

**UNIVERSIDADE DO VALE DO RIO DOS SINOS – UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS CONTÁBEIS
NÍVEL MESTRADO**

MICHEL VERZELETTI

**IDENTIFICAÇÃO DE RUPTURAS DE ESTOQUE NO VAREJO ALIMENTAR
ATRAVÉS DE TÉCNICAS DE SOFT COMPUTING**

Porto Alegre - RS

2024

MICHEL VERZELETTI

**IDENTIFICAÇÃO DE RUPTURAS DE ESTOQUE NO VAREJO ALIMENTAR
ATRAVÉS DE TÉCNICAS DE SOFT COMPUTING**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciências Contábeis: Controle de Gestão, pelo Programa de Pós-Graduação em Ciências Contábeis da Universidade do Vale do Rio dos Sinos - UNISINOS

Orientador: Prof. Dr. Alessandro Marian Carvalho

Porto Alegre - RS

2024

V574i Verzeletti, Michel.
Identificação de rupturas de estoque n varejo alimentar através de técnicas de soft computing / Michel Verzeletti. – 2024.
78 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Ciências Contábeis, 2024.
“Orientador: Prof. Dr. Alexsandro Marian Carvalho”

1. Algoritmo de detecção. 2. Identificação de ruptura. 3. Lógica difusa. 4. Ruptura de estoque. 5. Varejo alimentar. I. Título.

CDU 657

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Silvana Dornelles Studzinski – CRB 10/2524)

MICHEL VERZELETTI

**IDENTIFICAÇÃO DE RUPTURAS DE ESTOQUE NO VAREJO ALIMENTAR
ATRAVÉS DE TÉCNICAS DE SOFT COMPUTING**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciências Contábeis: Controle de Gestão, pelo Programa de Pós-Graduação em Ciências Contábeis da Universidade do Vale do Rio dos Sinos - UNISINOS

BANCA EXAMINADORA

Dr. Luis Felipe Dias Lopes – UFSM

Dr. Andre L Korzenowski – UNISINOS

Dr. Roberto Frota Decourt – UNISINOS

Dr. Aleksandro Marian Carvalho – UNISINOS

Em memória da Luna.

AGRADECIMENTOS

Ao meu orientador, professor Doutor Alexsandro Carvalho, pelos ensinamentos, dedicação e paciência durante esse processo.

À minha amada esposa Fran, por ter me incentivado a ingressar no mestrado e pelo apoio e compreensão durante este período.

Aos meus filhos felinos, Muffin, Ágatha e Amora que, com seus ronronados reconfortantes e travessuras encheram meus dias de leveza e inspiração.

RESUMO

Rupturas de estoque geram perdas significativas para o varejo e, apesar de décadas de estudo, as taxas de ruptura não têm diminuído consideravelmente. O primeiro passo para melhoria deste indicador é uma correta medição e detecção de rupturas. Desta forma, o objetivo deste estudo é determinar a eficácia de um método para identificação de rupturas no varejo alimentar baseado exclusivamente em dados transacionais. O foco da pesquisa foi exclusivamente a identificação de casos de ruptura total de estoque causadas pela falta efetiva do produto no ponto de venda, ou seja, quando a ausência de vendas em um determinado período foi causada exclusivamente pela indisponibilidade de estoque. O algoritmo desenvolvido utilizou técnicas de lógica difusa para determinação da ocorrência de ruptura, a partir de probabilidades baseadas na ocorrência de venda nula e em sua possível sequência, na venda acumulada e no saldo a partir da última entrada. Este algoritmo foi aplicado em 2.022 itens selecionados no período de 12 meses e obteve uma taxa de identificação correta de 70,2% com apenas 16,4% de alarmes falsos. Embora alguns estudos anteriores tenham obtido resultados superiores, suas abordagens estavam restritas a um número muito reduzido de itens. Além disso, a aplicação do modelo mostrou-se promissora, identificando a ruptura em média apenas 4 dias após sua ocorrência, o que pode proporcionar um incremento estimado de 1,3% no faturamento total da empresa.

Palavras-chave: ruptura de estoque; varejo alimentar; identificação de rupturas; lógica difusa; algoritmo de detecção.

ABSTRACT

Out-of-stock (OOS) leads to significant losses for retailers, and despite decades of study, out-of-stock rates have not decreased considerably. The first step to improve this indicator is accurate measurement and detection of OOS. Therefore, the aim of this study is to determine the effectiveness of a method for identifying OOS in the grocery retail sector based solely on transactional data. The research focus was exclusively on identifying cases of total OOS caused by the actual absence of the product at the point of sale, i.e., when the lack of sales in a certain period was solely caused by stock unavailability. The developed algorithm used fuzzy logic techniques to determine the occurrence of OOS based on probabilities derived from zero sales occurrence and its possible sequence, accumulated sales, and balance from the last entry. This algorithm was applied to 2,022 selected items over a 12-month period and achieved a correct identification rate of 70.2% with only 16.4% false alarms. Although some previous studies have achieved superior results, their approaches were restricted to a very small number of items. Furthermore, the application of the model proved promising, identifying OOS on average only 4 days after their occurrence, which may result in an estimated 1.3% increase in the company's total revenue.

Keywords: out-of-stock; grocery retail; out-of-stock identification; fuzzy logic; detection algorithm.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de Identificação Inicial de Rupturas	41
Figura 2 – Funções de Pertinência.....	47
Figura 3 – Diagrama de análise de um produto p no dia d	51

LISTA DE TABELAS

Tabela 1 – Produtos	27
Tabela 2 – Operações de Estoque.....	28
Tabela 3 – Registro de Operações (exceto venda).....	28
Tabela 4 – Operação de Venda	28
Tabela 5 – Saldo da Última Compra	29
Tabela 6 – Prazo Médio de Reposição	31
Tabela 7 – Venda Acumulada	31
Tabela 8 – Identificação da Ruptura.....	40
Tabela 9 – Amostra sem Ruptura.....	43
Tabela 10 – Intervalos de Probabilidades “Baixas”	44
Tabela 11 – Amostra com Ruptura.....	44
Tabela 12 – Intervalos de Probabilidades “Altas”	44
Tabela 13 – Desvio Padrão (Amostra Total)	45
Tabela 14 – Intervalos de Probabilidades “Médias”	45
Tabela 15 – Intervalos da Variável Consequente (R).....	45
Tabela 16 – Pontos Médios.....	46
Tabela 17 – Regras de Inferência	49
Tabela 18 – Definição do Ponto de Corte (m)	50
Tabela 19 – Rupturas Identificadas.....	52
Tabela 20 – Matriz de Confusão Binária	53
Tabela 21 – Real x Modelo.....	57
Tabela 22 – Matriz de Confusão do Modelo.....	58
Tabela 23 – Rupturas Identificadas.....	58
Tabela 24 – Teste de Regressão Linear Ordinária (OLS).....	60
Tabela 25 – Análise das Variáveis (OLS).....	61
Tabela 26 – Ruptura por Dia da Semana	62
Tabela 27 – Duração da Ruptura de Estoque	63
Tabela 28 – Dias para Identificação da Ruptura	64
Tabela 29 – Prazo Médio de Reposição	65
Tabela 30 – Prazo Médio de Reposição	65
Tabela 31 – Rupturas por Categoria de Produtos.....	67
Tabela 32 – Diminuição das Rupturas	68

LISTA DE SIGLAS

<i>ERP</i>	<i>Enterprise Resource Planning</i>
<i>IRI</i>	Imprecisões no Registro de Inventário
<i>OOS</i>	<i>Out-Of-Stock</i>
<i>SKU</i>	<i>Stock Keeping Unit</i>

SUMÁRIO

1	INTRODUÇÃO	8
1.1	TEMA	10
1.2	DELIMITAÇÃO DO TEMA.....	10
1.3	PROBLEMA	10
1.4	OBJETIVOS	10
1.4.1	Objetivo Geral.....	10
1.4.2	Objetivos Específicos	10
1.5	JUSTIFICATIVA	11
2	FUNDAMENTAÇÃO TEÓRICA.....	14
2.1	RUPTURA DE ESTOQUE.....	14
2.1.1	Reação do consumidor.....	15
2.1.2	Causas da ruptura de estoque	16
2.1.3	Identificação e Medição	18
2.2	ACURACIDADE DE ESTOQUE.....	20
2.3	PREVISÃO DE DEMANDA	22
2.4	<i>DATA MINING E SOFT COMPUTING</i>	23
2.5	CLASSIFICAÇÃO DA PESQUISA.....	25
2.6	DADOS: COLETA E PREPARAÇÃO	25
2.6.1	Seleção de produtos	26
2.6.2	Movimentação dos produtos	27
2.7	PROCESSAMENTO DOS DADOS	29
2.8	MODELO PARA IDENTIFICAÇÃO.....	32
2.8.1	Variáveis de entrada.....	32
2.8.2	Identificação Inicial dos Dados	40
2.8.3	Algoritmo	42
2.8.4	Avaliação do Modelo	52
2.8.5	Limitações do Modelo	56
3	ANÁLISE DOS RESULTADOS	57
3.1	ANÁLISE DO MODELO	57
3.2	ANÁLISE DAS VARIÁVEIS.....	60
3.3	IMPACTO REAL.....	62
4	CONSIDERAÇÕES FINAIS	70

REFERÊNCIAS.....	74
------------------	----

1 INTRODUÇÃO

O tema da ruptura de estoque tem sido reconhecido como um problema crítico cada vez mais importante no varejo. As estimativas indicam que os níveis de ruptura de estoque têm consistentemente ultrapassado a média de 8,3%, prejudicando o desempenho das empresas (GRANT; FERNIE, 2008; GRUEN; CORSTEN, 2007). Embora a falta de um produto possa aumentar a probabilidade de um cliente considerar a compra de um produto similar como substituto, não há necessariamente um aumento proporcional na chance de compra de outros produtos (BARRICHELLO et al., 2019). Portanto, o gerenciamento eficaz da ruptura de estoque e a redução dos índices podem se tornar uma vantagem competitiva sustentável para as empresas (COELHO et al., 2015).

Além da redução imediata de receita, a ruptura pode impactar em aspectos de médio e longo prazo nas operações e no resultado das organizações. A falta de um produto pode deformar a demanda, não só deste item, mas também de itens a ele relacionados (CADAVID; LAMOURI; GRABOT, 2018). Além de impactar negativamente nas projeções futuras deste item e superestimar necessidades dos produtos substitutos, ainda existe a possibilidade de impacto na demanda de itens da cesta de compra associados à compra do item faltante (BARRICHELLO et al., 2019). Em casos extremos, a insatisfação do cliente no momento da compra pode levá-lo a procurar a concorrência, resultando na perda da venda atual e de possíveis vendas futuras, além de prejudicar a imagem da empresa perante outros clientes (COELHO et al., 2015).

No varejo alimentar brasileiro, a ruptura de estoques traz perdas significativas. Em um setor que representa aproximadamente 7,03% do Produto Interno Bruto nacional (RUIZ, 2023) a melhoria deste indicador é essencial como diferencial em um ambiente com cada vez mais concorrentes. Essa ruptura no varejo alimentar é definida como a não disponibilidade para o cliente de um item na área de vendas (COELHO et al., 2015). Embora, a medição dos níveis de estoque nas lojas seja facilmente avaliada através de sistemas de informação, normalmente existem inconsistências, ou seja, há uma diferença entre o estoque físico real e o encontrado no sistema (AGUIAR; SAMPAIO, 2013). Além disso, em alguns casos, embora haja estoque tanto físico quanto no próprio sistema, erros operacionais podem fazer com

que o item não esteja disponível no ponto de venda para o cliente (GRUEN; CORSTEN, 2007).

A grande quantidade de estudos existentes sobre o tema corrobora com a relevância de seu impacto. Porém, dentre os estudos se destacam aqueles com foco em investigar o comportamento dos consumidores frente à ruptura, avaliando a falta de itens pontuais e seu impacto, sem a preocupação de identificar os itens não disponíveis (BARRICHELLO et al., 2019; HAUSRUCKINGER, 2006). Outra parcela significativa está voltada para identificação das causas e possíveis soluções da ruptura (AGUIAR; SAMPAIO, 2013; COELHO et al., 2015; GRANT; FERNIE, 2008). Apesar das pesquisas que investigam as causas da ruptura de estoque destacarem sua natureza multifacetada e desafiadora em termos de identificação e correção, há uma escassez significativa de estudos relacionados à sua medição e identificação. Devido à complexidade, os estudos neste sentido fazem ensaios com apenas um item e diversas verificações manuais que não possuem garantias de confiabilidade (AGUIAR; SAMPAIO, 2013) e, de forma mais ampla, ainda fica restrito a itens de alto giro (GRUEN; CORSTEN, 2007).

A movimentação diária de milhares de itens no varejo alimentar é uma das principais razões para a complexidade envolvida nesse contexto. O grande número de itens torna inviável uma verificação manual e produz um volume de dados que, trabalhado de forma eficiente, pode trazer informações essenciais para o negócio. Nesse contexto, o uso de técnicas de mineração de dados (*data mining*) pode ser útil para explorar esses dados e identificar padrões e tendências (HUNG; CHUANG; WU, 2021; MITRA; PAL; MITRA, 2002). Entretanto, a identificação de rupturas, principalmente de itens com baixo volume de vendas, pode ser desafiadora e nem sempre precisa. Por isso, métodos de *soft computing*, que são desenvolvidos para lidar com informações incertas (CHANDRASEKARAN et al., 2010; MITRA; PAL; MITRA, 2002) são uma alternativa viável para esse tipo de verificação.

Assim, em um cenário no qual existe uma dificuldade na medição dos índices de ruptura, especialmente quando se trata de itens de menor giro no ponto de venda, os próprios índices divulgados podem estar subdimensionados, pois sua assertividade está voltada apenas para produtos de maior relevância. Desta forma, o objetivo deste estudo é construir um modelo eficaz de identificação de rupturas de estoque para o varejo alimentar, que considere itens de alto e baixo giro e forneça indicadores

abrangentes deste problema. Este modelo será baseado na exploração dos dados produzidos pelas empresas e sua exploração através de ferramentas e técnicas de *data mining* e *soft computing*.

1.1 Tema

Ruptura de estoque.

1.2 Delimitação do Tema

Identificação de itens em ruptura de estoque no varejo alimentar.

1.3 Problema

Qual é a eficácia de um método autônomo, baseado em dados transacionais, para monitorar a ruptura de estoque no setor de varejo alimentar?

1.4 Objetivos

1.4.1 Objetivo Geral

O objetivo desta pesquisa é determinar a eficácia de um método autônomo, baseado apenas em dados transacionais, para monitorar a ruptura de estoque no setor de varejo alimentar.

1.4.2 Objetivos Específicos

Para obtenção da resposta perante a questão abordada, tem-se como objetivos específicos:

- a) Elaborar um método analítico autônomo para detecção de ruptura, baseado apenas em dados transacionais;
- b) Implementar o método desenvolvido em um ambiente de varejo real;

c) Estimar o impacto financeiro e operacional da utilização do método desenvolvido.

1.5 Justificativa

A ruptura de estoque é uma preocupação constante para o setor varejista, e tem sido objeto de estudos que analisam esta questão. Peckham (1963) realizou o primeiro estudo sobre o assunto, relatando uma taxa de ruptura de estoque de 8,5%. Apesar dos avanços na gestão e da implementação de controles informatizados, estudos mais recentes indicam uma taxa de ruptura de estoque global em torno de 8,3% (GRUEN; CORSTEN, 2007). A similaridade destes índices demonstra que, apesar dos esforços para melhorar os níveis de serviço logístico e a gestão em geral, a taxa de ruptura de estoque permanece elevada (AGUIAR; SAMPAIO, 2013).

Embora as taxas de ruptura de estoque possam variar consideravelmente entre diferentes varejistas, cada um com suas características próprias, a maioria delas apresenta taxas de ruptura entre 5 e 10% (PAPAKIRIAKOPOULOS; PRAMATARI; DOUKIDIS, 2009; USMAN; CAPLICE, 2008). Dependendo do cenário, também não é incomum encontrarmos taxas mais elevadas, podendo chegar até a 15% (BERGER, 2003). Estes níveis elevados possuem impacto direto no desempenho das empresas. Apenas em termos de faturamento, estes índices podem resultar em perdas de aproximadamente 1,5% de suas receitas totais (PAPAKIRIAKOPOULOS; PRAMATARI; DOUKIDIS, 2009).

Neste cenário, tanto a literatura nacional quanto a internacional têm se concentrado em duas áreas principais de pesquisa: o comportamento do consumidor diante da ruptura de estoque e as causas subjacentes da própria ruptura. No campo do marketing, os pesquisadores têm se dedicado a investigar os efeitos negativos e as reações dos consumidores diante da indisponibilidade de um produto (PAPAKIRIAKOPOULOS, 2012). O comportamento do consumidor ajuda a entender e dimensionar as consequências futuras através de suas reações. Essas reações abrangem desde a mudança de compra até a troca do local de compra (BARRICHELLO et al., 2019; HAUSRUCKINGER, 2006; MONTOYA; GONZALEZ, 2019; PAPAKIRIAKOPOULOS; PRAMATARI; DOUKIDIS, 2009), o que reforça a relevância do tema para o setor varejista.

No que diz respeito às causas da ruptura de estoque, estudos têm identificado problemas na cadeia de abastecimento e no processo de reposição como fatores contribuintes (AGUIAR; SAMPAIO, 2013; BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006; COELHO et al., 2015; HAUSRUCKINGER, 2006). Além disso, também são apontadas como causas projeções de demanda equivocadas devido à sazonalidade ou distorções dos dados históricos, inclusive resultantes de rupturas anteriores (ANDAUER; RUZ; GOYCOOLEA, 2021; LUISA; SILVA, 2008). Esta linha de pesquisa é essencial para orientar melhoria nos processos e buscar a minimização de índices de ruptura.

Apesar da relevância do assunto e da própria retroalimentação de dados que possibilita a recorrência de ruptura nos mesmos itens, poucos estudos têm se dedicado à identificação e mensuração deste fenômeno. O aumento no número de itens disponíveis no varejo demanda o desenvolvimento de técnicas apropriadas (HAUSRUCKINGER, 2006). No entanto, a grande maioria das empresas ainda utiliza técnicas manuais de verificação, que vão desde a auditoria visual do ponto de venda até a perguntas diretas aos clientes (COELHO et al., 2015; GRUEN; CORSTEN, 2007).

Com o objetivo de avançar neste contexto, alguns estudos têm se concentrado em métodos alternativos de monitoramento, como a utilização de etiquetas de radiofrequência (RFID) (HEESE, 2007; UÇKUN; KARAESMEN; SAVAŞ, 2008) ou a análise de imagens do ponto de venda (ANDAUER; RUZ; GOYCOOLEA, 2021). Alternativas estas que exigem investimentos consideráveis para implementação. Nesse sentido, pesquisas que empregam métodos analíticos, com base em informações de estoque e dados de vendas, têm apresentado resultados promissores (VASCONCELOS E SAMPAIO, 2009).

A presente pesquisa assume relevância ao ampliar o escopo dos estudos que se baseiam em métodos analíticos, visando superar limitações identificadas em pesquisas anteriores. A presença de um elevado número de inconsistências nos registros de inventário constitui um desafio relevante, uma vez que isso inviabiliza a utilização desses dados como fonte confiável para a identificação de rupturas (PAPAKIRIAKOPOULOS; PRAMATARI; DOUKIDIS, 2009). Além disso, as avaliações de eficácia dos métodos existentes dependem de confirmações manuais dos alertas gerados. Esse processo pode apresentar falhas operacionais ou resultar em omissões

de informações que podem comprometer a avaliação de eficiência das equipes envolvidas (CHUANG; OLIVA; LIU, 2016). No caso de terceirização dessa etapa de confirmação manual, os custos envolvidos nem sempre estão alinhados com os resultados obtidos (CHUANG; OLIVA; LIU, 2016).

Diante do exposto, a definição de um método baseado em dados transacionais, sem a utilização de registros de inventário e que não necessite de interferência ou mesmo conferência humana mostra-se relevante no ambiente acadêmico. Ao permitir uma medição mais ampla e menos distorcida por motivações pessoais, os níveis atualizados podem auxiliar em pesquisas futuras que visem a diminuição de tais índices. Além de propiciar uma análise de ocorrências de forma mais rápida, o modelo deve ser capaz de avaliações retroativas, permitindo que estudos posteriores façam comparativos com situações passadas, sem a necessidade de levantamentos preliminares.

O estudo também se mostra útil para o ambiente empresarial, propiciando o aprimoramento das operações do varejo alimentar. O método, ao identificar rapidamente um produto indisponível, auxilia na redução do tempo que este permanece em ruptura, o que, conseqüentemente, representa uma tendência de diminuição nos níveis gerais de indisponibilidade. Além disso, ao identificar itens em ruptura, o método facilita melhorias futuras, como o aperfeiçoamento das métricas de previsão de demanda, proporcionando uma gestão mais eficiente dos estoques.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo apresentam-se quatro seções essenciais para a pesquisa. A primeira traz os principais conceitos da Ruptura de Estoque, tema principal do estudo. A segunda e terceira seções tratam sobre acuracidade de estoque e a Previsão de Demanda, que são fundamentais para a construção de um modelo adequado de identificação de rupturas de estoque. Já a quarta seção, traz conceitos de *Data Mining* e *Soft Computing*, como ferramentas úteis para a construção do modelo.

2.1 Ruptura de Estoque

No âmbito do varejo, diversos estudos têm se dedicado a abordar a problemática da ruptura de estoque sob a perspectiva do cliente. Nessa abordagem, a ruptura de estoque é caracterizada como uma situação na qual um item regularmente disponibilizado para venda em um local específico das prateleiras não está disponível para o cliente no momento da compra (VASCONCELLOS; SAMPAIO, 2009).

Na literatura internacional, a ruptura de estoque, ou *out-of-stock* (OOS) é abordada em um conceito mais amplo do que a mera falta de estoque. O termo OOS engloba os casos nos quais o produto está presente na loja (ou seja, não há falta de estoque), porém não está disponível, por algum motivo, na posição correta da prateleira. Dessa forma, uma falta de estoque sempre implica em OOS, mas o contrário nem sempre é verdadeiro (PAPAKIRIAKOPOULOS; PRAMATARI; DOUKIDIS, 2009).

Neste contexto, autores como Gruen e Corsten (2007) diferenciam as formas pelas quais ocorre a ruptura de estoque. A ruptura na gôndola ocorre quando o produto não está disponível para o cliente, mas está presente na área de estoque da loja. Por outro lado, a ruptura na loja ocorre quando não há produto em estoque na própria loja. Além disso, a ruptura no centro de distribuição ocorre quando o produto está em falta tanto na loja quanto no centro de distribuição responsável pelo abastecimento da mesma.

Embora a definição de OOS possa parecer evidente, ela revela limitações quando considerada a perspectiva dos vendedores, sejam eles fabricantes ou

varejistas. Essa definição não incorpora o impacto econômico das vendas perdidas decorrentes da ruptura, nem leva em conta elementos como frequência, duração ou a importância do item em questão (BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006). Devido a esses fatores, juntamente com a importância econômica do item, o nível de vendas perdidas é utilizado para classificar o tipo de ruptura em duas categorias distintas. Por estes motivos, além da importância econômica do item, nível de vendas perdidas classificam o tipo de ruptura em duas categorias. Enquanto a ruptura total refere-se à falta do produto ao longo de um dia inteiro, a ruptura parcial corresponde à falta do produto em um ou mais períodos durante o dia (BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006). Dessa forma, enquanto a ruptura total resulta em vendas iguais a zero para o item no dia, a ruptura parcial implica apenas em um volume de vendas abaixo do esperado.

2.1.1 Reação do consumidor

A literatura aborda amplamente as reações dos consumidores diante da falta de produtos, com foco nas pesquisas desenvolvidas na área de marketing. Esses estudos investigam as diversas respostas dos consumidores a essa situação, destacando que tais reações variam consideravelmente e são influenciadas por variáveis situacionais, psicográficas, bem como por características específicas do produto e da loja em questão (BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006).

No estudo realizado por Papakiriakopoulos (2011) foram identificadas as reações mais comuns dos consumidores diante da falta de produtos. Uma dessas reações é a mudança do item de compra, que pode ocorrer ao optar por outro formato ou tamanho do mesmo produto, mantendo a mesma marca, ou ao escolher um produto similar de outra marca. Em alguns casos, o consumidor pode até mesmo decidir mudar a categoria do produto, buscando uma alternativa diferente. Outra possível reação é o adiamento da compra para uma visita futura, e em situações mais extremas, o consumidor pode desistir efetivamente da compra. Para os varejistas, os casos mais preocupantes envolvem os consumidores que optam por mudar de loja para encontrar o produto desejado. Nesses casos, existe a possibilidade de que essa mudança seja permanente, resultando na perda do cliente para a empresa.

As diferentes reações dos consumidores diante da ruptura de estoque não resultam em um aumento proporcional na chance de compra de outros produtos. Embora haja um aumento na escolha de produtos similares, a falta de um produto específico não implica necessariamente em um aumento significativo na probabilidade de compra dos demais produtos disponíveis (DIELS; WIEBACH; HILDEBRANDT, 2013).

A ocorrência de rupturas de estoque desencadeia diversas situações que podem comprometer os resultados e a imagem da empresa no mercado. Quando um cliente não encontra o produto desejado disponível na prateleira, é possível que ele acabe adquirindo um produto similar, porém de menor valor. Essa substituição pode levar a uma série de consequências negativas para a empresa, como a redução de receita, a distorção da demanda devido à substituição de produtos, a percepção negativa em relação à empresa e até mesmo a perda de clientes (COELHO et al., 2015).

2.1.2 Causas da ruptura de estoque

Assim como as próprias definições de ruptura de estoque são amplas, suas causas também podem ser as mais diversas. No caso em que um item possui estoque na loja, mas não está disponível no ponto de venda, as causas geralmente são relacionadas a questões operacionais. No entanto, quando ocorre a falta efetiva de estoque, as razões por trás da ruptura podem ser muito mais complexas e abrangentes (KUCUK, 2004).

No primeiro caso, embora a empresa possua o produto, as consequências para o negócio são semelhantes às observadas em casos de ruptura de estoque. Além dos problemas operacionais, existem outros fatores que podem ter as mesmas consequências, ou seja, impactar negativamente as vendas. Por exemplo, quando o cliente encontra dificuldades físicas para obter o produto, como obstruções ou problemas de visualização na gôndola, isso pode levar à perda de vendas e configurar um cenário de ruptura de estoque. Da mesma forma, quando o produto não está no local usual de exposição ou quando os itens restantes estão em condições pouco atrativas, como embalagens danificadas, isso também pode resultar em uma situação

com consequências semelhantes às da ruptura de estoque. (BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006).

Já para os casos em que o produto realmente não possui estoque físico na loja, as causas são muito mais amplas. Essas causas são impactadas até mesmo pela estrutura da própria empresa e sua cadeia de abastecimento. Causas de ruptura de estoque de operações com central de distribuição (CD), são diferentes de operações sem CD. Neste caso, as rupturas podem ser decorrentes de práticas ineficientes nas operações de loja, no centro de distribuição do varejista ou mesmo no próprio fornecedor (AGUIAR; SAMPAIO, 2013)

Além dos problemas operacionais, Gruen e Corsten (2007) e Kucuk (2004) listam outras causas importantes que contribuem para a ocorrência de rupturas de estoque:

- Frequência de compra do produto em desacordo com a eficiência da cadeia de suprimentos: quando a frequência de compra de um produto não está alinhada com a capacidade de abastecimento da cadeia de suprimentos, pode ocorrer a falta de estoque, resultando em rupturas;
- Grande número de *stock keeping units* (SKUs) ou unidade de manutenção de estoque: quanto maior for o número de SKUs (unidades de manutenção de estoque), maior será a complexidade do processo de reabastecimento. Isso pode dificultar o gerenciamento adequado do estoque, reduzir o espaço disponível para exposição e armazenamento de cada produto individualmente e aumentar as chances de ocorrerem rupturas;
- Informações imprecisas do ponto de venda: a falta de informações precisas sobre as vendas e estoques no ponto de venda pode levar a problemas na previsão de demanda e na geração de novos pedidos. Isso pode resultar em rupturas de estoque devido à falta de reposição adequada e oportuna;
- Perdas inerentes a danos ou furtos: danos aos produtos durante o transporte, armazenamento ou exposição, bem como furtos, podem resultar na redução do estoque disponível para venda e, conseqüentemente, em rupturas de estoque;
- Promoções não programadas ou mal dimensionadas: promoções repentinas, não planejadas ou mal dimensionadas podem causar um

aumento repentino na demanda por determinado produto. Se o estoque não estiver preparado para atender a essa demanda adicional, pode ocorrer uma ruptura de estoque.

Além destes, alguns fatores externos ao varejo também contribuem para a existência de rupturas. O atraso na entrega dos produtos por parte dos fornecedores é apontado como uma das principais causas de rupturas de estoque. Quando os fornecedores não conseguem entregar os produtos dentro do prazo esperado, ocorre a falta de mercadorias nas lojas, o que pode resultar em rupturas. Além disso, a reposição do estoque nas lojas muitas vezes é realizada por repositores externos, que são responsáveis por abastecer as prateleiras com os produtos. A ineficiência nesse processo de reposição pode levar a atrasos, erros ou até mesmo omissões na reposição adequada dos produtos, o que contribui para a ocorrência de rupturas de estoque (COELHO et al., 2015).

2.1.3 Identificação e Medição

A disponibilidade na prateleira é um indicador crucial de desempenho no setor varejista. Ao longo do tempo, várias técnicas têm sido utilizadas para verificar essa disponibilidade. Essas técnicas evoluíram ao longo do tempo de acordo com o número de produtos monitorados e o desenvolvimento de melhorias tecnológicas (MONTROYA; GONZALEZ, 2019).

A avaliação de rupturas de estoque é comumente realizada por meio de auditorias de loja, que consistem em inspeções visuais do ponto de venda. No entanto, esse tipo de processo tem limitações. Ele avalia a disponibilidade (ou falta) do produto apenas momentaneamente, não sendo adequado para avaliar a duração da falta de estoque observada. Além disso, esse processo está sujeito a erros de medição, uma vez que é baseado na observação e registro humano. Em alguns casos, os gestores adotam medidas para evitar a existência de espaços vazios que podem passar despercebidos por estas auditorias, o que diminui a taxa de identificação. Além disso, essa abordagem manual também tem limitações físicas, pois geralmente se restringe a amostras limitadas de categorias de produtos e de lojas específicas (BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006).

Em conjunto com o método de detecção visual, são empregadas outras formas manuais para verificar a disponibilidade dos produtos. Algumas empresas realizam entrevistas diretas com os clientes, questionando-os sobre os produtos não encontrados. Além disso, elas solicitam informações aos seus fornecedores sobre indisponibilidades, relacionadas às entregas ou às rupturas específicas nas lojas onde são responsáveis pelo abastecimento. Outras verificações, que embora possam ser automatizadas, ainda são consideradas manuais, envolvem a simples verificação de vendas igual a zero durante um determinado período ou a identificação de estoques zerados (MONTROYA; GONZALEZ, 2019).

Além da auditoria manual, existe uma crescente utilização de métodos analíticos de detecção (BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006; HAUSRUCKINGER, 2006). Esses métodos utilizam técnicas de análise de dados de vendas e estoque para identificar padrões e indicadores de indisponibilidade de produtos. Por meio da análise dos dados de vendas, é possível identificar períodos em que o volume de vendas apresenta comportamento atípico. Já informações relativas ao nível de estoque são utilizadas para identificar produtos que apresentam um registro de estoque igual a zero no sistema, ou que não apresentam movimentações consistentes com sua realidade durante um determinado período de tempo (VASCONCELLOS; SAMPAIO, 2009).

Outras alternativas para a identificação de rupturas de estoque também são desenvolvidas. Ampliando os métodos analíticos, técnicas de *machine learning*, através de aprendizado supervisionado surgem como uma alternativa (ROSADO et al., 2016). Outros métodos, porém, ainda pouco viáveis economicamente, também já são objeto de estudos. O mais popular faz o uso de etiquetas de radiofrequência (RFID), que além de indicar de forma segura a presença física ou não do produto, são úteis também para a acuracidade do estoque (KUCUK, 2004). A utilização de sensores de peso nas prateleiras e o processamento de imagens, embora dependam de fatores iniciais de configuração, também apresentam resultados promissores (BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006).

Aliado à problemática da definição e da identificação da ruptura de estoque, o próprio índice calculado sofre questionamentos. Embora usualmente a taxa de ruptura relacione os itens efetivamente disponíveis com o total de itens comercializados, essa abordagem pode conter algumas limitações. Desta forma, a taxa de ruptura de uma

loja l em um momento específico t , definida por Gruen e Corsten (2007) através da equação (1), pode omitir aspectos importantes na quantificação e no impacto real da indisponibilidade de produtos (LAMBERT; STOCK; ELLRAM, 1998).

$$\text{Taxa de Ruptura} = 1 - \frac{\text{Número de SKUs nas gôndolas}_{t,l}}{\text{Total de SKUs comercializados pela loja}_{t,l}} \quad (1)$$

Esta limitação, levou Hausruckinger (2006) a propor uma taxa de ruptura calculada com base nas vendas perdidas decorrentes da indisponibilidade do produto e nas vendas perdidas:

$$\text{Taxa de Ruptura} = \frac{\text{Vendas perdidas (em unidades) por período}}{\text{Vendas esperadas (em unidades) por período}} \quad (2)$$

De forma similar, Bayle-Torutoulou et al. (2006), propõem medir a ruptura de estoque apenas em termos absolutos, pela diferença entre as vendas esperadas e as vendas efetivamente realizadas:

$$\text{Vendas perdidas} = \text{Vendas esperadas} - \text{Vendas realizadas} \quad (3)$$

Embora as taxas de ruptura definidas pelas equações (2) e (3) sejam alternativas para as limitações da definição usual, as mesmas trazem outra limitação importante. Ambas as equações se baseiam em estimativas de vendas, o que requer um alto grau de confiabilidade nas previsões de vendas esperadas. Como as próprias demandas geradas já são influenciadas por rupturas anteriores, um índice baseado nestas projeções já possui sua assertividade comprometida (MONTROYA; GONZALEZ, 2019).

2.2 Acuracidade de Estoque

Os avanços nos sistemas de informação têm proporcionado uma maior facilidade na medição dos níveis de estoque. No entanto, é comum ocorrer um certo percentual de inconsistências entre o estoque físico real e o registro no sistema (AGUIAR; SAMPAIO, 2013; USMAN; CAPLICE, 2008). Essas imprecisões no registro

de inventário (IRI), que consistem em divergências entre os estoques reais e os registros de estoque, prejudicam as operações das empresas, principalmente porque afetam a reposição de estoque, podendo resultar em níveis desnecessariamente altos de estoque ou em rupturas (CHUANG; OLIVA; LIU, 2016).

Para Drohomeretski e Favaretto (2010), os efeitos da falta de acuracidade do estoque escondem custos que são gerados por erros operacionais e até roubos. A diminuição da produtividade, o baixo nível de serviço e expedições emergenciais de mercadorias, são custos muitas vezes não contabilizados dentro das divergências de estoques.

Os efeitos e os custos (diretos ou indiretos) da falta acuracidade dos estoques varia de acordo com as características operacionais da empresa e dos próprios níveis destas divergências. O estudo de Dehoartius et al. (2008) lista os principais fatores que impactam no aumento das inconsistências de estoque:

- Volume de vendas: quanto maior o volume de vendas do produto maior é a imprecisão dos registros de inventário;
- Registro de saída incorreta do produto: produtos similares movimentados como um único produto;
- Valor do produto: produtos mais baratos tendem a apresentar menor acuracidade do que produtos mais caros, tendo em vista que os produtos mais caros possuem maior controle;
- Fonte de fornecimento: produtos recebidos de centros de distribuição próprios têm maior chance de serem entregues com erro do que materiais recebidos de terceiros;
- Variedade de produtos: erros nos registros aumentam conforme aumenta a variedade de produtos, podendo ocasionar erros na expedição e movimentação de materiais;
- Intervalo de tempo para realização do inventário: quando maior for o intervalo, maior será a divergência de estoque.

Aliado a isso, problemas operacionais também possuem papel importante nas inconsistências de estoque (CHUANG; OLIVA, 2015). Falta de treinamento do pessoal, que leva a retiradas de material sem autorização ou registros incorretos de

entrada e saída dos produtos representam uma das principais causas das divergências. Aliado a isso, as perdas também contribuem significativamente com as inconsistências. Essas perdas podem ser oriundas de roubos internos e externos ou de produtos danificados, especialmente durante sua movimentação (DROHOMERETSKI; FAVARETTO, 2010).

Neste contexto, alguns estudos já consideram as inconsistências de estoque como parte do processo. Desta forma, passam a assumir que os erros de registro crescem, independentemente dos níveis de inventário, mas proporcionalmente às movimentações. Isso permite que sejam utilizadas técnicas estatísticas suficientes para minimizar a distribuição de erro, tornando o registro mais semelhante à sua realidade (KÖK; SHANG, 2007).

2.3 Previsão de demanda

Previsão de demanda é uma estimativa provável e plausível de um valor futuro de uma grandeza de interesse. Este tipo de cálculo somente é necessário quando se está diante de um cenário de incertezas a respeito do futuro. Assim, tomadas de decisão envolvendo resultados empresariais normalmente utilizam métodos formais de previsão de demanda (ARMSTRONG, 2001).

A maioria dos métodos de previsão de demanda está baseado na premissa de que os padrões ou os comportamentos passados serão refletidos no futuro. Desta forma, padrões do passado são modelados e projetados para o futuro (FRANK et al., 2003).

Para Armstrong (2001), as principais etapas de uma previsão de demanda são: formular o problema, obter dados, selecionar o método de previsão, implementar o método, avaliar o método, e por fim utilizar a previsão de demanda.

A literatura classifica os métodos existentes de previsão de demanda em quantitativos e qualitativos (HANKE; WICHERN, 2013; MAKRIDAKIS, 1993). Técnicas qualitativas não requerem manipulação de dados: somente julgamentos são utilizados para criar a previsão. Técnicas quantitativas não necessitam de julgamento: são procedimentos padronizados que produzem resultados objetivos (HANKE; WICHERN, 2013)

Os métodos qualitativos são considerados métodos de predição, baseados no julgamento, intuição ou experiência dos envolvidos. De acordo com Frank et al. (2003), estes métodos são úteis quando existem poucos dados para apoiar métodos quantitativos (previsão de novos produtos e novas tecnologias, por exemplo). Desta forma, geralmente são usados para criar cenários futuros, quando apenas poucos dados relevantes para o objetivo da previsão estão disponíveis, como padrões e relações causais.

Já os métodos quantitativos de previsão utilizam padrão histórico de dados para extrapolar o comportamento futuro. Ainda de acordo com Frank et al. (2003), estes métodos utilizam ferramentas quantitativas, como séries temporais e modelos causais, além de inteligência artificial para realizar as previsões. Enquanto as séries temporais utilizam apenas o tempo como variável, os modelos causais estabelecem relações causa-efeito em um sistema, através de uma análise multivariada.

2.4 Data Mining e Soft Computing

O desenvolvimento de *hardware* e *software* propicia uma coleta e armazenagem de dados cada vez maior. Técnicas tradicionais que unem ferramentas de gerenciamento de dados com ferramentas estatísticas já não são adequadas para analisar a vasta coleção de dados, especialmente no ambiente empresarial (MITRA; PAL; MITRA, 2002).

A mineração de dados ou *data mining* pode ser definida como uma forma de garimpagem das informações relevantes, através da aplicação de algoritmos específicos. De acordo com Palestino (2001), o *data mining* trata-se de uma forma de produção de novas informações a partir dos dados já existentes, através da descoberta de padrões das informações transacionais armazenadas.

Tecnicamente, *data mining* é um conjunto estruturado de processos de exploração e análise de grandes volumes de dados, por meio de métodos automáticos e semiautomáticos, com o intuito de descobrir e determinar padrões, tendências, relacionamentos e regras, que permitam converter tais dados em conhecimento de alto valor estratégico (SINGH et al., 2011).

Por sua característica, a mineração de dados tornou-se uma forma de descoberta essencial para resolver problemas em diferentes domínios. A partir de

dados coletados originalmente com outras finalidades, novos conhecimentos podem ser obtidos no processo, diminuindo tempo e custos para estas descobertas (MITRA; PAL; MITRA, 2002).

Neste sentido, a mineração de dados se utiliza de diversas ferramentas, como inteligência artificial e ferramentas estatísticas e de visualização, para extrair informações anteriormente desconhecidas, inesperadas e potencialmente úteis a partir dos dados (GORDON; GORDON, 2000).

Seguindo a trilha de técnicas de *data mining*, os modelos de *soft computing* desempenham um papel importante no campo do reconhecimento, classificação e previsão de dados em vários campos de aplicação. Enquanto as técnicas de mineração de dados incluem regras de associação, agrupamentos e árvores de decisão para extrair informações das bases de dados, métodos de *soft computing*, se utilizam destas técnicas para extração de informações ocultas nestes dados.

A principal característica da computação flexível ou *soft computing* é a exploração da tolerância à imprecisão e à incerteza para alcançar a rastreabilidade e a robustez, com um baixo custo para a solução. Apesar das características, não se trata de um método simples, mas de uma combinação de diversos métodos (IBRAHIM, 2016). Na verdade, trata-se de uma espécie de consórcio de metodologias que fornecem capacidade flexível de processamento de informações para lidar com situações ambíguas da vida real (MITRA; PAL; MITRA, 2002).

Embora existam inúmeras metodologias dentro das técnicas de *soft computing*, as principais são divididas em (SINGH et al., 2011):

- Lógica difusa (*fuzzy logic*): é uma abordagem da computação que se aproxima da notável capacidade da mente humana de raciocinar e aprender em um ambiente de incerteza e imprecisão;
- Redes neurais (*Artificial Neural Networks*): inspirado no sistema nervoso e em como o cérebro funciona.
- Algoritmos genéticos: modelo inspirado no processo de evolução da natureza, onde existe um mecanismo de busca adaptativa baseado em informações de custo e recompensa.
- Sistemas especialistas: sistemas baseados em regras, muito dependentes da base de conhecimento disponível.

METODOLOGIA

Neste capítulo são descritos os procedimentos metodológicos utilizados no estudo, com a finalidade de responder ao questionamento da pesquisa, bem como satisfazer os objetivos geral e específicos. Primeiramente será detalhada a coleta e o tratamento dos dados e, em seguida, serão descritas as técnicas utilizadas para desenvolvimento do modelo proposto.

2.5 Classificação da Pesquisa

A classificação da pesquisa presente no estudo é a seguinte: quanto à natureza uma pesquisa aplicada, quanto a abordagem quantitativa, quanto aos objetivos explicativa e quanto ao procedimento técnico empregado estudo de caso.

2.6 Dados: coleta e preparação

A pesquisa foi conduzida como um estudo de caso em uma rede supermercadista localizada na região da Serra Gaúcha. A rede é composta por 9 lojas e uma central de distribuição.

Os dados utilizados no estudo foram extraídos diretamente do banco de dados do *Enterprise Resource Planning* (ERP) da rede supermercadista. O acesso irrestrito diretamente aos dados transacionais, propiciou maior facilidade na obtenção, pois independe da construção de interfaces dentro do próprio ERP. Além disso, facilitou a verificação de possíveis inconsistências nos dados a serem tratados, uma vez que estes foram conciliados com relatórios gerenciais específicos.

A empresa possui um banco de dados PostgreSQL, com registros que iniciam em janeiro de 2009, propiciando uma grande quantidade de informações a serem extraídas e tratadas para os fins do estudo. Para esta extração e tratamento dos dados, utilizamos a linguagem Python, devido à sua extensa biblioteca voltada para manipulação de dados, especialmente em casos que envolvem uma grande massa de informações (MACKINNEY, 2018). Após o processamento, os dados foram armazenados em um banco de dados MySQL externo à organização. Isso permite a manipulação e análise dos dados sem impacto nos sistemas internos da empresa, além de garantir a imutabilidade dos dados históricos.

Nesta pesquisa, os dados analisados ficaram restritos à filial de maior faturamento da rede supermercadista durante o período de 01/06/2022 a 31/05/2023. O estudo considerou um período de 12 meses, pois abrange todas as sazonalidades presentes nos supermercados e proporciona uma análise abrangente e robusta.

2.6.1 Seleção de produtos

A empresa possui em seu cadastro um total de 12.396 itens ativos, e estes foram efetivamente movimentados no período da análise. A média mensal de vendas da filial ao longo dos 12 meses avaliados foi de 8.421 *stock keeping units* (SKUs). Porém, diariamente foram vendidos, em média, apenas 2.312 destes itens, o que demonstra a grande quantidade de itens não movimentados diariamente.

Dada a considerável quantidade de itens, foram estabelecidas restrições quanto ao número de produtos, visando viabilizar o processamento e a análise eficaz dos dados. Para a amostra, foram selecionados apenas produtos que cadastrados pelo menos um ano antes da data inicial de análise (01/06/2021) e que permaneceram ativos no mix até o final do período (31/05/2023). Este período anterior ao início efetivo da análise foi necessário para a construção de um histórico de movimentação que proporcionasse uma análise mais confiável.

Itens altamente perecíveis, como hortifrutigranjeiros, carnes e produtos de padaria, foram excluídos da análise devido aos altos índices de perdas e à reposição frequente. Além disso, produtos de bazar, como linha têxtil, utilidades domésticas, brinquedos, ferramentas e vidraçaria, bem como produtos específicos para datas como Natal e Páscoa, também não foram considerados, devido às baixas frequências de venda e às sazonalidades específicas que não estão dentro do escopo deste estudo.

Estas restrições foram implementadas com o objetivo de obter uma amostra mais apropriada para a análise dos dados, resultando em uma redução do número de itens a serem analisados para 2.022. Embora essa quantidade represente apenas 16% dos itens ativos, ela corresponde a 24% dos itens vendidos mensalmente e respondem por 37% do faturamento total no período. Como as restrições não consideraram itens sazonais, a amostra pode ser considerada mais adequada para refletir o número mensal de SKUs movimentados.

Em relação às características dos itens selecionados na filial, é importante observar que todos são abastecidos pela central de distribuição da empresa, o que torna esta variável irrelevante em nossa análise. No entanto, o intervalo de reposição desses produtos é uma importante variável a ser considerada tanto para os níveis de ruptura quanto para o impacto no modelo de identificação proposto.

Os itens analisados foram extraídos do ERP com informações sobre a sua estrutura mercadológica. A estrutura mercadológica adotada pela rede de supermercados para a análise dos itens é composta por um sistema de categorização em quatro níveis. Essa estrutura permite agrupar os produtos de acordo com critérios que aumentam a similaridade à medida que se avança nos níveis de classificação.

Tabela 1 – Produtos

Campo	Descrição
Código	Código do produto utilizado no ERP
Descrição	Descrição do produto
Departamento	Nível 1 da classificação
Categoria	Nível 2 da classificação
Subcategoria	Nível 3 da classificação
Segmento	Nível 4 da classificação

Fonte: Elaborado pelo autor.

Cada nível da estrutura mercadológica representa uma camada de classificação que visa organizar os produtos de forma mais específica e segmentada. À medida que se avança nos níveis, os critérios utilizados para agrupar os itens se tornam mais refinados e precisos, resultando em categorias mais similares entre si. Além da própria similaridade dos produtos, outros fatores como exposição conjunta e a árvore de decisão do cliente no momento da compra também são considerados na definição dessa estrutura.

2.6.2 Movimentação dos produtos

A partir do banco de dados, foram extraídas e sintetizadas informações relativas a todas as transações dos produtos, com o propósito de otimizar o processo de análise. A extração considerou apenas as operações observáveis que envolvem movimentação real de estoque. Ajustes de inventário e perdas estimadas não foram

incluídos na análise. Desta forma, foram consideradas as operações conforme tabela abaixo:

Tabela 2 – Operações de Estoque

Operações de Saída	Operações de Entrada
Venda (VE)	Compra (CO)
Transferência de Saída (TS)	Transferência de Entrada (TE)
Devolução de Compra (DC)	Devolução de Venda (DV)
Consumo Interno (CI)	Bonificação (BO)

Fonte: Elaborado pelo autor.

Os dados extraídos foram condensados através de uma totalização diária para cada uma das operações. Com exceção das informações relativas às vendas, as demais operações armazenam apenas a data da movimentação, o código do produto e a quantidade movimentada:

Tabela 3 – Registro de Operações (exceto venda)

Campo	Descrição
Data	Data em que ocorreu a movimentação
Código	Código do produto movimentado
Quantidade	Quantidade movimentada pela operação nesta data

Fonte: Elaborado pelo autor.

Na operação de venda, além dos dados existentes em outras operações, foi coletado o valor financeiro total da operação diária na respectiva data. Além destes, registrou-se um indicador de oferta para o item, sinalizando se houve alguma ação promocional no dia:

Tabela 4 – Operação de Venda

Campo	Descrição
Data	Data em que ocorreu a venda
Código	Código do produto vendido
Quantidade	Quantidade da venda na data
Valor	Valor da venda na data
Operações	Número de operações de venda na data
Oferta	Indicador de oferta na data (0 = Não, 1 = Sim)

Fonte: elaborado pelo autor.

Essas informações fornecem uma visão mais abrangente do desempenho do item no contexto da venda diária. O valor total da operação diária permite avaliar o impacto financeiro do item, considerando não apenas a quantidade vendida, mas também o valor monetário gerado por suas transações. Já o indicador de oferta sinaliza se o item estava sendo promovido ou se havia alguma ação promocional em vigor no dia da venda. Isso é relevante, pois as promoções podem influenciar o comportamento do consumidor e impactar as vendas e a demanda pelo item.

2.7 Processamento dos dados

A partir dos dados obtidos e consolidados, foi criada uma tabela reunindo informações relevantes para a análise aprofundada. A tabela contém todos as datas do intervalo analisado, independentemente de haver ou não movimentação neste dia, o dia da semana correspondente, o código do produto, total de entradas e o total de saídas do produto na filial, além do saldo de entradas e saídas desde a última entrada registrada:

Tabela 5 – Saldo da Última Compra

Campo	Descrição
Data	Data da movimentação
Dia da Semana	Dia da semana correspondente à data
Código	Código do produto movimentado
Entradas	Total de entradas na data (quantidade)
Saídas	Total de saídas na data (quantidade)
Saldo Última Entrada	Saldo a partir da última entrada (quantidade)

Fonte: Elaborado pelo autor.

Essa tabela agregada propicia uma análise mais abrangente do fluxo de produtos na filial ao longo do tempo. O dia da semana se faz necessário uma vez que um item pode apresentar grandes variações de movimentação nos diferentes dias da semana. Já a quantidade de entradas e de saídas em determinado dia d (respectivamente $Entradas_d$ e $Saídas_d$), além do saldo a partir da última entrada neste dia ($Saldo_d$), são definidos conforme as equações:

$$Entradas_d = CO_d + TE_d + BO_d \quad (4)$$

$$Saídas_d = VE_d - DV_d + DC_d + CI_d + TS_d \quad (5)$$

onde, para um determinado produto p no dia d , temos:

- CO_d = quantidade de compras;
- TE_d = quantidade de entradas de transferência;
- BO_d = quantidade de entradas;
- VE_d = quantidade de vendas;
- DV_d = quantidade de devolução de vendas;
- DC_d = quantidade de devolução de compra;
- CI_d = quantidade de consumo interno e
- TS_d = quantidade de saídas de transferência.

Conforme observado na equação (4), embora a operação de devolução de vendas (DV) seja considerada como um movimento de entrada, ela não é tratada como uma entrada efetiva. Como se trata de um estorno da operação de venda, ela é considerada apenas como um redutor deste tipo de movimento e, por este motivo, aparece na equação (5), que descreve as saídas.

Já o saldo a partir da última entrada em um determinado dia d ($Saldo_d$), tem seu cálculo reiniciado em toda nova entrada. Desta forma, caso exista entradas do produto neste dia ($Entradas_d > 0$), teremos:

$$Saldo_d = Entradas_d - Saídas_d \quad (6)$$

No caso de não haver entradas em d , ou seja, $Entradas_d = 0$, o saldo será calculado a partir do saldo dia anterior ($Saldo_{d-1}$), conforme a equação:

$$Saldo_d = Saldo_{d-1} - Saídas_d \quad (7)$$

O saldo da última entrada é utilizado como alternativa aos registros de inventário. Sua utilização se justifica pelo menor índice de divergências quando consideramos um menor intervalo de tempo (DEHORATIUS; MERSEREAU; SCHRAGE, 2008), o que pode demonstrar, em tese, a existência de estoque do item.

Neste sentido, ainda de acordo com Dehoratius e Raman (2008), o intervalo de reposição também influi sobre as inconsistências de estoque e, conseqüentemente,

nos casos de ruptura. Para esta análise, foi criada uma tabela contendo as informações referentes ao prazo médio de recebimento de cada produto no intervalo de 01/06/2021 a 31/05/2023:

Tabela 6 – Prazo Médio de Reposição

Campo	Descrição
Código	Código do produto movimentado
Reposição	Prazo médio de reposição (em dias)

Fonte: Elaborado pelo autor.

Nesta tabela, para cada produto analisado, foi informado o prazo médio de reposição do item na filial. Este prazo médio foi calculado como a média de dias entre cada recebimento do item durante o período considerado. Ele serve como uma referência não apenas para avaliar os índices de ruptura, mas também para a precisão do modelo proposto.

Baseado no estudo de Papakiriakopoulos et al. (2009), a avaliação das vendas acumuladas em um intervalo também auxilia na verificação de existência de ruptura. Além da própria ruptura total, objeto do estudo, a diminuição das vendas acumuladas em um período pode indicar também a existência de rupturas parciais, que muitas vezes precedem a ruptura total. Pelas características das vendas de supermercados, que se mostram semelhantes nos mesmos dias da semana, as vendas são consolidadas em um período de 7 dias:

Tabela 7 – Venda Acumulada

Campo	Descrição
Data	Data da movimentação
Código	Código do produto movimentado
Quantidade	Total de vendas acumulada período (quantidade)
Oferta	Indicador de oferta no período (0 = Não, 1 = Sim)

Fonte: Elaborado pelo autor.

O campo oferta indica se houve algum tipo de oferta no item no período de 7 dias. Já a quantidade ($VE Acum_d$) representa a venda física acumulada nos últimos 7 dias, propiciando a verificação da constância de vendas do produto. Desta forma, em um determinado dia d , a venda acumulada nos últimos 7 dias ($VE Acum_d$) é dada por:

$$VE\ Acum_d = VE_d + VE_{d-1} + \dots + VE_{d-6} \quad (8)$$

onde, VE_d representa a quantidade vendida de um determinado produto p e um dia d .

2.8 Modelo para Identificação

O objetivo do estudo é a identificação de ocorrências de rupturas totais de estoque pela falta efetiva do produto. Neste cenário, o produto não está disponível no ponto de venda e a causa não está relacionada a falhas na operação de reposição, mas sim à ausência do estoque físico. Para isso, foi desenvolvido um algoritmo baseado em lógica difusa.

A lógica difusa é a abordagem adotada por lidar com valores imprecisos que podem variar entre verdadeiro e falso. Nesse cenário, em que buscamos identificar a existência ou não de rupturas de estoque para um determinado item, é necessário definir os eventos que serão considerados durante a avaliação. Além dos próprios eventos, seus valores e sua forma de associação com a saída esperada também possuem papel importante nos resultados obtidos.

2.8.1 Variáveis de entrada

O primeiro passo para a identificação deste tipo de ruptura é a inexistência de venda (VE) na data avaliada (d). Desta forma, se a venda na data for maior do que zero ($VE_d > 0$), a verificação já não se faz necessária. Somente quando a venda for igual a zero ($VE_d = 0$) existe a possibilidade de ocorrência de ruptura de estoque. No entanto, é importante considerar que alguns itens podem não apresentar vendas mesmo quando estão disponíveis para o cliente. Para lidar com esta situação, é necessário avaliar a probabilidade de não haver venda para um determinado item. A fim de aumentar a precisão das análises desenvolvidas por Chuang, Oliva e Liu (2016) e também por Papakiriakopoulos (2011), que consideraram uma média diária de vendas geral, optou-se pelo cálculo da média de vendas para cada um dos diferentes dias da semana (PAPAKIRIAKOPOULOS; PRAMATARI; DOUKIDIS, 2009).

Ainda baseado em Papakiriakopoulos (2011), que partiu de literaturas anteriores (AXSÄTER, 2000; HILL, 1999) e utilizou a distribuição de Poisson para

cálculo da probabilidade de venda nula, observamos um número de itens significativo que segue a distribuição normal. De acordo com este estudo, 63% dos itens possuíam vendas diárias que seguem a distribuição de Poisson e 37% seguem uma distribuição Normal. Embora seja válida a opção única pela distribuição de maior representatividade, o percentual significativo de itens mais adequados à distribuição Normal, mostrou ser válida a adoção da distribuição de maior similaridade para cada um dos itens analisados.

Para esta verificação, foram empregados os dados do ano anterior ao período efetivo de análise, ou seja, de 01/06/2021 a 31/05/2022. Este período, por apresentar características semelhantes de sazonalidade ao período analisado, representa de forma significativa os dados em estudo. Por este motivo, serviram como base para identificar a melhor adequação entre as distribuições teóricas examinadas.

O modelo de distribuição que melhor descreve os dados de venda diária de cada produto foi encontrado através do teste de *Kolmogorov-Smirnov (K-test)*. Esse teste baseia-se na medida da diferença entre duas distribuições cumulativas de probabilidades empíricas (CDFs), através da equação:

$$D = \max|F_1(x) - F_2(x)| \quad (9)$$

onde, $F_1(x)$ é a CDF empírica da primeira amostra, $F_2(x)$ é a CDF empírica da segunda amostra e o valor de D representa a maior discrepância entre as duas CDFs, ou seja, a maior diferença entre as probabilidades cumulativas atribuídas aos mesmos valores de x nas duas amostras.

Desta forma, o *K-test* compara duas amostras para verificar se elas provêm da mesma distribuição, calculando a maior discrepância entre suas funções de distribuição cumulativa e fornecendo um valor p que indica a significância estatística da comparação. Neste caso, foram utilizadas ferramentas estatísticas da linguagem Python para comparar a distribuição das vendas diárias de cada item com as distribuições de Normal e de Poisson. O valor de p com maior significância estatística, definiu o tipo de distribuição que melhor descreve os dados analisados. Porém, não foi verificado se realmente os dados descreviam o tipo de distribuição, apenas a melhor adequação aproximada.

Para os casos em que podemos considerar que as vendas seguem uma distribuição de Poisson, a probabilidade de ocorrência de ocorrer uma venda de x unidades em um determinado dia é dada por:

$$f1(x) = \frac{\mu^x e^{-\mu}}{x!} \quad (10)$$

onde $f1(x)$ refere-se à probabilidade de ocorrência de venda de x unidades, μ é a venda média diária, em unidades, no período e "e" é o número de Euler (2,71828).

Simplificando a equação (9) para o caso específico em que procuramos saber a probabilidade de não ocorrência de vendas, temos:

$$f1(0) = e^{-\mu} \quad (11)$$

Já para os casos em que podemos considerar que as vendas se aproximam de uma distribuição Normal, a probabilidade de ocorrência de uma venda de x unidades é dada por:

$$f1(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (12)$$

onde μ é a venda média, em unidades, no período, σ é o desvio padrão observado e π é o número de Pi (3,14159).

Simplificando a equação (12) para o caso específico em que procuramos saber a probabilidade de não ocorrência de vendas, temos:

$$f1(0) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}} \quad (13)$$

Em ambos os casos, a probabilidade da venda ser igual a zero é calculada para cada um dos diferentes dias da semana. Além disso, os valores utilizados para o cálculo (média de vendas e desvio padrão) é recalculado para cada uma das ocorrências de venda nula, considerando, além do dia da semana, os dados das vendas até a semana anterior. Esta atualização dos valores relativos à média e ao

desvio padrão (quando necessário) auxiliam na adequação da probabilidade diante de mudanças no histórico dos padrões de venda de cada produto.

Assim, diante da probabilidade de ocorrência de venda nula em um dia específico d , para ambos os tipos de probabilidade, podemos definir que a probabilidade de ocorrência de um evento de ruptura total de estoque no dia avaliado a partir das equações (11) e (13):

$$P1_d = 1 - f1_d(0) \quad (14)$$

onde $f1_d(0)$ refere-se à probabilidade de não ocorrência de vendas de um determinado produto em um dia d (equação (11) ou equação (13)) e $P1_d$ é a probabilidade de ocorrência de ruptura de estoque deste produto no dia d .

Posteriormente, foi verificada também a existência de venda nula em dias consecutivos. Ao considerarmos que as probabilidades de ocorrer uma venda igual a zero em dias sequenciais são menores do que quando avaliadas isoladamente, nota-se um aumento da probabilidade de ruptura nestes casos. Desta forma ao reconsiderar as equações (11) e (13) para o cálculo da probabilidade de venda nula em um determinado dia d , podemos estimar a probabilidade deste tipo de sequência para este dia.

Para os casos em que não houve venda em um determinado dia d ($VE_d = 0$), mas houve venda no dia anterior ($VE_{d-1} > 0$), temos:

$$f2_{d_1} = f1_d(0) \quad (15)$$

onde $f1_d(0)$ refere-se à probabilidade de não ocorrência de vendas de um produto em um dia d e $f2_{d_1}$ é a probabilidade de ocorrência de venda nula no primeiro dia de uma possível sequência sem vendas.

Assim, para os casos em que não houve venda nos n dias anteriores, ou seja:

$$VE_{d_n} = VE_{d_{n-1}} = \dots = VE_{d_2} = VE_{d_1} = 0 \quad (16)$$

onde VE_{d_n} refere-se à quantidade diária vendida de determinado produto em um determinado dia d_n de uma sequência de n dias iniciada em d_1 .

Considerando que as ocorrências de vendas nulas em dias sequenciais são eventos independentes, uma vez que, para o modelo, não temos a confirmação das ocorrências anteriores, teremos:

$$f2_{d_n} = f1_{d_n}(0) \times f1_{d_{n-1}}(0) \times \dots \times f1_{d_2}(0) \times f1_{d_1}(0) \quad (17)$$

onde $f2_{d_n}$ refere-se à probabilidade de ocorrência de uma sequência de n dias sem ocorrência de vendas iniciada em $n = 1$ e, por analogia a $f1_{d_1} = f1_d(0)$, temos que $f1_{d_2} = f1_{d+1}(0)$ e assim sucessivamente.

Assim, diante da probabilidade de ocorrência de venda nula em dias consecutivos em um dia d , podemos definir que a probabilidade de ocorrência de um evento de ruptura total de estoque a partir das equações (15) e (17):

$$P2_d = 1 - f2_{d_n} \quad (18)$$

onde $P2_d$ é a probabilidade de ocorrência de ruptura de estoque de um determinado produto em um dia d , indicada pela sequência de n dias sem vendas e $f2_{d_n}$ é descrito pela equação (17).

Em um terceiro momento, verificou-se a venda acumulada nos últimos 7 dias do item analisado (VE_{Acum_d}). Esta venda foi comparada com a venda média do item em intervalos de 7 dias (μ_{Acum}), diferenciando períodos que apresentam o indicativo de existência de oferta dos períodos que não tiveram o item ofertado.

Assim como nas abordagens de Hausruckingner (2006) e de Chuang, Oliva e Liu (2016), a adequação das vendas acumuladas à distribuição de Poisson foi muito significativa. Neste caso, embora 94% dos itens avaliados se adequassem melhor à distribuição de Poisson e apenas 6% à distribuição Normal, optou-se novamente pela verificação da distribuição de maior similaridade para cada um dos itens analisados, seguindo as mesmas condições de avaliação descritas para o caso das vendas diárias.

Para os casos em que podemos considerar que as vendas acumuladas seguem uma distribuição de Poisson, a probabilidade acumulada de ocorrência de uma venda de até x unidades é dada por:

$$f3(x) = \sum_{k=0}^x \frac{e^{-\mu_{Acum}} \mu_{Acum}^x}{x!} \quad (19)$$

onde $f3(x)$ refere-se à probabilidade acumulada ocorrência de vendas até x unidades ($x = VE_{Acum_d}$), μ_{Acum} é a venda média acumulada e "e" é o número de Euler (2,71828).

Já para os casos em que as vendas acumuladas se aproximam de uma distribuição Normal, a probabilidade acumulada é dada por:

$$f3(x) = \frac{1}{2} \left[1 + erf \left(\frac{x - \mu_{Acum}}{\sigma_{Acum} \sqrt{2}} \right) \right] \quad (20)$$

onde $f3(x)$ refere-se à probabilidade acumulada ocorrência de vendas até x unidades ($x = VE_{Acum_d}$), μ_{Acum} é a venda média acumulada, σ_{Acum} é o desvio padrão da amostra das vendas acumuladas analisada e erf é uma função de erro dada por:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (21)$$

Em ambos os casos, apesar dos cálculos relativamente complexos envolvidos, utilizamos bibliotecas estatísticas da linguagem Python. Essas bibliotecas fornecem, de maneira simples, os valores das probabilidades acumuladas para cada tipo de distribuição. Isso facilita a obtenção dessas informações e agiliza o processo de análise estatística.

Temos dois cenários são possíveis:

$$VE_{Acum_d} \leq \mu_{Acum} \quad (22)$$

$$VE_{Acum_d} > \mu_{Acum} \quad (23)$$

No primeiro caso, descrito pela equação (22), observamos uma venda acumulada menor ou igual à média de vendas, onde a ruptura pode ser a causa desta retração nas vendas. Para tal, a partir das equações (19) e (20) em que obtemos a

probabilidade da venda acumulada de um produto em um determinado dia d ser menor ou igual ao valor de $VE Acum_d$ obtido, a probabilidade de ruptura é dada por:

$$P3_d = 1 - f3_d \quad (24)$$

onde $P3_d$ é a probabilidade de ocorrência de ruptura de um determinado produto em um dia d , indicada pelo valor de sua venda acumulada nos últimos 7 dias e $f3_{d_n}$ é descrito pelas equações (19) ou (20), dependendo do tipo de distribuição identificada.

Já no caso descrito pela Equação (23), temos uma venda acumulada superior à média onde o aumento de vendas pode originar uma demanda não prevista, e a ruptura pode ser a consequência deste incremento de vendas. Neste cenário, a partir das equações (19) e (20), a probabilidade de ruptura pode ser considerada como sendo igual à probabilidade da venda acumulada de um produto em um determinado dia d ser menor ou igual ao valor de $VE Acum_d$ obtido, ou seja:

$$P3_d = f3_d \quad (25)$$

onde $P3_d$ é a probabilidade de ocorrência de ruptura de um determinado produto em um dia d , indicada pelo valor de sua venda acumulada nos últimos 7 dias e $f3_{d_n}$ é descrito pelas equações (19) ou (20), dependendo do tipo de distribuição identificada.

Como última variável de entrada, foi avaliado o saldo da movimentação deste item desde sua última entrada. Embora esta seja uma avaliação que mostra limitações, a informação pode trazer questionamentos quanto à real existência de rupturas, diminuindo assim seu grau de certeza. Os erros nos registros de inventário vão se acumulando no decorrer do tempo e das desta forma quanto menor o intervalo e o número de movimentações, menor o grau de incerteza operações (DEHORATIUS; MERSEREAU; SCHRAGE, 2008; TON; RAMAN, 2010).

O valor do saldo físico do produto desde a última entrada em um determinado dia d ($Saldo_d$) foi comparado com a média dos saldos, a partir da última entrada, no momento da reposição do produto (μ_{Saldo}). Novamente esta média foi recalculada até a entrada anterior à data verificada, fim de abranger possíveis mudança no comportamento do reabastecimento do produto.

Como o saldo não foi avaliado em um intervalo constante, mas apenas em datas anteriores à entrada do produto na filial, a variável foi considerada como pertencente a uma distribuição Normal. Além de características específicas, como possuir valor negativos (que inviabilizam a adequação à distribuição de Poisson, por exemplo) e a periodicidade de reposição que, em um cenário de vendas constantes, fará o saldo no momento do recebimento flutuar sobre um intervalo relativamente constante, testes estatísticos foram efetuados e mostraram uma adequação de 72% dos itens a este cenário. Novamente, foram utilizadas as ferramentas estatísticas da linguagem Python para verificar a adequação dos dados à distribuição normal, com um valor p de 0,05, por meio do K -test descrito na Equação (9).

Desta forma, considerando a distribuição do saldo no momento da reposição do produto adequado à distribuição Normal, a probabilidade acumulada de ocorrência de um saldo a partir da última compra de até x unidades é dada por:

$$f4(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x - \mu_{\text{Saldo}}}{\sigma_{\text{Saldo}} \sqrt{2}} \right) \right] \quad (26)$$

onde $f4(x)$ refere-se à probabilidade acumulada ocorrência de um saldo de até x unidades ($x = \text{Saldo}_d$), μ_{Saldo} é o saldo médio no momento da reposição do produto p , σ_{Acum} é o desvio padrão da amostra dos saldos analisados e erf é uma função de erro descrita pela equação (21).

Assim, a partir da equação (26) em que obtemos a probabilidade acumulada do saldo de um produto em um determinado dia d ($x = \text{Saldo}_d$) ocorrer, temos:

$$P4_d = 1 - f4_d \quad (27)$$

onde $P4_d$ é a probabilidade de ocorrência de ruptura de estoque de determinado produto em um dia d , indicada pelo saldo no momento da venda nula e $f4_{d_n}$ é descrito pela equação (26).

Ao analisarmos o comportamento das probabilidades acumuladas de ocorrência de determinado valor em relação à média de uma distribuição normal, observamos que valores menores do que a média indicam um saldo inferior ao habitualmente obtido no momento da reposição. Por outro lado, um valor maior do que

a média sugere um saldo mais elevado do que em situações anteriores. Considerando que a probabilidade de ruptura é o complemento dessa probabilidade, é razoável supor uma maior probabilidade de ruptura para saldos inferiores à média e uma probabilidade de ruptura menor para saldos superiores à média.

2.8.2 Identificação Inicial dos Dados

A fim de permitir maior precisão nos ajustes do algoritmo a ser criado, foi efetuada uma rotulagem dos dados, identificando os casos efetivos de ruptura total por falta de estoque. Além disso, a identificação inicial armazenada conforme a Tabela 8 permitirá também a avaliação efetiva do modelo criado.

Tabela 8 – Identificação da Ruptura

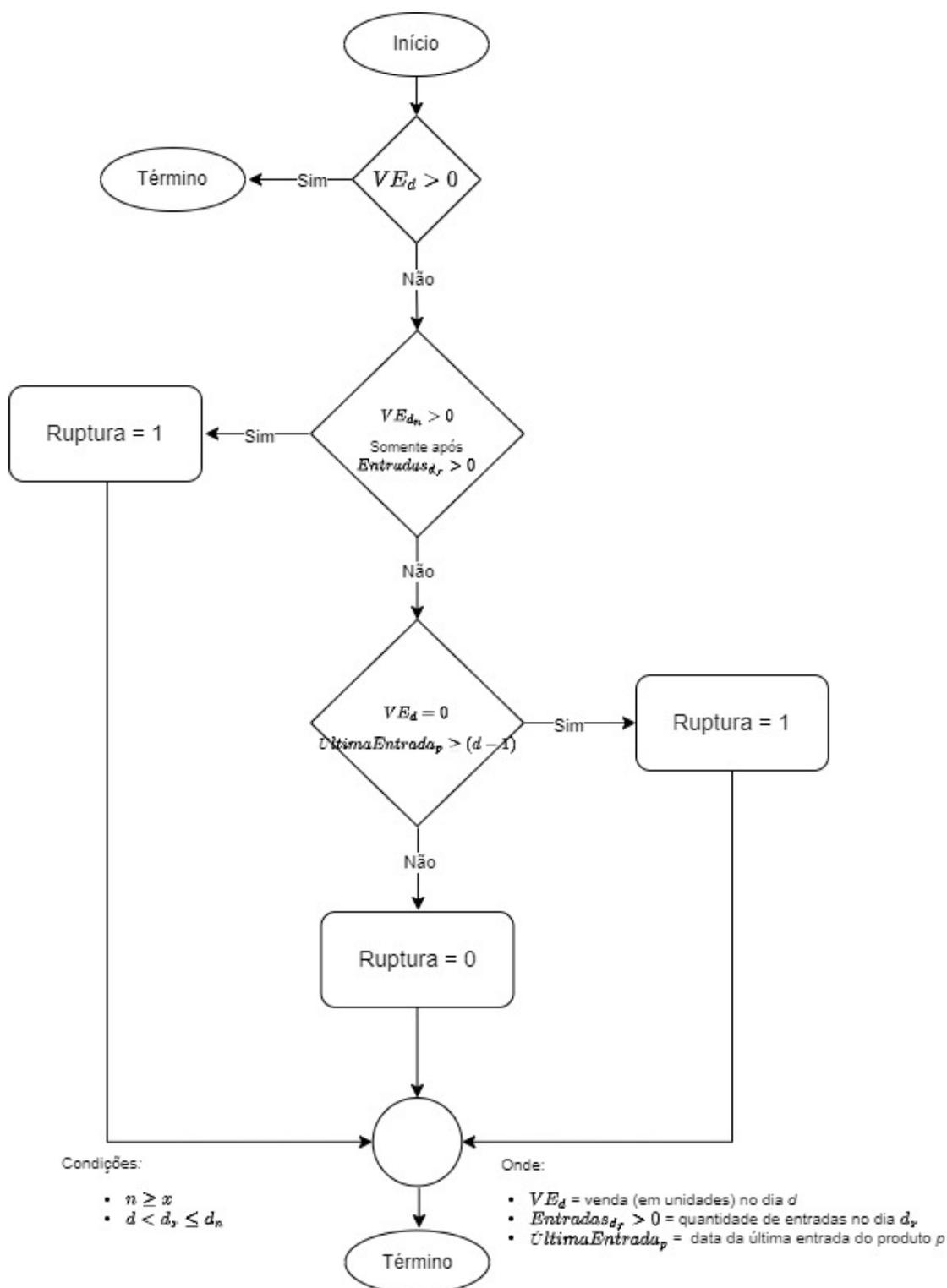
Campo	Descrição
Data	Data da movimentação
Código	Código do produto movimentado
Ruptura	Indicador de ruptura (0 = Não, 1 = Sim)

Fonte: Elaborado pelo autor.

Esta rotulagem foi realizada baseada exclusivamente na movimentação efetiva do produto, sem qualquer tipo de verificação ou ajuste manual. O processo seguiu a premissa de que um produto em ruptura por falta de estoque, somente voltará a ser vendido mediante uma nova reposição do estoque. Também foi considerado um período de um dia para o trânsito do produto entre o depósito da filial e o ponto de venda.

Desta forma, se a venda de um produto p em um determinado dia d (VE_d) for igual a zero e sua próxima venda ocorrer somente após a próxima entrada deste produto ($Entradas_{d_x}$), identificamos um cenário de ruptura por falta de estoque. Nos casos em que a venda é nula e a próxima venda não ocorre somente após a próxima entrada, mas a última entrada do produto ($Última Entrada_p$) ocorreu no dia ou no dia anterior (d ou $d - 1$), também é considerado um caso de ruptura total.

Figura 1 – Diagrama de Identificação Inicial de Rupturas



Fonte: Elaborado pelo autor.

A fim de alimentar a Tabela 8 de acordo com as informações da Figura 1 foram avaliados apenas os registros diários que apresentaram venda igual a zero para os

produtos selecionados. A análise dos 2.022 itens selecionados no período de 01/06/2022 a 31/05/2023 resultou em 299.342 registros com nula nos 361 dias de funcionamento da filial. Dentre esses registros, identificou-se que 124.953 correspondem realmente a casos de ruptura total de estoque causada pela falta efetiva de produtos em estoque. O código-fonte utilizado no desenvolvimento deste algoritmo está disponível no repositório Git sob o nome de “*Out-of-Stock-Detection*” (VERZELETTI, 2024).

Os 124.953 registros identificados como ruptura de estoque correspondem a apenas 42% do total de registros de venda diária igual a zero. Em outras palavras, em 58% dos casos em que não há registro de venda, a falta de estoque do item não é o motivo. Essa constatação ressalta a necessidade do desenvolvimento de um algoritmo capaz de distinguir essas situações com o máximo de antecedência possível. Nesse cenário, a identificação ocorreu apenas após a ocorrência de uma nova entrada do item, quando as perdas de venda já haviam ocorrido.

2.8.3 Algoritmo

Para execução da análise baseada nos critérios mencionados, foi desenvolvido um algoritmo fundamentado em lógica difusa (*fuzzy logic*) aplicado aos itens selecionados, que engloba os eventos a serem considerados. Nesta abordagem, o problema não é modelado de maneira precisa, mas sim aproximada, lidando com informações incertas.

O modelo foi desenvolvido utilizando biblioteca *scikit-fuzzy* da linguagem Python, voltada para aplicações dos conceitos de lógica difusa e foi construído seguindo a abordagem proposta por Bezdek (2014). A construção do modelo utilizou as probabilidades de ruptura pela ocorrência de venda nula em um dia ($P1_d$), pela sequência de vendas nulas ($P2_d$), pela venda acumulada nos últimos 7 dias ($P3_d$) e pelo saldo no produto a partir da última entrada ($P4_d$) como variáveis de entrada (antecedentes). A variável de saída (consequente) foi definida como R_d , representando a probabilidade de ruptura total de um produto p em um dia d devido à falta efetiva de estoque. Como, tanto as variáveis antecedentes quanto a variável consequente são probabilidades, o universo foi definido como os valores variando de 0 a 1 em intervalos dez milésimos, ou seja, 0,0001. Este intervalo, também conhecido

como “passo”, se refere à granularidade ou à resolução do espaço difuso, ou seja, é o tamanho do incremento usado para discretizar o universo contínuo em um conjunto de valores discretos.

Este método utiliza o conceito de termos e funções de pertinência para lidar com a incerteza do modelo, permitindo a representação eficaz de conceitos vagos. Os termos podem ser definidos como os “rótulos” atribuídos aos conjuntos difusos e são empregados para descrever os diferentes níveis das variáveis, tanto de entrada quanto de saída. Neste caso, os termos “baixa”, “média” e “alta” foram utilizados para descrever as probabilidades de entrada e de saída do modelo, proporcionando uma compreensão intuitiva diretamente relacionada à linguagem natural.

Cada conjunto difuso definido pelos rótulos estabelecidos é associado a uma função de pertinência, que determina o grau de pertencimento de um elemento a este conjunto. As funções de pertinência descrevem a incerteza associada à classificação de um elemento a este conjunto e, por este motivo, precisam ser definidas de forma consistente.

Neste estudo, os intervalos correspondentes às funções de pertinência associadas a cada termo das variáveis foram estabelecidos com base na amostra previamente classificada. Esta amostra, constituída por 20% do total de registros referentes à venda nula identificados e classificados anteriormente, foi selecionada utilizando técnicas comuns de treinamento de modelos empregadas em abordagens de aprendizado de máquina (AHN et al., 2019; AMUTHA; KHAN, 2023; ANDAUR; RUZ; GOYCOOLEA, 2021; CADAVID; LAMOURI; GRABOT, 2018).

A análise estatística dos 59.869 registros selecionados fundamentou a definição das funções de pertinência de cada um dos rótulos das variáveis de entrada:

1. Baixa: o primeiro intervalo de valores foi estabelecido a partir dos dados da amostra correspondente aos itens sem indicativo de ruptura. A análise estatística destes dados, não revelou uma diferença significativa entre a média e a mediana, razão pela qual os valores das médias foram utilizados:

Tabela 9 – Amostra sem Ruptura

	Registros	P1	P2	P3	P4
Sem Ruptura	34.728	0,680	0,855	0,796	0,276

Fonte: Elaborado pelo autor.

Desta forma, o intervalo teve início em zero (valor mínimo possível para as probabilidades) e seu limite superior foi estabelecido como a média das probabilidades encontradas para os itens sem ruptura na amostra selecionada:

Tabela 10 – Intervalos de Probabilidades “Baixas”

Variável	Início	Término
P1	0	0,68
P2	0	0,85
P3	0	0,80
P4	0	0,28

Fonte: Elaborado pelo autor.

2. Alta: o intervalo superior de valores foi estabelecido a partir dos dados da amostra correspondente aos itens com indicativo de ruptura. Novamente foram utilizados os valores das médias:

Tabela 11 – Amostra com Ruptura

	Registros	P1	P2	P3	P4
Com Ruptura	25.141	0,720	0,963	0,890	0,511

Fonte: Elaborado pelo autor.

Neste caso, o intervalo teve início no valor da média encontrado e seu limite superior igual a 1 (valor máximo possível para as probabilidades):

Tabela 12 – Intervalos de Probabilidades “Altas”

Variável	Início	Término
P1	0,72	1,0
P2	0,96	1,0
P3	0,89	1,0
P4	0,51	1,0

Fonte: Elaborado pelo autor.

3. Média: a definição do intervalo médio baseou-se, além da cobertura dos valores faltantes, no desvio padrão apresentado pela amostra. Como há variações tanto no intervalo inferior quanto no superior, foram utilizados os valores referentes à amostra total:

Tabela 13 – Desvio Padrão (Amostra Total)

	Registros	P1	P2	P3	P4
Geral	59.869	0,136	0,146	0,179	0,295

Fonte: Elaborado pelo autor.

Para este intervalo, o desvio padrão encontrado para cada uma das variáveis foi sobreposto aos demais intervalos. A sobreposição em cada um dos intervalos foi de 50% do desvio padrão identificado na amostra (σ).

$$\text{Limite inferior} = \text{Término do intervalo anterior} - \frac{1}{2}\sigma \quad (28)$$

$$\text{Limite superior} = \text{Início do intervalo posterior} - \frac{1}{2}\sigma \quad (29)$$

Desta forma:

Tabela 14 – Intervalos de Probabilidades “Médias”

Variável	Início	Término
P1	0,61	0,79
P2	0,78	1,00
P3	0,75	0,98
P4	0,19	0,60

Fonte: Elaborado pelo autor.

Já as funções de pertinência da variável de saída foram definidas de acordo com os objetivos propostos, buscando minimizar a identificação incorreta de rupturas, através da diminuição do intervalo superior. Desta forma, 50% do intervalo foi definido para as probabilidades “baixas” de ruptura e apenas 25% para as probabilidades “altas”. Como sobreposição, foi utilizado um percentual de 5% do intervalo médio em cada uma das extremidades:

Tabela 15 – Intervalos da Variável Consequente (R)

Rótulo	Início	Término
Baixa	0	0,50
Média	0,45	0,80
Alta	0,75	1

Fonte: Elaborado pelo autor.

Além dos intervalos, a função de pertinência pode adotar diferentes formatos para distribuir o grau de pertencimento dos valores ao intervalo. Neste estudo, optou-se pelo formato triangular tanto para as variáveis de entrada quanto para a variável de saída. Embora os formatos trapezoidal e gaussiano também sejam opções viáveis, o formato triangular foi selecionado devido à sua facilidade de visualização e interpretação, requerendo apenas um ponto, além dos limites já estabelecidos, para sua construção.

Com base nas definições empregadas para a determinação dos intervalos, a função triangular foi definida buscando equidade no intervalo. Desta forma, o triângulo foi definido através do ponto médio do intervalo. Nos casos em que a definição correta do intervalo ultrapassa o limite superior do universo estabelecido, utiliza-se o ponto médio calculado sem considerar este limite superior:

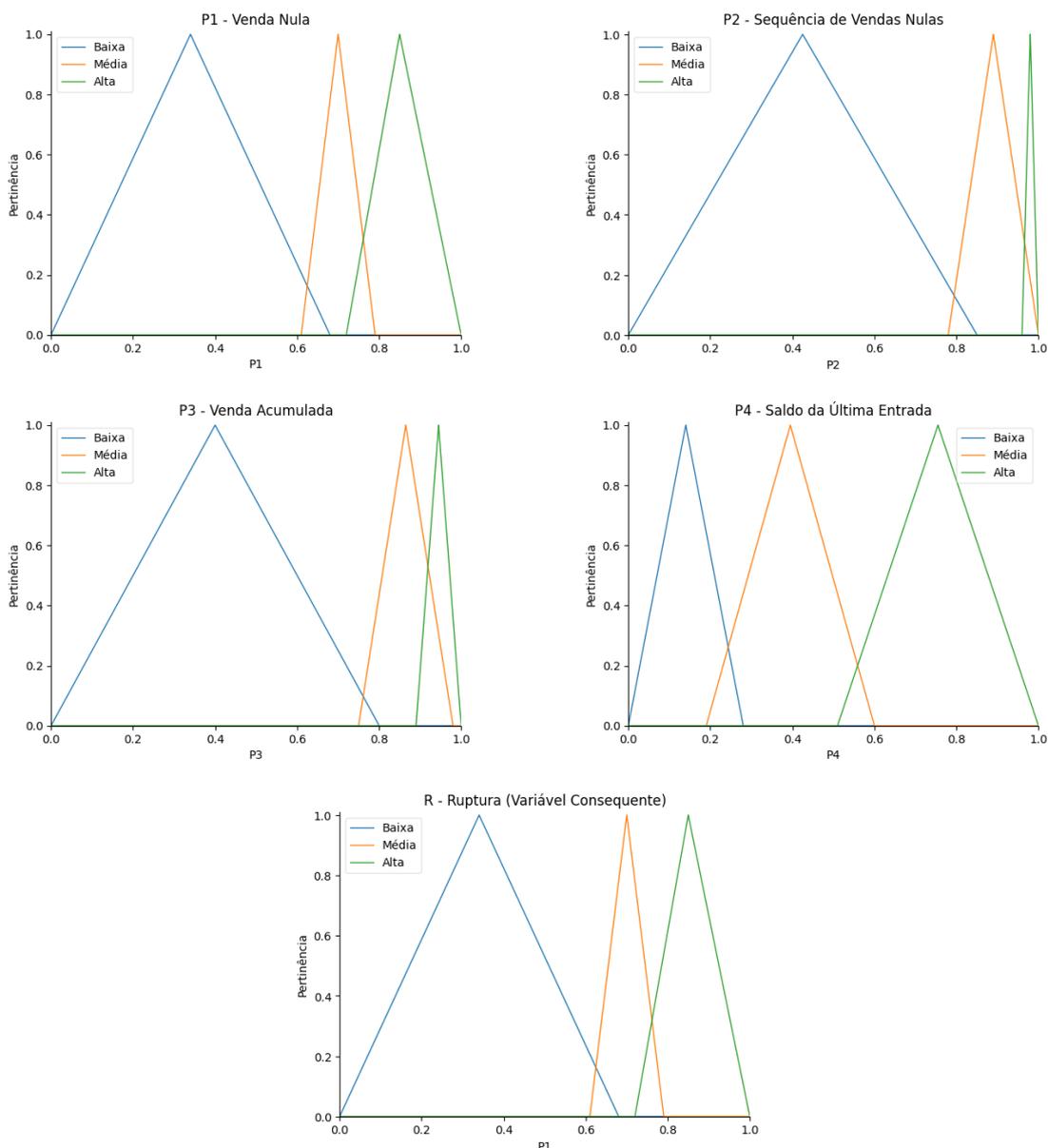
Tabela 16 – Pontos Médios

Rótulo	P1	P2	P3	P4	R
Baixa	0,340	0,425	0,400	0,140	0,250
Média	0,700	0,905	0,865	0,395	0,625
Alta	0,860	0,980	0,945	0,755	0,875

Fonte: Elaborado pelo autor.

A seguir, a visualização gráfica das funções de pertinência definida para cada uma das variáveis:

Figura 2 – Funções de Pertinência



Fonte: Elaborado pelo autor.

A partir da definição das variáveis e suas funções de pertinência, foi necessária a definição das regras que associam as variáveis de entrada à variável de saída. Essas regras descrevem como as variáveis antecedentes estão associadas à variável consequente e possuem impacto direto no resultado obtido. O número de antecedentes e a quantidade de termos atribuídos para cada um, define o número de regras necessário para cobertura total das possibilidades. Neste caso, em que temos 4 variáveis de entrada e cada uma possui 3 rótulos distintos, são necessárias 81 regras para um sistema completo de classificação.

A forma como as variáveis foram desenvolvidas, deixa clara algumas relações. Desta forma, quando todas as probabilidades de entrada forem “baixas”, a saída também será “baixa”. Quando as probabilidades de entrada forem “altas”, a saída também será “baixa”. Por analogia, entradas “médias” produzem saídas médias. Porém, as demais relações dependem do peso relativo a ser dado à variável de entrada. A forma como são definidas estas regras resulta em uma espécie de “peso” maior ou menor à variável considerada. Se duas regras possuem apenas mudança no termo de uma das variáveis, uma alteração na sua saída significa alta relevância para esta variável modificada. Por outro lado, não alterar o estado da saída, diminuiu sua importância no contexto.

Por este motivo, as regras foram definidas baseadas na análise dos valores obtidos na amostra e conhecimento do negócio. O desvio padrão elevado apresentado pela variável *P4*, representa um indicativo de baixo impacto no resultado final do algoritmo. Quanto às demais variáveis, embora apresentem desvios padrão semelhantes, baseiam-se em informações de diferentes significados. Enquanto a variável *P3* se baseia na venda acumulada de um período de 7 e, por este motivo, pode sofrer diversas influências, as variáveis *P1* e *P2* são construídas diretamente sobre a ocorrência de venda nula. Desta forma, as regras devem, em algum momento, considerar esta maior relevância no cálculo.

Tabela 17 – Regras de Inferência

P1	P2	P3	P4	R	P1	P2	P3	P4	R
Alta	Alta	Alta	Alta	Alta	Média	Média	Média	Baixa	Média
Alta	Alta	Alta	Média	Alta	Média	Média	Baixa	Alta	Média
Alta	Alta	Alta	Baixa	Alta	Média	Média	Baixa	Média	Média
Alta	Alta	Média	Alta	Alta	Média	Média	Baixa	Baixa	Média
Alta	Alta	Média	Média	Alta	Média	Baixa	Alta	Alta	Média
Alta	Alta	Média	Baixa	Alta	Média	Baixa	Alta	Média	Baixa
Alta	Alta	Baixa	Alta	Alta	Média	Baixa	Alta	Baixa	Baixa
Alta	Alta	Baixa	Média	Alta	Média	Baixa	Média	Alta	Baixa
Alta	Alta	Baixa	Baixa	Alta	Média	Baixa	Média	Média	Baixa
Alta	Média	Alta	Alta	Alta	Média	Baixa	Média	Baixa	Baixa
Alta	Média	Alta	Média	Alta	Média	Baixa	Baixa	Alta	Baixa
Alta	Média	Alta	Baixa	Alta	Média	Baixa	Baixa	Média	Baixa
Alta	Média	Média	Alta	Alta	Média	Baixa	Baixa	Baixa	Baixa
Alta	Média	Média	Média	Alta	Baixa	Alta	Alta	Alta	Alta
Alta	Média	Média	Baixa	Alta	Baixa	Alta	Alta	Média	Alta
Alta	Média	Baixa	Alta	Alta	Baixa	Alta	Alta	Baixa	Alta
Alta	Média	Baixa	Média	Alta	Baixa	Alta	Média	Alta	Média
Alta	Média	Baixa	Baixa	Alta	Baixa	Alta	Média	Média	Média
Alta	Baixa	Alta	Alta	Média	Baixa	Alta	Média	Baixa	Média
Alta	Baixa	Alta	Média	Média	Baixa	Alta	Baixa	Alta	Média
Alta	Baixa	Alta	Baixa	Média	Baixa	Alta	Baixa	Média	Média
Alta	Baixa	Média	Alta	Média	Baixa	Alta	Baixa	Baixa	Média
Alta	Baixa	Média	Média	Média	Baixa	Média	Alta	Alta	Alta
Alta	Baixa	Média	Baixa	Média	Baixa	Média	Alta	Média	Alta
Alta	Baixa	Baixa	Alta	Média	Baixa	Média	Alta	Baixa	Média
Alta	Baixa	Baixa	Média	Média	Baixa	Média	Média	Alta	Média
Alta	Baixa	Baixa	Baixa	Média	Baixa	Média	Média	Média	Média
Média	Alta	Alta	Alta	Alta	Baixa	Média	Média	Baixa	Baixa
Média	Alta	Alta	Média	Alta	Baixa	Média	Baixa	Alta	Baixa
Média	Alta	Alta	Baixa	Alta	Baixa	Média	Baixa	Média	Baixa
Média	Alta	Média	Alta	Alta	Baixa	Média	Baixa	Baixa	Baixa
Média	Alta	Média	Média	Alta	Baixa	Baixa	Alta	Alta	Baixa
Média	Alta	Média	Baixa	Alta	Baixa	Baixa	Alta	Média	Baixa
Média	Alta	Baixa	Alta	Alta	Baixa	Baixa	Alta	Baixa	Baixa
Média	Alta	Baixa	Média	Alta	Baixa	Baixa	Média	Alta	Baixa
Média	Alta	Baixa	Baixa	Média	Baixa	Baixa	Média	Média	Baixa
Média	Média	Alta	Alta	Alta	Baixa	Baixa	Média	Baixa	Baixa
Média	Média	Alta	Média	Média	Baixa	Baixa	Baixa	Alta	Baixa
Média	Média	Alta	Baixa	Média	Baixa	Baixa	Baixa	Média	Baixa
Média	Média	Média	Alta	Média	Baixa	Baixa	Baixa	Baixa	Baixa
Média	Média	Média	Média	Média	Baixa	Baixa	Baixa	Baixa	Baixa

Fonte: Elaborado pelo autor.

Após as definições dos universos das variáveis, das funções de pertinência e das regras associadas, é feita a inferência difusa. Esta inferência identifica a região do consequente correspondente às respectivas variáveis de entrada. Embora as regras sejam definidas através dos rótulos atribuídos, os valores de entrada resultam

em um valor de saída também numérico dentro das regras de inferência atribuídas para o consequente.

A saída gerada, foi definida a fim de representar a probabilidade de ruptura total por falta de estoque a partir da combinação das variáveis de entrada informadas. Como o objetivo do estudo trata da identificação dos casos em que existe a ruptura de estoque, precisamos encontrar um ponto de corte ideal (m) a fim de definir a sinalização ou não deste estado.

Utilizando-se novamente a amostra de 20% dos registros que definiram os intervalos de cada rótulo, foram efetuados testes quando ao valor de corte ideal. Pela função de pertinência definida para a variável consequente, espera-se que este ponto de corte esteja entre os valores os médios dos intervalos de média e alta probabilidade, ou seja entre 0,625 e 0,875. Desta forma, ao analisarmos os 59.869 registros, foram obtidos os seguintes resultados para os diferentes valores testados:

Tabela 18 – Definição do Ponto de Corte (m)

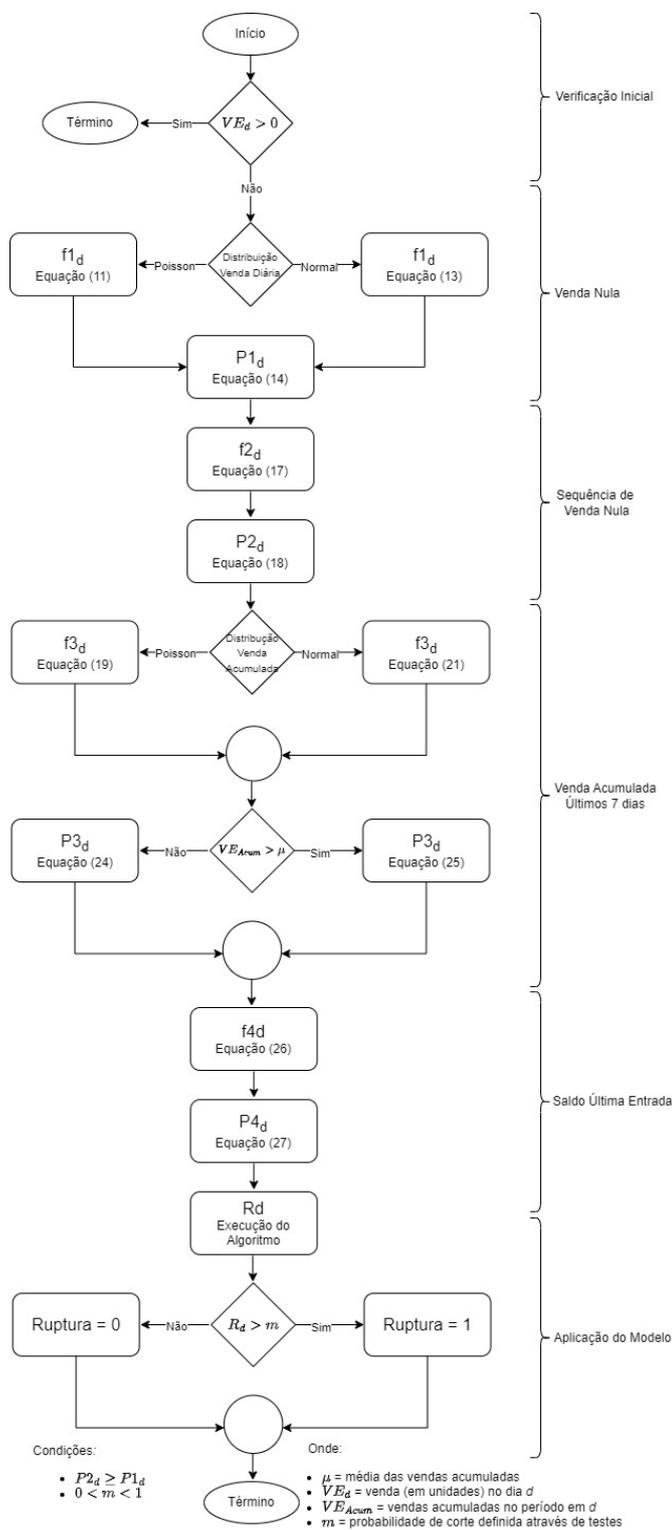
	Teste 1	Teste 2	Teste 3	Teste 4	Teste 5	Teste 6
M	0,625	0,750	0,825	0,850	0,860	0,875
Verdadeiros positivos	22.768	18.199	17.715	17.582	1.385	287
Falsos positivos	15.261	7.207	5.887	5.701	1.421	243
Verdadeiros negativos	19.467	27.521	28.841	29.027	33.307	34.485
Falsos negativos	2.373	6.942	7.426	7.559	23.756	24.854
Acurácia	0,71	0,76	0,78	0,78	0,58	0,58
Precisão	0,60	0,72	0,75	0,76	0,49	0,54
<i>Recall</i>	0,91	0,72	0,70	0,70	0,06	0,01

Fonte: Elaborado pelo autor.

Esta definição precisa estar de acordo com os resultados esperados, com o objetivo de alcançar maior precisão na identificação correta dos casos de ruptura. Neste caso, a prioridade é minimizar o número de falsos positivos identificados, já que a geração de alertas falsos de casos de ruptura pode ser operacionalmente custosa. Ao definir o ponto de corte como 0,85 ($m = 0,85$), observamos que há 5.701 casos de falsos positivos na amostra, representando apenas 9,5% do total de casos avaliados. Além disso, o número de 7.559 ocorrências de falsos negativos corresponde a apenas 12,6% do total de casos analisados. Os graus de precisão e acurácia, respectivamente 0,76 e 0,78, também se mostraram satisfatórios por estarem próximos, o que viabiliza a utilização efetiva desse ponto de corte na operacionalização do modelo.

As regras e condições foram encadeadas de acordo com o diagrama de fluxo da Figura 3:

Figura 3 – Diagrama de análise de um produto p no dia d



Fonte: Elaborado pelo autor.

A partir da execução do algoritmo e de uma probabilidade de corte definida, a existência de ruptura foi identificada diariamente para cada um dos itens no período analisado, conforme Tabela 19. O código-fonte utilizado no desenvolvimento deste modelo está disponível no repositório Git sob o nome de “*Out-of-Stock-Detection*” (VERZELETTI, 2024).

Tabela 19 – Rupturas Identificadas

Campo	Descrição
Data	Data da movimentação
Código	Código do produto movimentado
Ruptura	Indicador de ruptura (0 = Não, 1 = Sim)

Fonte: Elaborado pelo autor.

Desta forma, as informações contidas na Tabela 19, derivadas da execução do algoritmo desenvolvido, representam a predição do modelo, uma vez que este se baseia exclusivamente em informações anteriores à data em questão. Esses dados podem ser contrastados com as identificações presentes na Tabela 8, que refletem a situação real de ruptura do produto, pois são fundamentadas em dados futuros.

2.8.4 Avaliação do Modelo

O algoritmo desenvolvido tem como objetivo identificar casos de ruptura total de estoque. Em outras palavras, o modelo procura classificar os eventos em que as vendas são iguais a zero em um determinado dia como casos de ocorrência ou não de ruptura. Para avaliação de problemas que envolvem algum tipo de classificação, uma ferramenta amplamente utilizada é a Matriz de Confusão.

Uma Matriz de Confusão Binária de um classificador estabelece a relação entre as classes positivas (P) e negativas (N) atribuídas durante o processo de classificação. Neste estudo, a ocorrência de ruptura foi definida uma classe positiva enquanto a ausência de ruptura foi tratada como classe negativa. Assim, ao compararmos a saída do modelo (Tabela 19) com os dados rotulados obtidos na Tabela 8, podemos avaliar sua eficiência através da classificação destes resultados.

Neste caso, onde a presença de ruptura foi tratada como uma classe positiva, os eventos corretamente classificados pelo modelo como ruptura são denominados verdadeiros positivos (VP). Por outro lado, os eventos classificados incorretamente

como ruptura, ou seja, quando não houve ruptura real, mas a saída do algoritmo foi positiva, são chamados de falsos positivos (FP). Considerando ainda que a ausência de ruptura foi tratada como uma classe negativa, os eventos corretamente classificados pelo modelo como não ruptura são conhecidos como verdadeiros negativos (VN). Já, os eventos de não ruptura identificados erroneamente são denominados falsos negativos (FN).

A partir destas definições, a Matriz de Confusão para avaliação do modelo pode ser representada da seguinte forma:

Tabela 20 – Matriz de Confusão Binária

		Modelo	
		SIM	NÃO
Real	SIM	VP	FN
	NÃO	FP	VN

Fonte: Elaborado pelo autor.

A Matriz de Confusão evidencia o desempenho do algoritmo em números absolutos. A classificação correta de eventos é descrita pelos valores de verdadeiros positivos (VP) e de verdadeiros negativos (VN), que representam, respectivamente, os casos de ruptura e de não ruptura identificados de forma correta. Já as rupturas não identificadas pelo modelo (FN) e as classificações de eventos de ruptura de forma incorreta (FP) apontam as deficiências do algoritmo em cada um dos aspectos.

É importante ressaltar que as ocorrências não classificadas corretamente possuem impactos com pesos distintos na avaliação. Enquanto as rupturas não identificadas (FN) fazem parte da situação já existente, os casos de não ruptura erroneamente classificados como ruptura pelo modelo (FP) podem ter impactos negativos na operação. Nestes casos, podem surgir esforços adicionais por parte da empresa, seja em termos operacionais para verificar a ocorrência real ou mesmo financeiros, como a reposição de produtos desnecessários. Além disso, um número elevado de alarmes falsos pode comprometer a confiabilidade dos resultados perante os responsáveis pela cadeia de abastecimento. Desta forma, embora o ideal seja a minimização de qualquer tipo de erro de classificação, os valores de falsos positivos merecem uma atenção especial na avaliação do modelo.

Além da visão do número de eventos classificados corretamente ou incorretamente para cada um dos tipos de classes, a Matriz de Confusão também fornece métricas importantes para análise do desempenho do algoritmo. Estas métricas, são fundamentais para avaliação dos resultados obtidos pelo modelo em relação aos métodos utilizados pelos estudos anteriores.

Através da proporção das identificações corretas do modelo, podemos mensurar seu desempenho geral. Para isso, utilizamos a acurácia como métrica indicadora:

$$Acurácia = \frac{VP+VN}{VP+VN+FP+FN} \quad (30)$$

Já a precisão do modelo mede a proporção de eventos positivos classificados corretamente em relação ao total de eventos positivos:

$$Precisão = \frac{VP}{VP+FP} \quad (31)$$

Ainda avaliando a identificação de classes positivas, o *recall* avalia a capacidade do modelo de identificar todos os eventos positivos, medindo a proporção de eventos positivos classificados corretamente em relação ao total de ocorrências positivas verdadeiras:

$$Recall = \frac{VP}{VP+FN} \quad (32)$$

Vale ressaltar que um equilíbrio entre precisão e recall é essencial para garantir um desempenho satisfatório do modelo. Um alto valor de precisão indica que o modelo faz poucas classificações errôneas de casos de ruptura, enquanto um alto *recall* indica que o modelo identifica a maioria das ocorrências de ruptura. Neste cenário, aumentar a precisão muitas vezes leva a uma diminuição no *recall* e vice-versa. Por este motivo, um equilíbrio ideal entre precisão e *recall* depende do contexto específico do problema. Neste estudo, onde o impacto negativo gerado por classificações de eventos como ocorrências de ruptura de forma incorreta (FP) é muito superior aos casos eventualmente não detectados, opta-se por priorizar o valor da precisão. Esta

situação pode ser observada na definição do ponto de corte apresentada na Tabela 18.

Em relação à classe negativa, a especificidade do modelo representa a proporção dos casos de não ocorrência de ruptura identificados corretamente:

$$\text{Especificidade} = \frac{VN}{VN+FN} \quad (33)$$

Uma alta especificidade indica que o modelo é capaz de identificar corretamente a maioria dos casos negativos, minimizando assim os falsos positivos. Por outro lado, uma baixa especificidade indica que o modelo está classificando erroneamente muitos casos negativos como positivos, resultando em uma alta taxa de falsos positivos. Novamente, o contexto do problema e as consequências associadas são fundamentais para avaliação desta métrica.

Além destas, as taxas de identificações incorretas também são importantes para avaliação do desempenho do modelo. Em especial, níveis elevados de falsos positivos podem gerar alertas desnecessários, diminuindo a credibilidade do sistema.

$$\text{Taxa de FP} = \frac{FP}{FP+VN} \quad (34)$$

$$\text{Taxa de FN} = \frac{FN}{FN+VP} \quad (35)$$

Mais uma vez, embora o objetivo do algoritmo seja obter as menores taxas de erro possíveis em ambos os cenários, é necessário mitigar impactos operacionais e financeiros decorrentes. Por este motivo, o ponto de corte estabelecido priorizou a minimização das taxas de falsos positivos, visando reduzir o número de alarmes falsos potencialmente gerados.

Desta forma, o modelo busca obter métricas de desempenho superiores aos estudos prévios, melhorando níveis de acurácia e minimizando alarmes falsos através da diminuição da taxa de falsos positivos. Somente através destas métricas e de sua comparação com métodos já existentes, pode ser verificada a real contribuição do algoritmo ao desenvolvimento das pesquisas nesta área.

2.8.5 Limitações do Modelo

De acordo com proposta do estudo, o objetivo principal é identificar os casos de ruptura total consequentes da falta de estoque. Embora o modelo não aborde especificamente as indisponibilidades dos produtos devido a erros operacionais, esses casos também serão analisados pelo algoritmo.

Ao analisarmos um caso extremo, onde o produto possuía estoque na filial, porém não disponível no ponto de venda, e o mesmo teve ocorrência de venda somente após uma nova entrada no estoque, será considerado como ruptura pela forma de identificação utilizada. No entanto, as probabilidades de ruptura, especialmente do saldo a partir da última entrada, estarão distorcidas, podendo comprometer a própria avaliação do algoritmo.

Desta forma, apesar da possibilidade de haver pequenas distorções na predição do modelo e até na própria identificação inicial dos casos de ruptura, os objetivos finais de sua utilização ainda se mostram válidos.

3 ANÁLISE DOS RESULTADOS

A avaliação do modelo abrange não apenas o desempenho do algoritmo, mas também o impacto prático de sua aplicação. Isso envolve análises descritivas das variáveis de entrada e saída, bem como avaliações dos potenciais resultados econômicos decorrentes da identificação antecipada de cenários de ruptura de estoque.

Os resultados foram analisados comparando os dados gerados pelo modelo (Tabela 19) com a identificação dos casos de ruptura definidos na Tabela 8. Para esta análise, foram utilizadas ferramentas estatísticas da própria linguagem Python, além de consultas aos dados já previamente coletados.

3.1 Análise do Modelo

O algoritmo desenvolvido foi aplicado sobre 299.342 registros. Estes registros correspondem às ocorrências de venda igual a zero dos 2.022 itens analisados no período de 01/06/2022 a 31/05/2023. Destes, 116.388 foram identificados pelo modelo como casos de ruptura total de estoque decorrentes da falta efetiva do produto, ou seja, em 39% dos casos. Ao compararmos a saída do modelo com o estado real de ruptura dos produtos, temos:

Tabela 21 – Real x Modelo

	Real	Modelo
Com Ruptura	124.953	116.388
Sem Ruptura	174.389	182.954

Fonte: Elaborado pelo autor.

Podemos observar que a taxa de ruptura real, entre as ocorrências de venda nula, é de 42%. Embora a identificação do modelo esteja abaixo da taxa real, é importante destacar que a diferença não é significativa a ponto de desqualificar o resultado do modelo. Além disso, ao compararmos estas taxas com a amostra utilizada para definição do ponto de corte (Tabela 18), observamos que os percentuais de ruptura real e do modelo permanecem consistentes, indicando que a amostra utilizada

representa de forma precisa a totalidade dos registros. Essa consistência reforça a confiabilidade nas premissas utilizadas no desenvolvimento e ajustes do algoritmo.

Para analisarmos a efetividade do modelo, podemos utilizar a Matriz de Confusão:

Tabela 22 – Matriz de Confusão do Modelo

		Previsto		
		SIM	NÃO	TOTAL
Real	SIM	87.764	37.189	124.953
	NÃO	28.624	145.765	174.389
	TOTAL	116.388	182.954	299.342

Fonte: Elaborado pelo autor.

A partir da análise da Matriz de Confusão, podemos observar que o modelo apresentou um bom desempenho na identificação dos casos de ruptura, com uma alta proporção de verdadeiros positivos (VP), totalizando 87.764 casos. Isso significa que o modelo conseguiu identificar corretamente a maioria das ocorrências reais de ruptura de estoque. No entanto, também houve um número significativo de falsos negativos (FN), representando 37.189 casos, o que indica que o modelo deixou de identificar algumas rupturas reais. Por outro lado, o modelo teve um alto número de verdadeiros negativos (VN), totalizando 145.765 registros, ou seja, identificou corretamente a ausência de ruptura quando ela não ocorreu. Entretanto, também houve uma quantidade considerável de falsos positivos (FP), representando 28.624 registros, nos quais o modelo identificou erroneamente a presença de ruptura quando na verdade não houve.

Ainda analisando a matriz de confusão, podemos extrair métricas cruciais para avaliar o desempenho do modelo de identificação de casos de ruptura de estoque:

Tabela 23 – Rupturas Identificadas

Métrica	Valor
Acurácia	0,780
Precisão	0,754
Especificidade	0,836
Recall	0,702
Taxa de FP	0,164
Taxa de FN	0,298

Fonte: Elaborado pelo autor.

As métricas apresentadas na Tabela 23 proporcionam uma visão abrangente do desempenho do modelo em prever corretamente as instâncias positivas e negativas. A acurácia, que representa a proporção de previsões corretas em relação ao total de previsões feitas pelo modelo, foi calculada como 0,780. Isso significa que o modelo acertou aproximadamente 78% de todas as previsões feitas. Já a precisão, que mede a proporção de previsões positivas corretas em relação ao total de previsões positivas feitas pelo modelo, foi calculada como 0,754. Isso indica que cerca de 75,4% das previsões positivas feitas pelo modelo foram corretas.

A especificidade do modelo (taxa de verdadeiros negativos), indica que aproximadamente 83,6% dos casos verdadeiramente negativos foram corretamente identificados como negativos pelo modelo. Por outro lado, o *recall* do modelo (taxa de verdadeiros positivos) significa que aproximadamente 70,2% de todos os casos verdadeiramente positivos foram corretamente identificados como positivos pelo modelo. Estas métricas medem, respectivamente, a capacidade do algoritmo de capturar corretamente casos negativos e positivos.

Ao avaliarmos as taxas de falsos positivos e de falsos negativos, observamos que enquanto apenas em 16,4% dos casos verdadeiramente negativos foram classificados incorretamente pelo modelo, 29,8% dos casos positivos foram identificados erroneamente pelo algoritmo.

As métricas de desempenho do modelo de identificação de ruptura de estoque oferecem uma avaliação detalhada de sua eficácia, destacando suas principais características e limitações. Embora a taxa de verdadeiros positivos possa ser considerada relativamente baixa (aproximadamente 70%), o modelo mostrou-se mais eficiente do que a abordagem utilizada por Montoya e Gonzalez (2019), que obteve uma taxa de detecção de 63,48% a partir de uma avaliação de apenas 10 SKUs.

No entanto, o desempenho do algoritmo foi inferior à utilização de regras heurísticas utilizadas por Papakiriakopoulos, Pramataris e Doukidis (2009), que na avaliação de 109 produtos, apresentou uma taxa de detecção próxima a 85%. É importante destacar que, no referido estudo, a verificação e confirmação da eficácia foram realizadas manualmente em diferentes unidades monitoras. Dado que a confiabilidade dos dados depende de uma verificação oportuna das ocorrências, é plausível esperar algum grau de distorção neste indicador.

Também podemos destacar que a minimização de falsos positivos foi uma prioridade durante o estudo. O processo de definição do ponto de corte do algoritmo foi planejado para reduzir estes casos, resultando em apenas 16,4% de alarmes falsos. O resultado ficou muito próximo ao obtido por Montoya e Gonzalez (2019), que obteve uma taxa de apenas 15,52% de falsos alertas em sua avaliação.

Desta forma, além dos elevados índices de precisão e de acurácia do modelo, os resultados apresentados são muito próximos, em alguns casos superiores, aos obtidos em estudos anteriores. Estes fatos fortalecem a viabilidade para aplicação prática do algoritmo desenvolvido.

3.2 Análise das Variáveis

Embora a validade do modelo para os objetivos do estudo tenha sido estabelecida, é crucial verificar o grau de significância das variáveis de entrada. Para isso, foram conduzidos testes de regressão linear ordinária (OLS) usando bibliotecas estatísticas disponíveis na linguagem de programação Python.

Esses testes analisaram o impacto das variáveis preditoras associadas às probabilidades de ruptura indicadas pela presença de vendas nulas ($P1$), pela sequência de vendas nulas ($P2$), pela variação acumulada de vendas nos últimos 7 dias ($P3$) e pelo saldo desde a última entrada ($P4$) no valor da variável de saída do modelo de lógica difusa, que representa a probabilidade de ruptura do item (R).

Tabela 24 – Teste de Regressão Linear Ordinária (OLS)

Teste Estatístico	Valor
R-quadrado	0,850
Estatística F	4,206e+05
p -valor (Estatística F)	0,000
Omnibus	23.672,247
p -valor (Omnibus)	0,000
Skew	0,242
Kurtosis	5,981
Durbin-Watson	1,963

Fonte: Elaborado pelo autor.

Tabela 25 – Análise das Variáveis (OLS)

Variável	Coefficiente	Erro Padrão	t	P > t	[0,025	0,975]
Constante	-0,7946	0,001	-635,300	0,000	-0,797	-0,792
P1	0,0966	0,001	74,937	0,000	0,094	0,099
P2	1,2658	0,001	1.075,720	0,000	1,264	1,268
P3	0,2620	0,001	288,210	0,000	0,260	0,264
P4	0,0591	0,001	104,092	0,000	0,058	0,060

Fonte: Elaborado pelo autor.

Ao analisarmos os resultados da OLS através da Tabela 24, verificamos que 85% da alteração da variável dependente (R) é explicada pelas variáveis independentes incluídas no modelo, conforme indicado pelo coeficiente de determinação (R -quadrado). Isso sugere que o modelo tem uma capacidade substancial de explicar a variação no valor de R . Além disso, observamos um valor elevado da estatística F , juntamente com um valor p próximo de zero, o que indica que o modelo como um todo é altamente significativo. O mesmo vale para o teste de *Omnibus*. O valor de p muito próximo de zero obtido para este teste mostra que o modelo é estatisticamente significativo

Ainda de acordo com os resultados, a avaliação dos coeficientes das variáveis de entrada fornece informações sobre como cada variável contribui para a previsão do valor de saída. A partir da Tabela 25, ao observamos que $P1 = 0,09$, $P2 = 1,26$, $P3 = 0,26$ e $P4 = 0,06$, percebemos que a variável de maior impacto no resultado do modelo é $P2$, seguida de $P3$ e $P1$. Por outro lado, a variável $P4$ possui a menor significância no modelo. Esses resultados estão em conformidade com os objetivos definidos na elaboração das regras do modelo, onde buscamos minimizar o impacto da probabilidade do saldo a partir da última entrada ($P4$), devido à sua alta variabilidade. Mesmo assim, todos os coeficientes têm valores p muito baixos (próximos de zero), indicando que todas as variáveis independentes são estatisticamente significativas para a previsão de R .

Desta forma, os resultados da regressão sugerem que o modelo tem um poder substancial para explicar a alteração na variável dependente (R), e todas as variáveis independentes são estatisticamente significativas na previsão da probabilidade. Esse resultado era esperado devido à estrutura das regras presentes no modelo de lógica difusa. No entanto, qualquer inconsistência nas regras teria um impacto direto nesses resultados, podendo reduzir significativamente a relevância das variáveis e a

confiabilidade do algoritmo. Além disso, o grau de importância dos valores de entrada se manteve de acordo com o que foi proposto durante o desenvolvimento do estudo.

3.3 Impacto Real

Enquanto sob a ótica de avaliação do modelo nosso foco está sobre o número total de registros identificados ou não identificados, no ambiente empresarial a identificação rápida e correta dos diferentes itens em estado de ruptura se mostra muito mais relevante. Como os diferentes dias da semana apresentam características próprias de fluxo de clientes e vendas, a análise da taxa de ruptura por dia da semana se mostra pertinente:

Tabela 26 – Ruptura por Dia da Semana

Dia da Semana	Itens vendidos	Itens em Ruptura	Taxa de Ruptura	Itens Identificados	Taxa Identificada
Domingo	999	335	16,6%	241	11,9%
Segunda	946	339	16,7%	235	11,6%
Terça	946	360	17,8%	241	11,9%
Quarta	1.049	369	18,2%	260	12,9%
Quinta	965	358	17,7%	241	11,9%
Sexta	1.028	336	16,6%	237	11,7%
Sábado	1.130	326	16,1%	248	12,2%
Geral	1.009	342	16,9%	243	12,0%

Fonte: Elaborado pelo autor.

Os dados da Tabela 26 consideram apenas os 2.022 itens analisados no período de 01/06/2022 a 31/05/2023. Podemos observar que a taxa de ruptura de estoque calculada de acordo com a Equação (1), é de aproximadamente 17%, porém, a taxa identificada pelo modelo é de apenas 12%. Apesar do modelo identificar, em média, 243 itens em ruptura, 99 casos passam despercebidos pelo algoritmo.

Embora uma taxa de ruptura de 17% esteja muito acima dos percentuais médios que variam entre 5 e 10% (PAPAKIRIAKOPOULOS; PRAMATARI; DOUKIDIS, 2009; USMAN; CAPLICE, 2008) o cenário avaliado não é o mesmo. Nestes estudos, não foi avaliado especificamente o varejo alimentar, mas sim o varejo em geral. Mesmo assim, estudos mostram que existem casos em que os índices de ruptura podem chegar a 15% (BERGER, 2003), muito próximo dos valores obtidos.

Outra observação importante diz respeito quanto aos índices de ruptura no decorrer da semana. Embora a menor taxa de ruptura ocorra no sábado, que é o dia com maior número de itens vendidos, essa relação não pode ser interpretada como causa e efeito, pois a quarta-feira, o segundo dia com maior volume médio de itens, apresenta a taxa de ruptura mais elevada. Quanto à eficácia da identificação, observa-se que ela varia ao longo da semana, com melhor desempenho no sábado, quando o modelo deixa de identificar, em média, apenas 78 itens. No entanto, nas terças e quintas-feiras, o modelo mostra-se menos eficiente, deixando de identificar 119 e 117 itens, respectivamente.

O principal benefício da identificação rápida e eficiente das ocorrências de ruptura de estoque é a possibilidade de diminuição do período de indisponibilidade de produto. O período de duração da ruptura de estoque pode ser de apenas um dia ou durar meses, sendo que o mesmo produto pode ter diferentes casos de ruptura com as mais diversas variações. Considerando cada período de ruptura de um determinado produto p , independente de sua duração, como uma ocorrência de ruptura, temos um total de 18.877 casos de ruptura, distribuídos da seguinte forma:

Tabela 27 – Duração da Ruptura de Estoque

Ruptura	Ocorrências	%
Até 7 dias	11.455	61%
7 a 14 dias	2.512	13%
14 a 28 dias	3.275	17%
28 a 56 dias	1.061	6%
Mais de 56 dias	574	3%

Fonte: Elaborado pelo autor.

A Tabela 27 mostra que 91% dos casos de ruptura são de períodos inferiores a 28 dias. Embora este seja um período relativamente longo, observamos que 61% dos casos se referem a rupturas de até 7 dias. Este fato torna ainda mais importante a necessidade de identificação rápida por parte do modelo, pois, além da identificação, ainda é necessário um período para a efetiva correção.

Ao compararmos os intervalos de ruptura com os prazos médios da identificação pelo modelo, temos:

Tabela 28 – Dias para Identificação da Ruptura

Ruptura	Ocorrências	Média Ruptura	Dias para Identificação	Não identificados	% Não Identificado
Até 7 dias	11.455	3	2	5.189	45%
8 a 14 dias	2.512	11	4	554	22%
15 a 28 dias	3.275	20	5	543	17%
29 a 56 dias	1.061	39	10	373	35%
Mais de 56 dias	574	85	18	298	52%
Geral	18.877	11	4	6.957	37%

Fonte: Elaborado pelo autor.

De acordo com a Tabela 28, em média, o modelo identificou casos de ruptura de estoque apenas 4 dias após sua ocorrência. Este resultado, embora ligeiramente inferior ao obtido por Papakiriakopoulos, Pramadari e Doukidis (2009) que alcançou a identificação da ruptura nos 109 produtos analisados em apenas 3 dias após sua ocorrência, ainda demonstra uma resposta rápida do algoritmo em detectar situações de falta de estoque.

Podemos observar ainda que o algoritmo se mostra menos preciso em intervalos de ruptura muito curtos e muito longos. A imprecisão em intervalos curtos pode ser atribuída ao menor impacto da variável que considera a sequência de vendas igual a zero, a qual foi identificada como a mais significativa no modelo. Já em intervalos longos, onde 52% das rupturas com duração superior a 56 dias não são previstas pelo modelo, uma explicação possível é a retroalimentação do sistema. Nesses casos, o longo período de ruptura afeta as próprias médias consideradas e distorce a distribuição teórica utilizada.

No entanto, ao analisarmos os casos identificados, que correspondem a 63% das ocorrências, observamos resultados promissores. Embora não haja uma redução significativa para rupturas de até 7 dias, nos demais cenários, o modelo pode reduzir o tempo médio sem vendas em até 79%, como no caso de rupturas com mais de 56 dias. Para as rupturas de 8 a 28 dias, que representam 39% do total de rupturas, o tempo médio de identificação é de aproximadamente 5 dias, o que implica em uma possível redução deste período em até 11 dias. Desta forma, considerando-se períodos de ruptura superiores a 7 dias, o modelo tem o potencial de diminuir o número de dias sem vendas em até 74%, reduzindo significativamente as perdas decorrentes desse tipo de ocorrência.

No caso específico deste estudo, em que identificamos os casos de ruptura por falta efetiva de estoque, além da própria identificação, a correção deste tipo de evento depende também de uma nova entrada do produto na filial. Por este motivo, o prazo médio de reposição também é significativo ao avaliarmos o modelo:

Tabela 29 – Prazo Médio de Reposição

Reposição	Número de Itens	%
Até 7 dias	21	1%
8 a 14 dias	234	12%
15 a 21 dias	181	9%
22 a 56 dias	1.035	51%
57 a 84 dias	341	17%
Mais de 84 dias	210	10%

Fonte: Elaborado pelo autor.

Conforme os critérios adotados para a seleção dos itens a serem analisados, produtos altamente perecíveis e, conseqüentemente, sujeitos a reposição frequente não foram incluídos no estudo. Assim, apenas 13% dos itens avaliados apresentam um prazo médio de reposição de até 14 dias. Por outro lado, os itens com prazo de reposição superior a 21 dias compreendem 78% do total analisado, sendo que mais da metade desse total está dentro do intervalo entre 22 e 56 dias.

Ao relacionarmos os prazos de reposição dos produtos, com as ocorrências de ruptura e sua identificação, temos:

Tabela 30 – Prazo Médio de Reposição

Reposição	Prazo Médio	Ocorrências de Ruptura	Dias Ruptura	Dias para Identificação	Não Identificados	% Não Identificado
Até 7 dias	4	219	1	1	4	2%
8 a 14 dias	11	2.895	3	2	903	31%
15 a 21 dias	18	1.829	6	2	612	33%
22 a 56 dias	39	8.453	10	3	2.974	35%
57 a 84 dias	68	2.799	17	6	1.440	51%
Mais de 84 dias	112	1.621	24	9	1.024	63%

Fonte: Elaborado pelo autor.

De acordo com o esperado, quanto maior o intervalo para a reposição, maior também será o número de dias em ruptura. Porém, se considerarmos o número de

ocorrências de ruptura de cada intervalo e dividirmos pelo número de itens constante na Tabela 29, percebemos um número muito próximo de casos por produto. Embora o intervalo de reposição entre 8 e 14 dias apresente aproximadamente 12 ocorrências por produto, os intervalos a partir de 22 dias de reposição se mantêm em 8 casos. Se por um lado, o menor intervalo implique em maior quantidade de reposições e, conseqüentemente, mais oportunidades de ocorrência de ruptura, a diminuição dos intervalos superiores não impacta neste sentido. Este fato sugere que o índice de ruptura aumenta com intervalos de reposição maiores.

Mais uma vez, observamos que a acurácia do modelo diminui à medida que o tempo em que o item permanece em ruptura aumenta. No entanto, ao analisarmos a relação entre o número de dias em ruptura e o prazo médio de reposição, observamos que o período de ruptura representa entre 21% e 33% do prazo de reposição, indicando que, em média, os itens estão indisponíveis por cerca de 26% do tempo entre as reposições. Já o tempo médio para identificar essa ruptura, para intervalos de reposição acima de 7 dias, corresponde a apenas 11% do prazo de reposição. Esses resultados, combinados com a consistência na identificação da ruptura em diferentes intervalos, destacam a eficácia do algoritmo e sugerem que sua aplicação no ambiente empresarial pode resultar em melhorias significativas.

A seleção de itens para o estudo considerou suas características a fim de evitar distorções causadas por reposições constantes ou por sazonalidades. Porém, podemos considerar a avaliação de acordo com as características dos itens selecionados, utilizando a estrutura mercadológica da empresa:

Tabela 31 – Rupturas por Categoria de Produtos

Categoria	Itens Analisados	Ocorrências de Ruptura	Dias para Reposição	Ruptura Identificadas	% Identificado
LIMPEZA	199	1.805	54	955	53%
BOMBONIERE	193	1.207	61	807	67%
BISCOITOS	175	1.280	46	864	68%
HIGIENE E BELEZA	158	1.106	70	453	41%
BEBIDAS	147	1.265	39	845	67%
CONSERVAS	145	1.158	55	699	60%
IOGURTES	128	2.145	16	1245	58%
EMPACOTADOS	83	702	32	442	63%
MARGARINAS	63	834	24	525	63%
CEREAIS	60	516	61	265	51%
SUCOS	60	509	44	313	61%
CHAS E ERVAS	58	513	50	342	67%
MASSAS	54	408	49	232	57%
ANIVERSÁRIO	50	347	47	212	61%
CONGELADOS	50	667	21	394	59%
SALGADINHOS	48	499	23	344	69%
CAFÉ	45	323	44	187	58%
BEBIDAS QUENTES	62	481	44	288	60%
PET SHOP	34	275	44	177	64%
LEITES	33	398	25	319	80%
DESCARTÁVEIS	26	168	58	77	46%
DOCES E GELEIAS	26	216	48	135	63%
PAPEL HIGIÊNICO	22	167	26	145	87%
MATINAIS	21	131	41	100	76%
DEMAIS	82	696	44	494	71%

Fonte: Elaborado pelo autor.

A análise da Tabela 31, novamente observamos o impacto do prazo médio de reposição no nível de precisão do modelo. Enquanto as categorias “Papel Higiênico” e “Leites”, com prazos de reposição de 26 e 25 dias, respectivamente, alcançaram uma taxa de identificação superior a 80%, a categoria “Higiene e Beleza”, que possui o maior intervalo entre as reposições (70 dias), teve uma taxa de identificação de apenas 41% das ocorrências. Embora algumas categorias com prazos médios de reposição próximos de 60 dias tenham obtido taxas de identificação superiores a 60%, fica evidente mais uma vez uma relação entre o intervalo de reposição e a taxa de identificação. Outras características das categorias de produtos também podem influenciar essa identificação, como os níveis de estoque e seu valor unitário, no entanto, esses aspectos não estão incluídos no escopo do estudo.

Quanto ao impacto financeiro do modelo, devemos observar o incremento com cautela. A identificação tem a taxa de 12%, o que significa que este é o percentual máximo que pode ser sanado pelo modelo. Se deixarmos de lado o percentual real, de 16,9% e trabalharmos com o que foi efetivamente identificado, temos que 88% dos itens estavam disponíveis no ponto de venda. Através de uma regra simples de proporção, chegamos em um cenário de 13,6% de incremento no faturamento dos itens selecionados.

Porém, outros fatores também impactam nesta avaliação. O valor médio diário dos itens com e sem ruptura é diferente. Em geral, itens de maiores vendas costuma ter acompanhamentos constantes para minimização de indisponibilidades, o que não é verdade para itens de menor procura. Em nossa análise, o valor médio diário dos itens sem ruptura foi de R\$ 31,90 enquanto o valor dos itens com ruptura foi de R\$ 27,23. Desta forma, a venda diária dos itens geralmente em ruptura representa apenas 85% do valor dos itens disponíveis, o que reduz o incremento de faturamento para 11,6%.

Adotando um cenário conservador, onde após a identificação da ruptura ainda seja necessário um intervalo de 5 dias até a reposição efetiva do produto na filial podemos estimar, a partir da Tabela 28, o percentual de diminuição a partir das identificações do modelo:

Tabela 32 – Diminuição das Rupturas

Ruptura	Ocorrências	Média Ruptura	Dias para Correção	Não Identificados	% Diminuição
Até 7 dias	11.455	3	7	5.189	0%
8 a 14 dias	2.512	11	9	554	14,0%
15 a 28 dias	3.275	20	10	543	43,0%
29 a 56 dias	1.061	39	15	373	40,5%
Mais de 56 dias	574	85	23	298	35,4%
Geral	18.877	32	13	6.957	30,6%

Fonte: Elaborado pelo autor.

Desta forma, ao estimarmos uma melhoria de 30,6% na correção dos casos de ruptura, chegamos a um incremento de faturamento, apenas nos itens analisados de 3,6%. Ainda em termos de resultados, se considerarmos que estamos analisando itens que representam apenas 37% das vendas do período, o incremento no faturamento total pode ser superior a 1,3%.

De acordo com estudos anteriores, as perdas estimadas para o varejo em virtude das rupturas são de aproximadamente 1,5% do total de receitas (BAYLE-TOURTOULOU; LAURENT; MACÉ, 2006; PAPAKIRIAKOPOULOS; PRAMATARI; DOUKIDIS, 2009). Embora nosso resultado esteja muito próximo deste índice, é importante ressaltar que a comparação não é direta devido à diferença nas taxas de ruptura encontradas. A taxa de ruptura em nosso estudo de caso é significativamente superior àquelas utilizadas nas estimativas anteriores. Além disso, nosso cálculo considerou a possibilidade de melhoria, ao passo que as pesquisas anteriores foram realizadas em um cenário idealizado. Essa diferença de contextos deve ser levada em consideração ao interpretar e comparar os resultados.

Assim, após avaliação sob diferentes aspectos, além do modelo ser eficiente na rápida e correta identificação dos casos de ruptura de maneira geral, sua utilização prática também se mostra positiva a nível financeiro. Mesmo adotando uma abordagem conservadora, uma vez que a estrutura da empresa contendo uma central de distribuição pode ser mais eficiente e aumentar o percentual de melhoria, o modelo pode trazer incrementos significativos no faturamento. Portanto, ele demonstra eficácia em atingir os objetivos propostos tanto a nível acadêmico quanto para o ambiente empresarial.

4 CONSIDERAÇÕES FINAIS

Os índices de ruptura desempenham um papel crucial no contexto do varejo, pois refletem a capacidade das empresas em gerenciar suas cadeias de suprimentos de forma eficaz. Uma gestão eficiente nesse aspecto pode se traduzir em uma vantagem competitiva significativa para as organizações. Assim, reduzir as ocorrências de ruptura permite às empresas melhorarem a satisfação do cliente, garantindo a disponibilidade dos produtos e respondendo de maneira ágil e eficaz às demandas do mercado. Essa abordagem não apenas fortalece o relacionamento com os clientes, mas também contribui para a maximização do desempenho e dos resultados financeiros das empresas no setor varejista.

A redução da ruptura é um desafio que requer, primeiramente, sua identificação precisa e oportuna. No entanto, como revela a revisão de literatura, poucos estudos se dedicaram a abordar essa questão de forma abrangente. A maioria desses estudos ainda dependem de métodos manuais para identificar a ruptura, o que pode ser suscetível a erros humanos e inconsistências nos dados. Além disso, mesmo os estudos que avançaram para o uso de métodos analíticos enfrentam limitações, muitas vezes restringindo suas avaliações a itens de alto giro ou a um número reduzido de produtos. Essas restrições podem limitar a eficácia das análises e dificultar a identificação precisa de suas ocorrências.

Ainda que a linha de pesquisa apoiada em métodos analíticos esteja avançando e possa minimizar suas limitações, especialmente no que diz respeito ao número de itens analisados, sua implementação muitas vezes requer recursos computacionais significativos. Isso se deve principalmente ao emprego de algoritmos robustos de aprendizado de máquina, os quais requerem considerável capacidade computacional para funcionar de maneira eficaz. Quando necessário o processamento em grande escala, estes recursos estão associados a altos investimentos, que muitas vezes não estão ao alcance tanto de empresas quanto de pesquisadores.

No entanto, como alternativa a esse cenário que demanda grandes investimentos em infraestrutura computacional, o presente estudo adotou técnicas de *soft computing*, especificamente algoritmos de lógica difusa, como ferramenta para o desenvolvimento dos métodos analíticos. Essa abordagem permitiu obter resultados

consistentes na identificação e análise de ocorrências de ruptura sem a necessidade de investimentos significativos em poder computacional.

Outra vantagem da metodologia empregada neste estudo é o controle explícito das variáveis de entrada e de saída. Em contraste com certos métodos de aprendizado de máquina, nos quais não há controle direto sobre a saída, uma vez que as regras são definidas internamente pelo algoritmo, a abordagem baseada em lógica difusa permite a definição explícita das regras a serem aplicadas. Embora mais trabalhosa, essa abordagem facilita a detecção de inconsistências e possibilita ajustes conforme as regras de negócio previamente observadas.

O estudo de caso realizado por meio desta abordagem impôs limitações quanto ao número de itens, com base em características que poderiam potencialmente distorcer seus resultados. A análise de 2.022 itens representa uma significativa ampliação em relação aos estudos anteriores, que estavam principalmente limitados a verificações ou confirmações manuais. Apesar das restrições impostas, o estudo alcançou uma avaliação abrangente, uma vez que esses itens correspondem a aproximadamente 37% do faturamento total do período analisado.

A pesquisa empregou exclusivamente dados transacionais para a identificação da ruptura durante o período avaliado. Essa identificação prévia partiu de conceitos relacionados à ruptura total, definida como a falta efetiva de estoque, o que permitiu o uso de dados históricos para a construção e avaliação do modelo proposto. Com base nessa identificação, observou-se que a empresa apresentou uma taxa de ruptura de 16,9% durante o período, considerando apenas os itens analisados.

Embora essa taxa seja significativamente superior aos percentuais entre 5% e 10% relatados em estudos anteriores, é importante ressaltar que o contexto não é diretamente comparável. Esses estudos geralmente abordam o varejo em geral, enquanto nossa análise se concentra especificamente no varejo alimentar, onde as características dos produtos são distintas. Além disso, é necessário considerar que esses dados podem não refletir a situação atual. Desde a publicação destes estudos, ocorreram diversas mudanças no cenário. Por um lado, podem ter ocorrido melhorias nos sistemas de gestão durante o período, mas também houve um incremento significativo no número de itens disponíveis. Essas mudanças podem ter impactado negativamente no controle de estoque e, conseqüentemente, podem refletir em um aumento na taxa de ruptura.

Quanto ao objetivo principal do estudo, que era a elaboração de um algoritmo para identificação dos casos de ruptura total de estoque causada pela falta efetiva do produto, o mesmo obteve uma taxa de identificação correta de 70,2% dos casos. O modelo, baseado em lógica difusa, utilizou como variáveis de entrada probabilidades já associadas a casos de ruptura em estudos anteriores, obtendo 75,4% de precisão e 78% de acurácia, com apenas 16,4% de falsos positivos em sua utilização.

Embora a taxa de identificação tenha ficado abaixo da obtida em estudos anteriores, a abordagem considerou um maior número de itens e não restringiu pelo volume de vendas. Além disso, as confirmações dos estudos anteriores dependiam de confirmações manuais, o que pode levar a distorções significativas nos resultados. A obtenção de um baixo índice de identificações falsas foi uma preocupação central desenvolvimento do algoritmo, uma vez que altos níveis de alarmes falsos, além de custosos, tendem a invalidar a utilidade prática de determinados modelos.

Outro ponto positivo do modelo foi a rápida identificação das ocorrências de ruptura. Em média, uma ruptura foi identificada apenas 4 dias após o seu início. Considerando que o tempo médio em que um produto permanece em estado de ruptura é de 11 dias, isso representa um potencial de melhoria de até 64%. Este índice é fundamental para a redução da taxa de ruptura, uma vez que somente após a identificação inicia-se o processo de reposição, que pode levar vários dias para ser concluído.

Em termos do impacto financeiro, em cenários conservadores, estimamos um impacto no faturamento total de aproximadamente 1,6%, considerando apenas os itens analisados. Em comparação, estudos anteriores que consideram a ruptura geral do varejo indicam perdas de faturamento em torno de 1,5%. É importante notar que as taxas de ruptura citadas nestes estudos estão abaixo das identificadas pelo modelo. Enquanto os estudos indicam taxas entre 5% e 10%, o modelo aponta uma taxa de ruptura de aproximadamente 12% em sua execução. Portanto, os resultados obtidos não só são promissores, mas também válidos, indicando um possível incremento no faturamento da empresa.

É importante destacar que os benefícios financeiros derivados da implementação do modelo desenvolvido estão intrinsecamente ligados à sua integração efetiva com os sistemas de gestão da empresa. Para garantir o pleno funcionamento do modelo, os dados transacionais devem ser avaliados diariamente,

possibilitando a geração precisa de alertas que serão automaticamente encaminhados para a área responsável. Essa integração contínua e coordenada entre o modelo desenvolvido e os sistemas existentes é fundamental para maximizar os benefícios financeiros e operacionais resultantes da sua implementação.

Em termos de limitações da presente pesquisa, a limitação do número de itens precisa ser destacada. Trabalhos futuros podem buscar formas eficazes de lidar com itens não considerados neste estudo, como itens sazonais e de altamente perecíveis. Embora a identificação desses itens em tais cenários possa não ter impactos significativos no curto prazo, ela será fundamental para ajustes nas previsões de demanda de pedidos futuros, prevenindo assim novas ocorrências de ruptura. Desta forma, estudos visando uma implementação conjunta de previsão de demanda e identificação de ruptura parecem ser uma direção promissora para investigações futuras.

Estudos futuros ainda podem abordar o desenvolvimento de modelos de identificação de ruptura integrados ao restante da cadeia de suprimentos envolvida. Com isso, seria possível dimensionar com precisão o impacto da identificação precoce de casos de ruptura, permitindo uma compreensão mais completa do cenário real.

Além disso, como complemento a este estudo, pesquisas que aplicam o modelo desenvolvido em diferentes empresas e setores do varejo poderiam enriquecer ainda mais o algoritmo. Essas pesquisas podem explorar a inclusão de novas variáveis e servirão para validar e generalizar o modelo desenvolvido, ampliando assim sua aplicabilidade e confiabilidade em diferentes contextos e cenários do mercado.

Ao concluir esta pesquisa, almeja-se não apenas contribuições para o ambiente empresarial, mas também avanços significativos no campo acadêmico ao oferecer suporte para o desenvolvimento de modelos baseados em lógica difusa. Esses modelos têm o potencial de aprimorar não apenas a identificação de rupturas no estoque, mas também a resolução de uma ampla gama de problemas em diversos setores. Ao integrar abordagens de *soft computing*, esta pesquisa pode abrir novas perspectivas e oportunidades em diversos cenários tanto empresariais quanto acadêmicos.

REFERÊNCIAS

AGUIAR, F. H. O. DE; SAMPAIO, M. Definição de processo para tratar a ruptura de estoque no varejo de alimentos. **Revista GEPROS - Gestão da Produção, Operações e Sistemas**, v. 8, n. 2, p. 67–67, 23 maio 2013.

AHN, S. et al. A Fuzzy Logic Based Machine Learning Tool for Supporting Big Data Business Analytics in Complex Artificial Intelligence Environments. **IEEE International Conference on Fuzzy Systems**, v. 2019- June, 1 jun. 2019.

AMUTHA, R.; KHAN, A. A. Customer Segmentation using Machine Learning Techniques. **Tuijin Jishu/Journal of Propulsion Technology**, v. 44, n. 3, p. 2051–2061, 11 set. 2023.

ANDAUR, J. M. R.; RUZ, G. A.; GOYCOOLEA, M. Predicting Out-of-Stock Using Machine Learning: An Application in a Retail Packaged Foods Manufacturing Company. **Electronics 2021**, Vol. 10, Page 2787, v. 10, n. 22, p. 2787, 14 nov. 2021.

ARMSTRONG, J. S. Principles of forecasting: a handbook for researchers and practitioners. **Principles of forecasting: a handbook for researchers and practitioners**, International series in operations research & management science. 2001.

AXSÄTER, S. Exact Analysis of Continuous Review (R,Q) Policies in Two-Echelon Inventory Systems with Compound Poisson Demand. <https://doi.org/10.1287/opre.48.5.686.12403>, v. 48, n. 5, p. 686–696, 1 out. 2000.

BARRICHELLO, A. et al. COMPORTAMENTO DO CONSUMIDOR FRENTE À RUPTURA DE ESTOQUE. **South American Development Society Journal**, v. 5, n. 13, p. 108, 7 abr. 2019.

BAYLE-TOURTOULOU, A.-S.; LAURENT, G.; MACÉ, S. Assessing the Frequency and Causes of Out-of-Stock Events Through Store Scanner Data. **Les Cahiers de Recherche**, v. 830, 2006.

BERGER, R. Optimal Shelf Availability-Increasing Shopper Satisfaction at the Moment of Truth. **Scientific Research Publishing**, p. 10–23, 2003.

BEZDĚK, V. Using Fuzzy Logic in Business. **Procedia - Social and Behavioral Sciences**, v. 124, p. 371–380, 20 mar. 2014.

CADAVID, J. P. U.; LAMOURI, S.; GRABOT, B. Trends in machine learning applied to demand & sales forecasting: A review. **International conference on information systems, logistics and supply chain**. 2018.

CHANDRASEKARAN, M. et al. Application of soft computing techniques in machining performance prediction and optimization: a literature review. **The International Journal of Advanced Manufacturing Technology**, v. 46, p. 445–464, 2010.

CHUANG, H. H. C.; OLIVA, R.; LIU, S. On-Shelf Availability, Retail Performance, and External Audits: A Field Experiment. **Production and Operations Management**, v. 25, n. 5, p. 935–951, 1 maio 2016.

CHUANG, H. H.-C.; OLIVA, R. Inventory record inaccuracy: Causes and labor effects. **Journal of Operations Management**, v. 39–40, n. 1, p. 63–78, 1 nov. 2015.

COELHO, A. G. D. C. et al. Inovando a Gestão de Rupturas e Percepção do Risco no Varejo de Autosserviço. **Simpósio Internacional de Gestão de Projetos, Inovação e Sustentabilidade**, 2015.

DEHORATIUS, N.; MERSEREAU, A. J.; SCHRAGE, L. Retail Inventory Management When Records Are Inaccurate. **Manufacturing & Service Operations Management**, v. 10, n. 2, p. 257–277, 11 abr. 2008.

DEHORATIUS, N.; RAMAN, A. Inventory Record Inaccuracy: An Empirical Analysis. <https://doi.org/10.1287/mnsc.1070.0789>, v. 54, n. 4, p. 627–641, 1 abr. 2008.

DIELS, J. L.; WIEBACH, N.; HILDEBRANDT, L. The impact of promotions on consumer choices and preferences in out-of-stock situations. **Journal of Retailing and Consumer Services**, v. 20, n. 6, p. 587–598, 1 nov. 2013.

DROHOMERETSKI, E.; FAVARETTO, F. UM LEVANTAMENTO DAS CAUSAS E EFEITOS DA FALTA DE ACURACIDADE NOS ESTOQUES: UM ESTUDO EXPLORATÓRIO. **Revista Gestão Industrial**, v. 6, n. 2, 30 jun. 2010.

FRANK, C. et al. Forecasting women's apparel sales using mathematical modeling. **International Journal of Clothing Science and Technology**, v. 15, n. 2, p. 107–125, 2003.

GORDON, S. R.; GORDON, J. R. **Sistemas de Informação: Uma Abordagem Gerencial**. [s.l.] Grupo Gen-LTC, 2000.

GRANT, D. B.; FERNIE, J. Exploring out-of-stock and on-shelf availability in non-grocery, high street retailing. **International Journal of Retail and Distribution Management**, v. 36, n. 8, p. 661–672, 2008.

GRUEN, T. W.; CORSTEN, D. A Comprehensive Guide To Retail Out-of-Stock Reduction In the Fast-Moving Consumer Goods Industry. **Grocery Manufacturers of America**, 2007.

HANKE, J. E.; WICHERN, D. W. **Business Forecasting**. 9. ed. Nova York: Pearson, 2013.

HAUSRUCKINGER, G. Approaches to measuring on-shelf availability at the point of sale. **Preuzeto sa http://ecr-all.org/content/ecropedia_element.php**, 2006.

HEESE, H. S. Inventory Record Inaccuracy, Double Marginalization, and RFID Adoption. **Production and Operations Management**, v. 16, n. 5, p. 542–553, 10 set. 2007.

HILL, R. M. On the suboptimality of (S-1, S) lost sales inventory policies. **International Journal of Production Economics**, v. 59, n. 1-3, p. 387-393, 1 mar. 1999.

HUNG, H.-C.; CHUANG, Y.-J.; WU, M.-C. Customizable and committee data mining framework for stock trading. **Applied Soft Computing**, v. 105, p. 107277, 1 jul. 2021.

IBRAHIM, D. **An Overview of Soft Computing**. Procedia Computer Science. **Anais...Elsevier B.V.**, 2016.

KÖK, A. G.; SHANG, K. H. Inspection and Replenishment Policies for Systems with Inventory Record Inaccuracy. <https://doi.org/10.1287/msom.1060.0136>, v. 9, n. 2, p. 185-205, 1 abr. 2007.

KUCUK, S. U. Reducing the Out-of-Stock Costs in a Developing Retailing Sector. **Journal of International Consumer Marketing**, v. 16, n. 3, p. 75-104, 9 ago. 2004.

LAMBERT, D. M.; STOCK, J. R.; ELLRAM, L. M. **Fundamentals of logistics management**. [s.l.] McGraw-Hill/Irwin, 1998.

LUISA, A.; SILVA, B. **A aplicação do modelo de redes bayesianas para o Gerenciamento de risco de ruptura em cadeias de suprimento**. São Paulo : FGV-EAESP, 2008.

MACKINNEY, W. **Python para Análise de Dados: Tratamento de dados com pandas, numpy e ipython**. 4ª ed. ed. São Paulo: Editora Novatec, 2018.

MAKRIDAKIS, S. Accuracy measures: theoretical and practical concerns. **International Journal of Forecasting**, v. 9, n. 4, p. 527-529, 1 dez. 1993.

MITRA, S.; PAL, S. K.; MITRA, P. Data mining in soft computing framework: a survey. **IEEE Transactions on Neural Networks**, v. 13, n. 1, p. 3-14, 2002.

MONTOYA, R.; GONZALEZ, C. A Hidden Markov Model to Detect On-Shelf Out-of-Stocks Using Point-of-Sale Data. **Manufacturing & Service Operations Management**, v. 21, n. 4, p. 932-948, 6 out. 2019.

PALESTINO, C. B. **BI-Business Intelligence: modelagem e tecnologia**. [s.l.] Axcel Books, 2001.

PAPAKIRIAKOPOULOS, D. Developing a mechanism to support decisions for products missing from the shelf. **Journal of Decision Systems**, v. 20, n. 4, p. 417-441, 2011.

PAPAKIRIAKOPOULOS, D. Predict on-shelf product availability in grocery retailing with classification methods. **Expert Systems with Applications**, v. 39, n. 4, p. 4473-4482, 1 mar. 2012.

PAPAKIRIAKOPOULOS, D.; PRAMATARI, K.; DOUKIDIS, G. A decision support system for detecting products missing from the shelf based on heuristic rules. **Decision Support Systems**, v. 46, n. 3, p. 685–694, 1 fev. 2009.

PECKHAM, J. O. The Consumer Speaks. **Journal of Marketing**, v. 27, n. 4, p. 21–26, 19 out. 1963.

ROSADO, L. et al. Supervised learning for Out-of-Stock detection in panoramas of retail shelves. **IST 2016 - 2016 IEEE International Conference on Imaging Systems and Techniques, Proceedings**, p. 406–411, 7 nov. 2016.

RUIZ, R. Ranking Abras. **SuperHiper**, p. 90–91, maio 2023.

SINGH, D. et al. A Comparative Study on the performance of Soft Computing models in the domain of Data Mining. **International Journal of Advancements in Computer Science and Information Technology**, v. 01, n. 1, 2011.

TON, Z.; RAMAN, A. The Effect of Product Variety and Inventory Levels on Retail Store Sales: A Longitudinal Study. **Production and Operations Management**, v. 19, n. 5, p. 546–560, 1 set. 2010.

UÇKUN, C.; KARAESMEN, F.; SAVAŞ, S. Investment in improved inventory accuracy in a decentralized supply chain. **International Journal of Production Economics**, v. 113, n. 2, p. 546–566, 1 jun. 2008.

USMAN, K.; CAPLICE, C. **Determination of Drivers of Stock-Out Performance of Retail Stores using Data Mining Techniques**. Thesis—[s.l.] Massachusetts Institute of Technology, 2008.

VASCONCELLOS, L. H. R.; SAMPAIO, M. The stockouts study: an examination of the extent and the causes in the São Paulo supermarket sector. **BAR - Brazilian Administration Review**, v. 6, n. 3, p. 263–279, 2009.

VERZELETTI, M. Out-of-Stock-Detection. **GitHub**. <https://github.com/mverzeletti/Out-of-Stock-Detection>, 2024