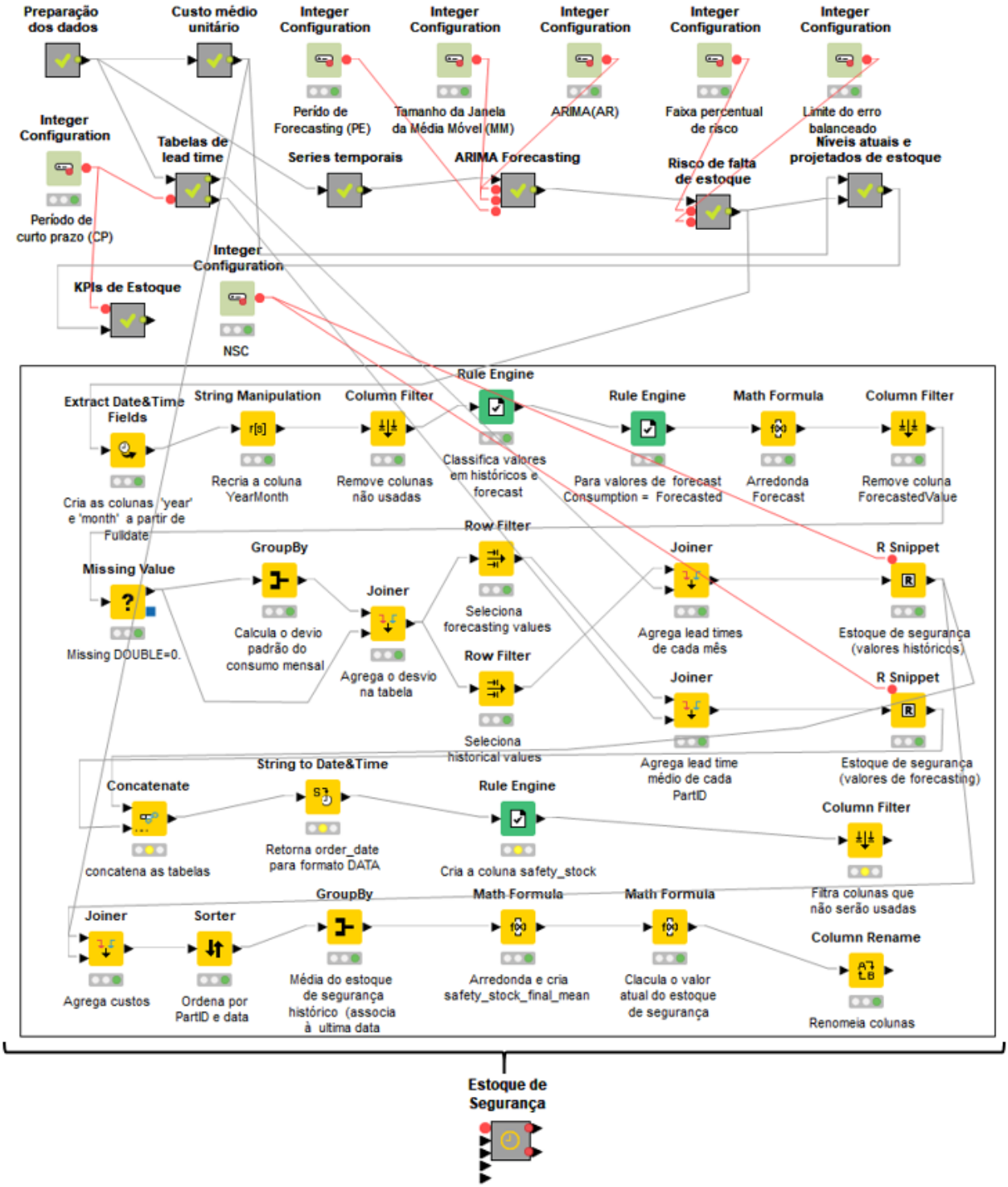


Figura 17: Fluxo de cálculo dos níveis de estoque de segurança de cada PartID

3.11 Custos de Excesso de Estoque

As estimativas de estoque de segurança quando comparadas ao níveis atuais de estoque e a previsão de demanda permitem estimar os custos atuais de excesso de estoque. Este é um indicador importante de eficiência de gerenciamento dos estoques. É possível a partir deste fluxo de dados verificar-se o valor atual do excesso de estoque para cada produto assim como verificar sua evolução no tempo.

A figura 18 representa o fluxo para cálculo do curo de excesso de estoque. Para isto são utilizadas informações sobre os custo dos produtos, o forecasting de demanda, o nível atual de estoques e o valor dos estoques de segurança. O fluxo foi encapsulado dentro do meta-bloco "Cálculo de excesso de estoque".

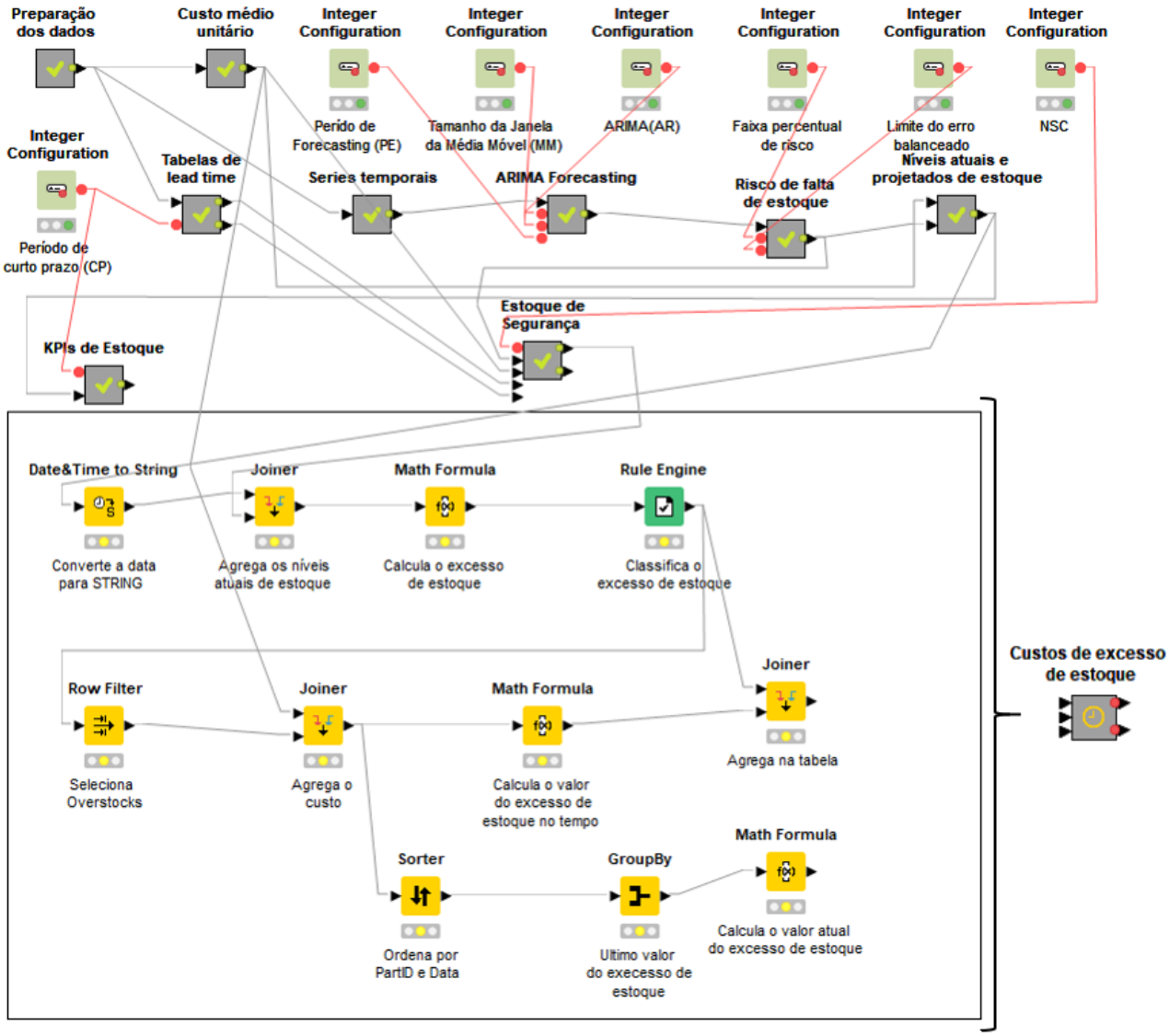


Figura 18: Fluxo de cálculo dos custos de excesso de estoque

3.12 KPIs de Consumo

O fluxo representado na figura 19 faz o agrupamento dos dados de consumo. São agregados as médias e os desvios padrão de consumo históricos e de curto prazo para cada produto. Estes dados serão utilizados para compor a tabela geral de KPIs dos dashboards de acompanhamento.

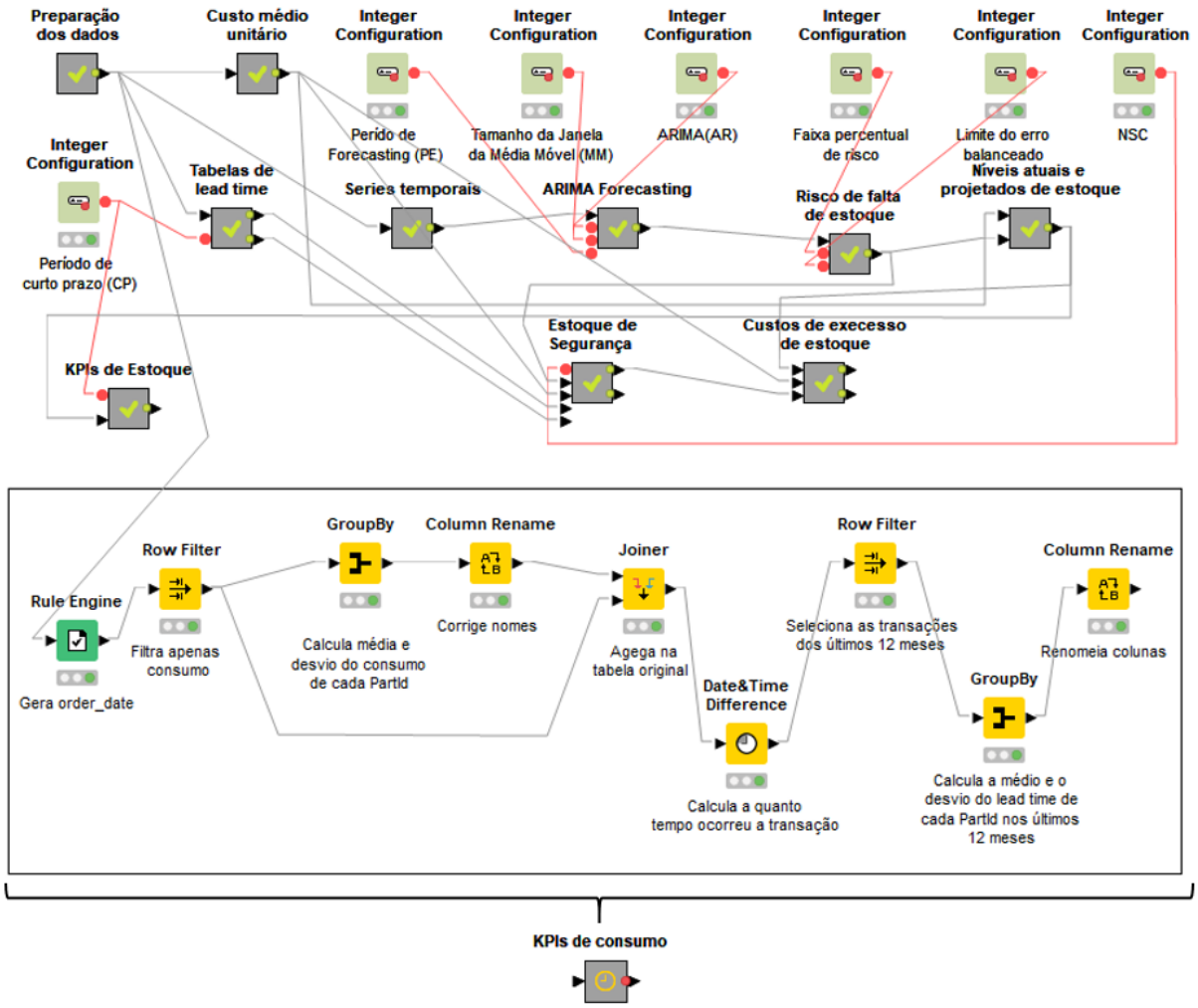


Figura 19: Fluxo de agrupamento dos KPIs de consumo (médias e desvios padrão)

3.13 Agregação das Tabelas de KPIs

O fluxo representado na figura 20 agrega todos os KPIs do modelo. Estes indicadores, calculados em tempo real pelo fluxo, podem ser utilizados pra alimentar sistemas de tomada de decisão de curto prazo (táticos) melhorando a agilidade de reposta à mudanças de comportamento, tanto da demanda, quanto do abastecimento. A tabela 6 faz um resumo de todos os KPIs calculados pelo modelo. Esta tabela pode ser agora ser utilizada por sistemas de tomada de decisão, inclusive de geração de pedidos automáticos de ressuprimento.

PROJETO TÉCNICO

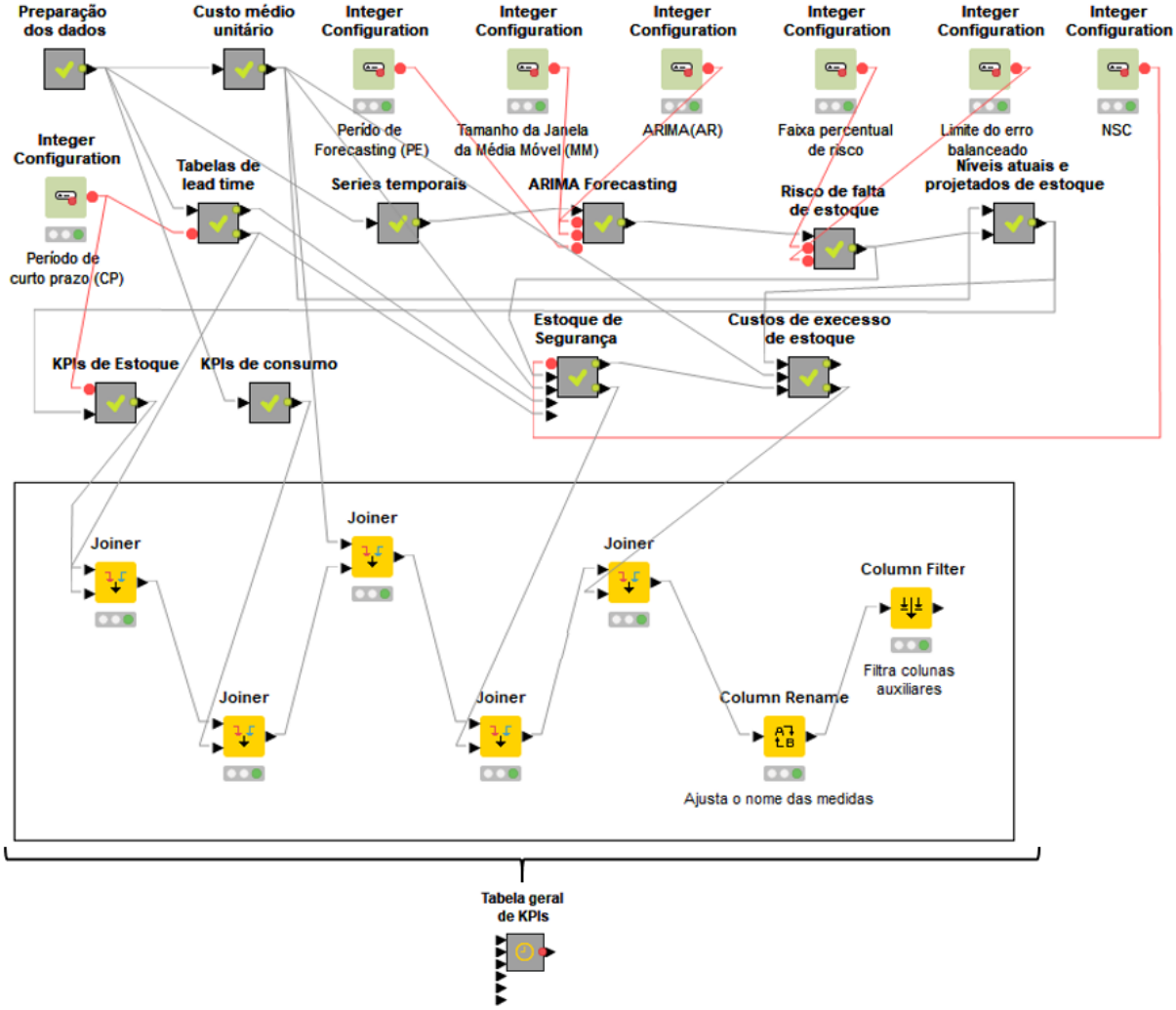


Figura 20: Fluxo de agregação de todos os KPIs calculados

Tabela 6: Tabela final de KPIs do modelo construído

PartId
unit_cost
lead_time_mean_all
lead_time_sd_all
lead_time_mean_recent
lead_time_sd_recent
inventory_mean_all
consumption_sum_all
inventory_turn_all
avg_days_to_sell_all
sold_for_last_short_periods
mean_inventory_for_last_short_periods
inventory_turn_recent
avg_days_to_sell_recent
inventory_value
consumption_mean_all
consumption_sd_all
consumption_mean_12_periods
sd_consumption_12_months
safety_stock_mean
safety_stock_value
inventory_mean_recent

3.14 Indicadores de Estoque

O fluxo da figura 20, agrega todos os indicadores por produto (PartID). Todavia, muitas vezes é necessário para se tomar decisões ter acesso às evoluções históricas dos indicadores para cada produto. Estas evoluções são perdidas na agregação de dados feita na figura 20. Desta forma, foi necessário construir um fluxo que permitisse um registro dos dados históricos de evolução do níveis de estoque de cada produto. Isto é feito por meio do fluxo representado na figura 21.

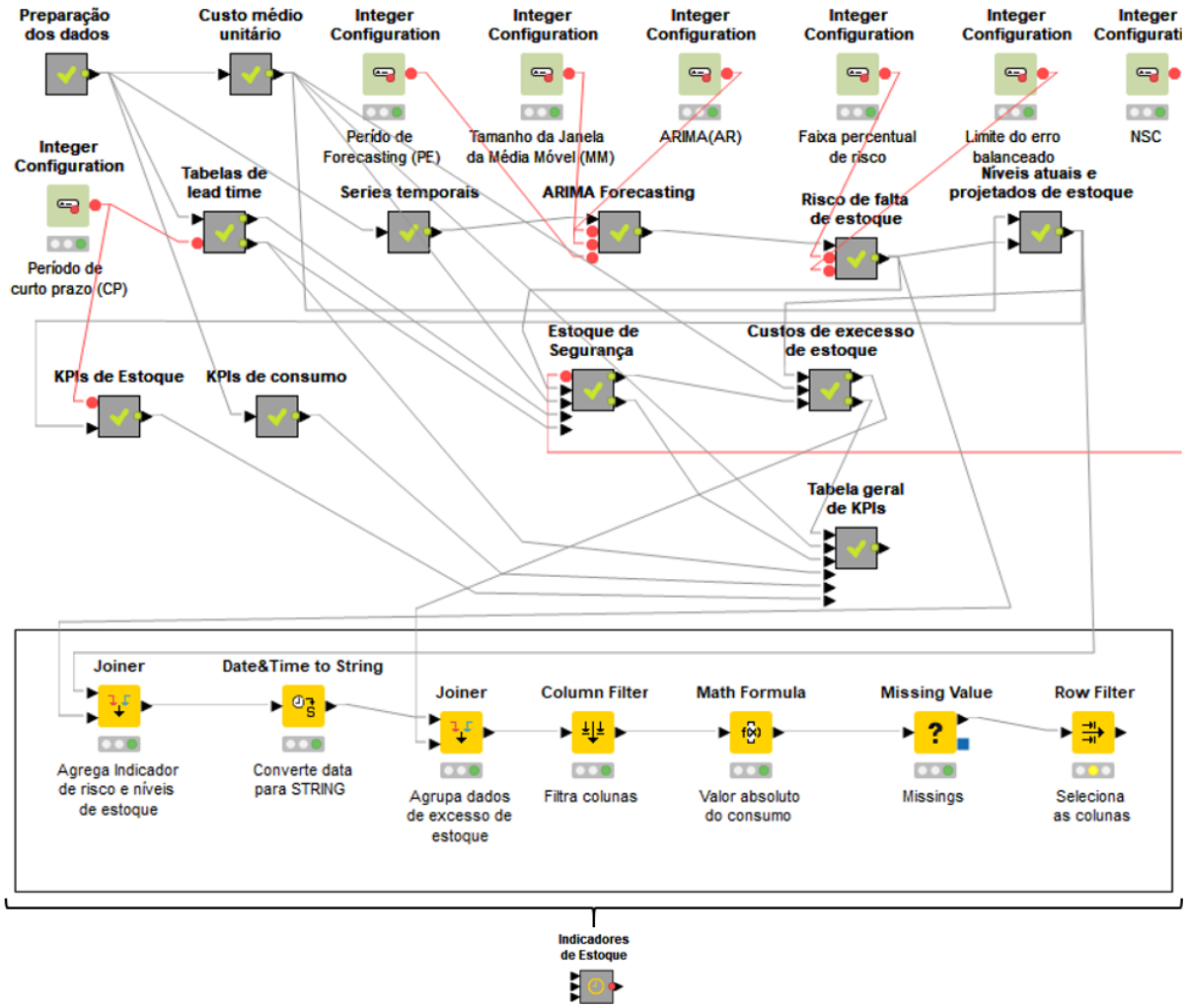


Figura 20: Fluxo de agregação dos dados de evolução temporal dos estoques para cada PartID

3.15 Classificação com Redes Neurais

Historicamente existem várias abordagens e heurísticas para o planejamento de estoques, o que que é comum entre elas em geral é o uso de simplificações significativas para maior velocidade no processo de tomado de decisão. Concluir análises no nível de componentes

individuais usando dados históricos detalhados é importante, por outro lado, tais análises tem de dispor de instrumentos de tempo real para melhoria das ações de resposta. Essa abordagem automatizada, conhecida como Real Time BI, simplifica e eleva a precisão do trabalho do processo de gerenciamento de estoques. Em geral, é necessária um tratamento e agregação de dados para que tais técnicas possam ser utilizadas. Isto feito nas seções anteriores culminado com os dados da tabela 6. Tais dados, quando preparados por fluxos ETL podem alimentar algoritmos de classificação e tomada de decisão.

A ideia nesta seção é mostrar com estes dados podem ser utilizados, por exemplo, para fornecer uma análise rápida de níveis de estoque com base em um algoritmo de classificação supervisionada com uso de redes neurais.

Em uma técnica de classificação supervisionada Dados N objetos caracterizados por atributos (variáveis preditoras) e um rótulo existente (variável classe) o objetivo é gerar a partir dos dados um classificador capaz de predizer com alta confiabilidade e velocidade as classes de futuros exemplos. Ou seja, existe um conjunto de dados de "aprendizado" a partir do qual o algoritmo "aprende" como classificar observações futuras.

Redes neurais podem ser utilizadas para classificação supervisionada. As redes neurais são modelos de inspirados no funcionamento do cérebro humano. São compostas por um número elevado de modelos de neurônios artificias trabalhando na resolução de um problema. O conhecimento adquirido por meio da aprendizagem é armazenado nas ligações deste neurônios. O neurônio artificial é, na realidade, uma função que mapeia diversas entradas em uma saída controlada por uma função de classificação, chamada de função de ativação. A figura 21 esquematiza o processo de treinamento de uma rede neural. É necessário se ter um conjunto de treino.

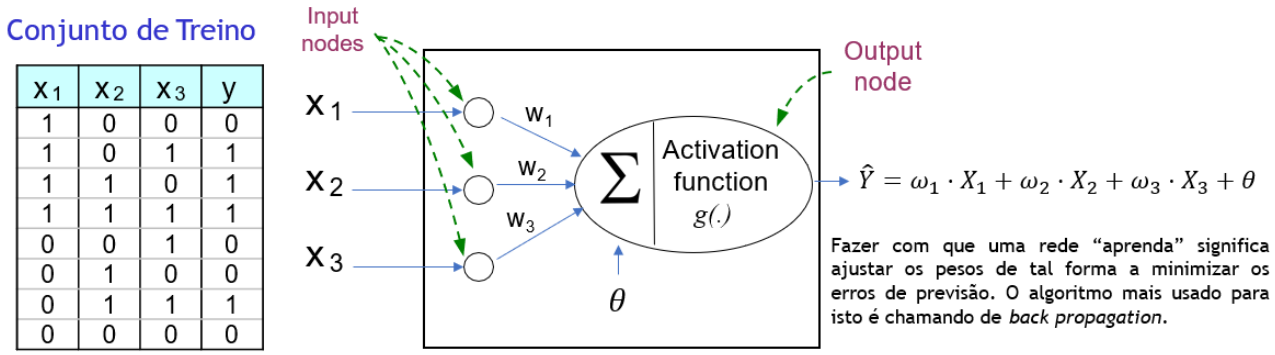


Figura 21: Processo de treinamento de uma rede neural de 1 neurônio.

PROJETO TÉCNICO

Na figura 22 está representado um fluxo que faz a classificação do nível de ocupação do estoque em duas faixas (alto e baixo) a partir de 3 váriáveis: lead_time_mean_all, avg_days_to_sell_all e consumption_mean_all. A ideia é verificar se é possível, dadas estas 3 características determinar se um produto terá nível alto ou baixo de ocupação do estoque. A determinação das faixas Alta e Baixa é feita a partir de um algoritmo de cálculo de faixas (binners).

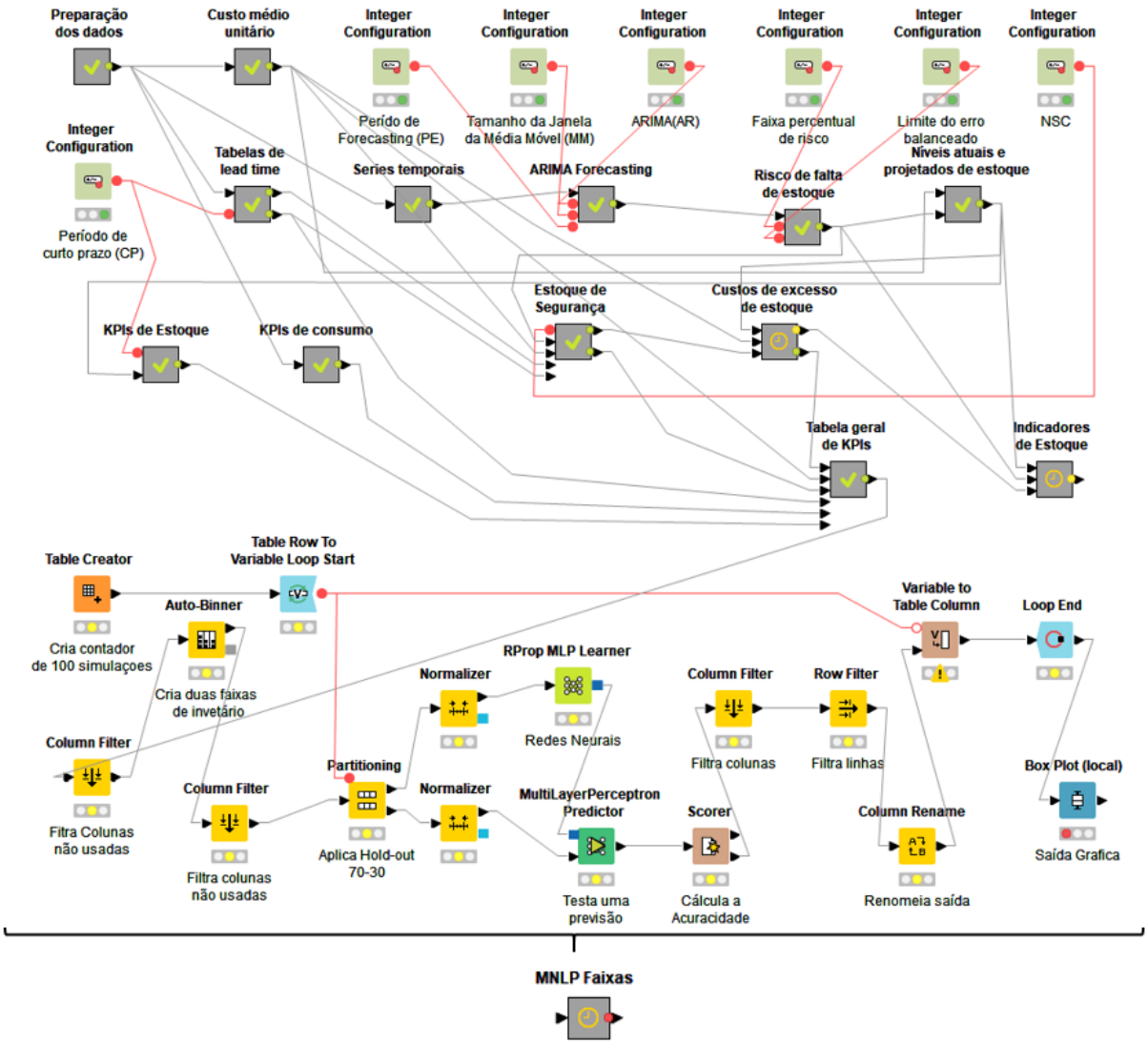


Figura 22: Processo de treinamento de uma rede neural de 1 neurônio.

Os da tabela 6 de indicadores foram divididos para uso de uma técnica de avaliação cruzada conhecida como “Hold-out”. O conceito central das técnicas de Hold-out é o particionamento do conjunto de dados em dois subconjuntos mutuamente exclusivos, e posteriormente, o uso de um destes subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento), sendo o subconjunto restantes (dados de validação ou de teste) é empregado

na validação do modelo. Foram então executadas 100 simulações para testar a acuracidade de previsão da rede neural para a faixa de ocupação considerando apenas as variáveis de lead time, giro e consumo. O resultado pode ser visto na figura 22. O modelo treinado permite uma taxa de acerto média de 85% para o nível de ocupação de estoque de um produto apenas com estas 3 informações. O valor mínimo de acurácia para as amostra de teste foi de 60%, chegando em alguns casos a 100%. Por fim, a tabela 7 apresenta, para uma amostra de 20 produtos, a previsão da faixa de ocupação do estoque e as probabilidades de pertencimento.

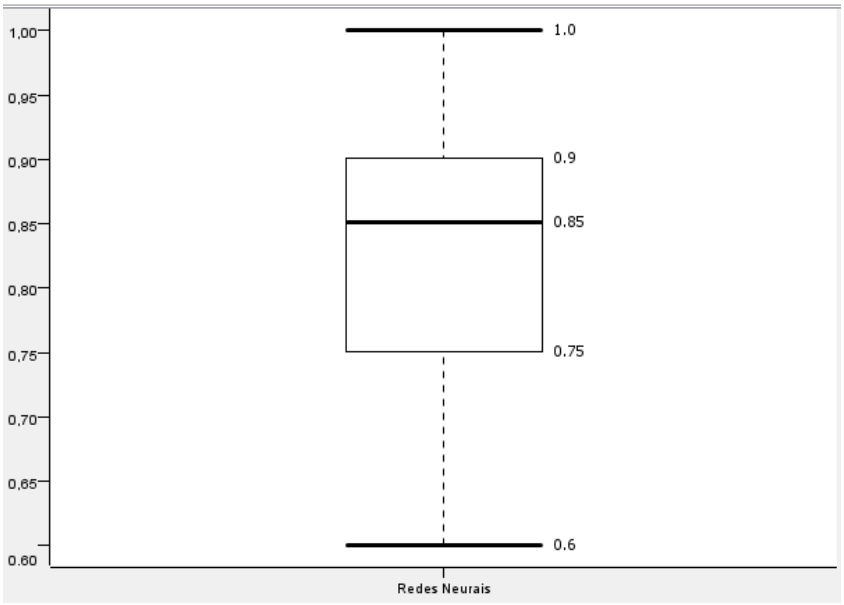


Figura 22: Processo de treinamento de uma rede neural de 1 neurônio.

Tabela 7: Tabela final de KPIs do modelo construído

D lead_time_...	D avg_days_t...	D consumption_...	S invento...	D P (inve...	D P (inve...	S Predicti...
0.15	0.507	0.259	Bin 2			Bin 2
0.408	0.274	0.435	Bin 1			Bin 2
0.098	0.932	0.646	Bin 2			Bin 2
0.122	0.8	0.808	Bin 2			Bin 2
0.303	0.347	0.678	Bin 2			Bin 2
0.513	0.123	0.075	Bin 1			Bin 1
0.166	0.015	0.013	Bin 1			Bin 1
0.212	1	0	Bin 2			Bin 2
0.067	0.143	0.502	Bin 2			Bin 2
0.04	0.151	0.1	Bin 1			Bin 1
0.337	0.382	0.84	Bin 2			Bin 2
0.285	0.292	0.079	Bin 1			Bin 1
0.12	0.59	0.188	Bin 1			Bin 2
0.456	0.08	0.275	Bin 1			Bin 1
0.36	0.753	0.726	Bin 2			Bin 2
0.157	0	1	Bin 1			Bin 2
0	0.263	0.472	Bin 2			Bin 2
0.097	0.442	0.377	Bin 2			Bin 2
0.576	0.114	0.553	Bin 1			Bin 2
1	0.205	0.303	Bin 1			Bin 1

4. Conclusões

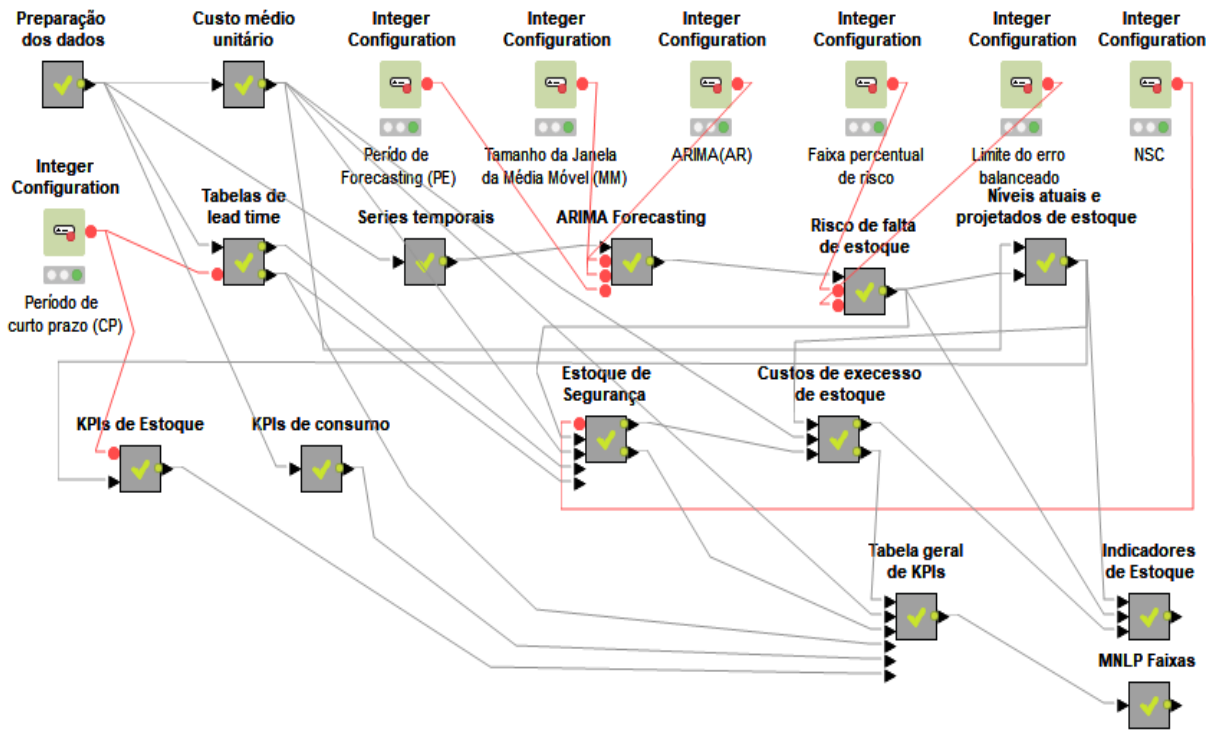


Figura 23: Fluxo final KNIME do modelo desenvolvido no projeto

A cadeia de suprimentos é uma pedra angular de competitividade para qualquer empresa, seja de manufatura ou de serviços. Ela é responsável direta por uma parte significativa dos os custos de operação. Definir os níveis ideais de estoque e gerenciar seus níveis para garantir eficiência e nível de serviço requer uma boa compreensão tanto do comportamento dos fornecedores como do uso e consumo dos componentes e partes. Finalmente, é importante ter claro o nível de serviço necessário para atendimento.

Neste sentido, o modelo desenvolvido neste projeto buscou automatizar, com uso em dados passados e algoritmos de análise de dados, o processo de seleção das estratégias de gestão de estoque com base nos de níveis de serviço almejados.

O atual ambiente de crise e pandemia, a complexidade dos processos de controle e gerenciamento de estoques tem crescido em função do crescimento das estruturas de e-commerce e pelo aumento natural da competição em nível local e internacional.

Estratégias de gerenciamento baseada em modelos estatísticos e determinísticos, em geral suportados por programação linear, tem tido pouca capacidade de adaptação às mudanças cada vez mais rápidas dos cenários.

É importante buscar desenvolver mecanismos que permitam não só a emissão de relatórios automatizados de dados para expandir os recursos de análise das posições de estoque, mas também automatizar os processos e tomada de decisão que envolvem compra e precificação e, neste sentido, ainda há um amplo espaço para que este projeto seja avançado endereçando estre problemas por meio da automatização do gerenciamento de estoques.

Espera-se com base no que foi apresentado que se possa perceber o potencial deste tipo de abordagem para expandir as capacidades de processamento analítico das informações de estoque. O uso de abordagens ETL representa um marco importante na jornada de aperfeiçoamento dos sistemas de gerenciamento de estoques. Ao automatizar o processo de interpretação de dados com ferramentas como o KNIME, os gestores podem ser capazes de identificar oportunidades potenciais e alavancar a competitividade das empresas.

5. Bibliografia

Chun-Tian Cheng, Lin Qiu, Kwok-Wing Chau, Wen-Chuan Wang, 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Available on: <https://www.sciencedirect.com/science/article/pii/S0022169409003527?casa_token=A1TWvdIhk-AAAAAA:LZmhyOg3eAwr-9YFxPaLOXZJP7rcnOmrXokM4QPNBas2DtByE9EwN-rijtCcodhqM1wVc3_NhA#!>.

Min, Hockey, 2009. Artificial intelligence in supply chain management: theory and applications. Available on: <https://www.tandfonline.com/doi/full/10.1080/13675560902736537?scroll=top&needAccess=true>.

Remko van Hoek, Brian Gibson & Mark Johnson, 2020. Talent Management For a Post-COVID-19 Supply Chain—The Critical Role for Managers. Available on: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7753586/>.

Timme, S. G. and Williams-Timme, C. 2003. The real costs of holding inventory. *Supply Chain Management Review*, 7: 30–37.

Zixin Hu, Qiyang Ge, Shudi Li, Li Jin and Momiao Xiong, 2020. Artificial Intelligence Forecasting of Covid-19 in China. Available on: <https://arxiv.org/pdf/2002.07112.pdf>.