



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO ACADÊMICO DO AGRESTE  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

ITALO MATHAUS CHAVES SANTOS

**DETECÇÃO DE ANOMALIAS A PARTIR DE APRENDIZADO DE MÁQUINAS  
PARA MARCAÇÃO DE MERCADO DA SOJA *COMMODITY***

CARUARU  
2023

ITALO MATHAUS CHAVES SANTOS

**DETECÇÃO DE ANOMALIAS A PARTIR DE APRENDIZADO DE MÁQUINAS  
PARA MARCAÇÃO DE MERCADO DA SOJA *COMMODITY***

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de mestre em Engenharia de Produção. Área de concentração: Otimização e Gestão da Produção.

Orientador: Prof. Dr. Caio Bezerra Souto Maior

CARUARU

2023

Catálogo na fonte:  
Bibliotecária – Paula Silva - CRB/4 - 1223

S237d Santos, Italo Mathaus Chaves.

Detecção de anomalias a partir de aprendizado de máquinas para marcação de mercado da soja *commodity*. / Italo Mathaus Chaves Santos. – 2023.  
52 f.; il.: 30 cm.

Orientador: Caio Bezerra Souto Maior.

Dissertação (Mestrado) – Universidade Federal de Pernambuco, CAA, Programa de Pós- Graduação em Engenharia de Produção, 2023.

Inclui Referências.

1. Gerenciamento de risco. 2. Cadeia de suprimentos. 3. Aprendizado de máquinas. 4. Soja – Comércio. I. Souto Maior, Caio Bezerra (Orientador). II. Título.

CDD 658.5 (23. ed.)

UFPE (CAA 2023-038)

ITALO MATHAUS CHAVES SANTOS

**DETECÇÃO DE ANOMALIAS A PARTIR DE APRENDIZADO DE MÁQUINAS  
PARA MARCAÇÃO DE MERCADO DA SOJA *COMMODITY***

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Pernambuco, Centro Acadêmico do Agreste, como requisito para a obtenção do título de Mestre em Engenharia de Produção. Área de concentração: Otimização e Gestão da Produção.

Aprovada em: 28/06/2023.

**BANCA EXAMINADORA**

---

Prof. Dr. Caio Bezerra Souto Maior (Orientador)  
Universidade Federal de Pernambuco – UFPE

---

Prof. Dr. Lucimário Gois De Oliveira Silva (Examinador interno)  
Universidade Federal de Pernambuco - UFPE

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Ana Cláudia Souza Vidal de Negreiros (Examinadora Externa)  
Universidade Federal Rural do Semi-Árido - UFERSA

## **AGRADECIMENTOS**

Início agradecendo a Deus por me guiar e capacitar, me habilitando a perseguir todos os meus objetivos.

Agradeço a minha família, em especial a minha querida esposa, de quem recebo todo o suporte e conselhos necessários para continuar minha jornada.

Lembro também dos bons amigos que tenho, aos quais constantemente tenho apoio e solidariedade.

Por fim, com imensa alegria agradeço a todos que fazem parte do PPGEP-CAA, que por meio destes recebi muitos ensinamentos e aprendizados. Ressalto o cuidado com uma educação de qualidade. Especialmente agradeço ao meu orientador, por suas ideias, planejamentos, preocupações e empatia.

## RESUMO

Nos últimos anos, eventos de escala global colocaram em risco variados modelos de negócios, obrigando-os a repensar suas estratégias e parcerias. Tal fragilidade pode ser enfrentada pela aplicação de metodologias capazes de prever prováveis eventos disruptivos por meio da análise de dados. Esse trabalho visa a adoção de técnicas de aprendizado de máquina no campo do gerenciamento de risco da cadeia de suprimentos aplicados na cadeia da soja. O objetivo é auxiliar a indústria nacional na tomada de decisão, a fim de mitigar os riscos de abastecimento com o propósito de manter as operações dessa *commodity*. Foi implementado o modelo *Isolation Forest*, um algoritmo de aprendizado de máquina não supervisionado desenvolvido para detecção de anomalias. As fontes de dados são índices futuros da soja, ligados aos mercados dos Estados Unidos e da China. O estudo utilizou dois tipos de verificação, uma dedicada a investigar eventos capazes de alterar o preço da soja, a outra comparando implementações de diferentes bancos de dados. Os resultados obtidos se mostraram satisfatórios, uma vez que as anomalias detectadas pelo modelo possuíam possíveis correlações com acontecimentos relativos ao mercado da soja.

**Palavras-chave:** gestão de risco da cadeia de suprimentos; aprendizado de máquina; soja.

## ABSTRACT

In recent years, events on a global scale have put various business models at risk, forcing them to rethink their strategies and partnerships. This fragility can be addressed by applying methodologies capable of predicting likely disruptive events through data analysis. This work aims to adopt machine learning techniques in the field of supply chain risk management applied to the soybean chain. The objective is to assist the national industry in decision-making, to mitigate supply risks to maintain operations for this *commodity*. The Isolation Forest model, an unsupervised machine learning algorithm developed for anomaly detection, was implemented. The data sources are soybean futures, indexes linked to the United States and China markets. The study used two types of verification, one dedicated to investigating events capable of changing the price of soybeans, the other comparing implementations of different databases. The results obtained were satisfactory, since the anomalies detected by the model had possible correlations with events related to the soybean market.

**Keywords:** supply chain risk management; machine learning; soybeans.

## LISTA DE ILUSTRAÇÕES

Figura 1-	Cadeia de suprimentos Visão Integrada e Visão Global	15
Figura 2-	Modelo de gerenciamento da cadeia de suprimentos	16
Diagrama 1-	Análise e avaliação de riscos	18
Figura 3-	Alocação global da produção de soja entre 2017 e 2019	19
Gráfico 1-	Produção da soja mensurada em toneladas	20
Gráfico 2-	Ranking: Agricultura, Valor da produção (2019)	21
Figura 4-	Subdisciplinas da IA	23
Figura 5-	Categorias de Aprendizado de Máquina	24
Figura 6-	Exemplo de isolamentos através de particionamento aleatório	26
Figura 7-	Construção <i>iForest</i> aprendida para um conjunto de dados	27
Figura 8-	Efeitos de inundação e mascaramento em um conjunto de dados	28
Fluxograma 1-	Etapas da revisão	29
Gráfico 3-	Número de citações e publicações ao longo do tempo	30
Figura 9-	Mapa de redes de ocorrências das palavras-chave	31
Diagrama 2-	Metodologia do estudo	34
Gráfico 4-	Histórico de preços da Soja em dólares	36
Gráfico 5-	Anomalias detectadas pelo <i>iForest</i>	38
Gráfico 6-	Produção de carne suína da China	39
Gráfico 7-	Ocorrência de anomalias na verificação cruzada	42

## LISTA DE TABELAS

Tabela 1-	Relação dos artigos selecionados	32
Tabela 2-	Etapas de treinamento <i>iForest</i>	37
Tabela 3-	Principais notícias referentes a segunda anomalia	40
Tabela 4-	Quadro de anomalias detectadas	43

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>10</b>
1.1	CONTRIBUIÇÕES .....	12
1.2	OBJETIVOS .....	13
1.2.1	Objetivo geral .....	13
1.2.2	Objetivos específicos.....	13
1.3	ESTRUTURA DA DISSERTAÇÃO.....	13
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA</b> .....	<b>14</b>
2.1	FUNDAMENTAÇÃO TEÓRICA.....	14
2.1.1	Gestão da Cadeia de Suprimentos.....	14
2.1.2	Gerenciamento de Riscos da Cadeia de Suprimentos .....	17
2.1.3	Soja .....	19
2.1.4	Inteligência artificial e Aprendizado de máquinas.....	22
2.1.5	<i>Isolation Forest</i> .....	25
2.2	REVISÃO DA LITERATURA .....	28
2.2.1	Classificação e seleção dos estudos.....	30
<b>3</b>	<b>ESTUDO DE CASO</b> .....	<b>34</b>
3.1	PROCEDIMENTOS METODOLÓGICOS.....	34
3.1.1	Caracterização da pesquisa .....	34
3.1.2	Origem e descrição dos dados .....	35
3.2	APLICAÇÃO DO MÉTODO <i>ISOLATION FOREST</i> .....	36
3.3	VERIFICAÇÃO CRUZADA.....	41
3.4	DISCUSSÕES DOS RESULTADOS.....	44
<b>4</b>	<b>CONCLUSÃO</b> .....	<b>45</b>
	<b>REFERÊNCIAS</b> .....	<b>46</b>

# 1 INTRODUÇÃO

A logística é o principal elo dentro das cadeias de produção, sua estratégia e implantação podem determinar o sucesso ou fracasso de uma atividade de mercado tal qual no atendimento das necessidades econômicas. Novaes (2001) afirma que a logística procura coligar todos os elementos do processo: prazos, integração de setores da empresa e formação de parcerias com fornecedores e clientes; para satisfazer as necessidades e preferências dos consumidores finais. Ballou (2006) aponta que a logística busca administrar eficientemente os fluxos de mercadorias, serviços e informações, também enfatiza que ela requer uma boa estratégia para o seu sucesso. A logística desempenha um importante papel na economia global, abastecendo mercados e consumidores, de modo que eventuais falhas seriam capazes de gerar grandes transtornos.

Durante os últimos anos, eventos e adversidades de escala global colocaram em risco variados modelos de negócios, obrigando-os a repensar suas estratégias e parcerias. As cadeias de suprimentos, especialmente as globais, são comumente afetadas por eventos previsíveis ou imprevistos que ameaçam sua lucratividade e continuidade (BARYANNIS *et al.*, 2019). Analistas como Arminio Fraga, ex-presidente do Banco Central, entendem que o mundo está repensando a maneira de gerir estoques que se encontravam dentro de uma cadeia produtiva internacional integrada que apresentou problemas em anos anteriores (CANZIAN, 2022).

Existem diversos riscos capazes que influenciar negativamente as operações de uma rede de abastecimento. Muitos autores, implícita ou explicitamente, correlacionam a eficiência de uma cadeia de suprimentos com os riscos inerentes a mesma. De fato, Behzadi *et al.* (2018) menciona que isso acontece devido à globalização somada a gestões cada vez mais enxutas, resultando em cadeias de suprimentos mais longas e complexas, expostas a maiores riscos que as tornam mais vulneráveis.

Dentro de uma cadeia de suprimentos, artigos acadêmicos voltados ao gerenciamento de risco da cadeia de suprimentos (GRCS) constantemente destacam o risco de interrupção. Conforme mencionam Aboutorab *et al.* (2022), os riscos de interrupção são definidos como os riscos que têm o potencial de impactar

negativamente as operações entre quaisquer duas empresas na cadeia de suprimentos (CS) Para evitar as consequências decorrentes de tais interrupções, os pesquisadores enfatizaram a necessidade dos parceiros da cadeia de suprimentos serem proativos e não reativos no gerenciamento de riscos operacionais (NIMMY *et al.*, 2022).

O GRCS é analisado em diversos estudos aplicados e metodologias variadas, fato destacado por Kosasih, Brintrup e Alexandra (2021) e Baryannis *et al.* (2019). Este também afirma que, apesar de receber boa atenção, não são tantos trabalhos que aderem ao viés da inteligência artificial ou do aprendizado de máquina como ferramentas para auxiliar o GRCS como, por exemplo, em uma identificação de riscos proativa, apontado por (ABOUTORAB *et al.* 2022) e (RAJESH, 2020). Uma quantidade ainda menor de estudos se propõe a uma abordagem onde apresenta aplicações em casos reais.

Neste contexto, o setor logístico tornou-se primordial para a competitividade empresarial, com diversos processos de transformações visando a otimalidade de suas operações amparadas pelos avanços tecnológicos. Pesquisas específicas sobre os impactos e aplicações das tecnologias da Indústria 4.0 na logística receberam a nomenclatura de “Logística 4.0” ou “Logística Inteligente” (*Smart Logistics*) de acordo com (VOIGT e MÜLLER 2018). Cimini *et al.* (2020) afirma que tal avanço é resultado do aumento do uso da Internet de Coisas, Serviços e Pessoas (IoTSP), que permite a comunicação em tempo real entre produtos, máquinas, serviços e humanos, e o uso de ferramentas digitais avançadas. Bu (2021) expõe que o setor de logística precisa urgentemente de métodos eficazes de mineração de dados, recomendando a utilização de inteligência artificial, e suas técnicas de aprendizado. Essas técnicas de aprendizado utilizam os dados coletados para extrair conhecimento que poderá ser utilizado para aprimorar um sistema ou realizar tarefas.

Deste modo, esse trabalho visa a adoção de aprendizado de máquina no campo do GRCS aplicada na cadeia de soja. A cadeia da soja é altamente relevante para a economia mundial por ter aplicações e poder percorrer tanto o caminho da indústria alimentícia como o da indústria energética. Conforme Tallman *et al.* (2008) apontam, a soja é a cultura mais valiosa dos grãos leguminosas e representa a mais importante fonte mundial de proteína e óleo vegetal. De fato, sua utilização abrange

diversos setores da economia com importante valoração para o tratamento das informações disponíveis que subsidiar seus investimentos e decisões.

Assim, o uso de uma ferramenta proativa para alertar sobre potenciais eventos que afetem os processos empresariais bem como a continuidade da cadeia de suprimentos desse produto será de grande valor. O objetivo desse estudo é auxiliar a indústria nacional que utiliza esse *commodity* como matéria-prima em seus processos na tomada de decisões afim de mitigar os riscos de abastecimento, com o propósito de manter as operações das empresas. A ideia é detectar eventos capazes de afetar a cadeia da soja. O modelo utilizado foi o *Isolation Forest*, um algoritmo criado para detecção de anomalias.

Aqui, bancos de dados com indicadores econômicos (e.g., bolsas de valores sobre a marcação da soja, produção dos países relevantes, além do câmbio de cada país em relação ao dólar (moeda de referência do preço da soja) são utilizados. Tais informações são úteis pois traduzem de maneira rápida as expectativas do mercado sobre o produto em questão. Logo, um evento relevante pode ocasionar em uma queda nas marcações desse *commodity*.

## 1.1 CONTRIBUIÇÕES

O risco de desabastecimento das cadeias globais tem recebido maior relevância, com impactos econômicos diretos. Dessa forma, a contribuição desse trabalho está na implementação e uso de um modelo de aprendizado de máquinas, especializado na detecção de anomalias, aplicado no campo do gerenciamento de risco da cadeia de suprimentos da soja. O propósito é mapear os eventos disruptivos através das cotações de mercado da soja, visto que essas cotações traduzem em tempo real os eventos ligados a soja através dos seus preços. Os resultados podem auxiliar o agronegócio e investidores da agroindústria a gerenciar os riscos dos seus negócios. Ainda, indiretamente, impactos sociais e ambientais podem ser vistos com questões relacionadas ao deslocamento de comunidades locais e concentração de terras além disso de desmatamento e perda de biodiversidade uma vez que o *commodity* utiliza grandes extensões territoriais.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo geral

O objetivo geral deste trabalho é auxiliar a gestão empresarial acerca das decisões sobre a soja como matéria-prima utilizando aprendizado de máquinas para detectar eventos que possam gerar interrupção na cadeia de abastecimento desse *commodity*.

### 1.2.2 Objetivos específicos

Para o alcance do objetivo geral, tem-se os seguintes objetivos específicos:

- a) Levantar a bibliografia referente ao GRCS com aplicações de aprendizado de máquina;
- b) Compreender quais modelos de ML podem ser mais adequados para detectar eventos disruptivos;
- c) Definir o modelo de aprendizado que melhor se aplique na detecção de anomalias do preço da soja;
- d) Implementar o modelo definido para detectar eventos capazes de gerar interrupção na cadeia de abastecimento da soja.

## 1.3 ESTRUTURA DA DISSERTAÇÃO

Este trabalho está estruturado em outros cinco capítulos a seguir:

O Capítulo II trata da fundamentação teórica sobre a qual foi desenvolvida esta dissertação. Inclui também uma revisão de literatura.

O Capítulo III apresenta os procedimentos metodológicos desse estudo.

No Capítulo IV é apresentado o estudo de caso.

O Capítulo V apresenta a aplicação do método e os resultados encontrados.

No Capítulo VI são descritas as conclusões e sugestões de trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA E REVISÃO BIBLIOGRÁFICA

Neste capítulo será apresentado a base conceitual que sustenta esta pesquisa, sendo apresentado no primeiro momento os conceitos de CS. Posteriormente, é abordado a soja como *commodity* e sua relevância para o mercado atualmente. Finalmente, são introduzidos os conceitos de inteligência artificial, aprendizado de máquinas e, mais especificamente, o modelo *Isolation Forest*, aplicado na problemática de detecção de anomalias.

### 2.1 FUNDAMENTAÇÃO TEÓRICA

#### 2.1.1 Gestão da Cadeia de Suprimentos

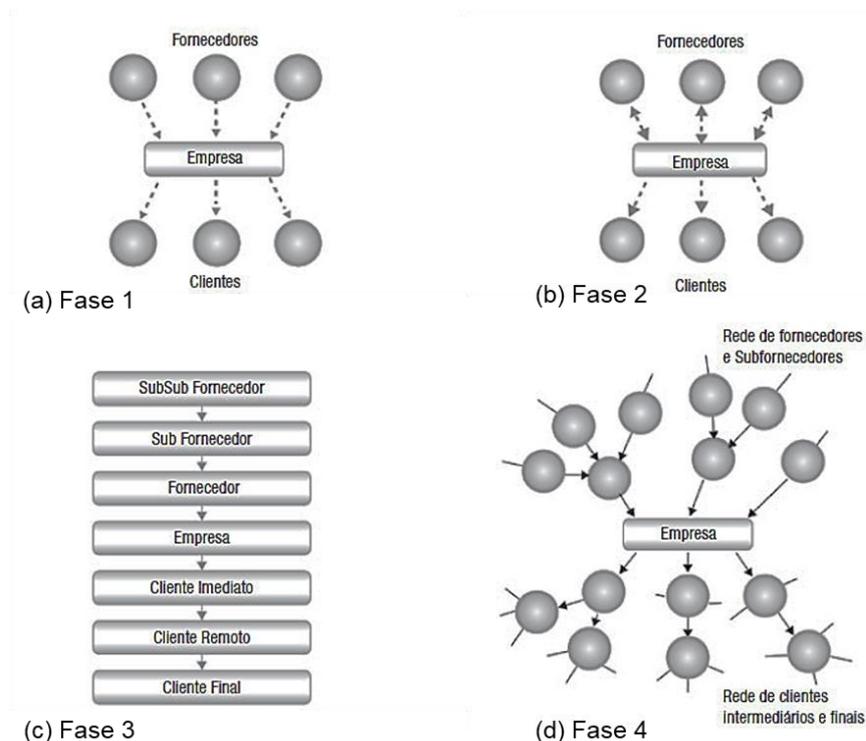
A CS é um sistema de cooperação interempresarial no qual se organizam atividades, pessoas, informações e recursos, tradicionalmente associados a atividades logísticas como o transporte de fornecedores a clientes nas fases posteriores da prestação de serviços. Ballou (2006) afirma que “CS é um conjunto de atividades funcionais (transportes, controle de estoques etc.) que se repetem inúmeras vezes ao longo do canal pelo qual matérias-primas vão sendo convertidas em produtos acabados, aos quais se agrega valor ao consumidor”.

Bowersox *et al.* (2014) menciona que o conceito geral de uma CS integrada costuma ser ilustrado por um diagrama linear que liga as empresas participantes formando uma unidade competitiva coordenada. Este autor também afirma que a vantagem competitiva só será obtida a partir do alinhamento entre os distribuidores, fornecedores, a empresa e os clientes, através de sua estrutura e estratégia da cadeia de suprimentos. Bertaglia (2009) por sua vez aponta que nos últimos anos o conceito de CS vem evoluindo, atualmente é apresenta uma visão mais ampla do que pode se chamar de cadeia logística, pois esta é mais limitada à aquisição e transporte de materiais e distribuição dos produtos finais.

Para Scavarda e Hamacher (2001), a CS é uma rede que engloba todas as empresas que participam das etapas de formação e comercialização de determinado produto ou serviço, que será entregue a um cliente final. Machline (2011) aponta que enquanto a logística concentra-se nas operações da própria empresa, a cadeia de suprimentos olha desde o início até os elos finais da corrente de fornecedores e

clientes. E com uma visão mais ampla e panorâmica do que a visão logística. A Figura 1 traz a evolução da visão empresarial em relação aos parceiros e clientes.

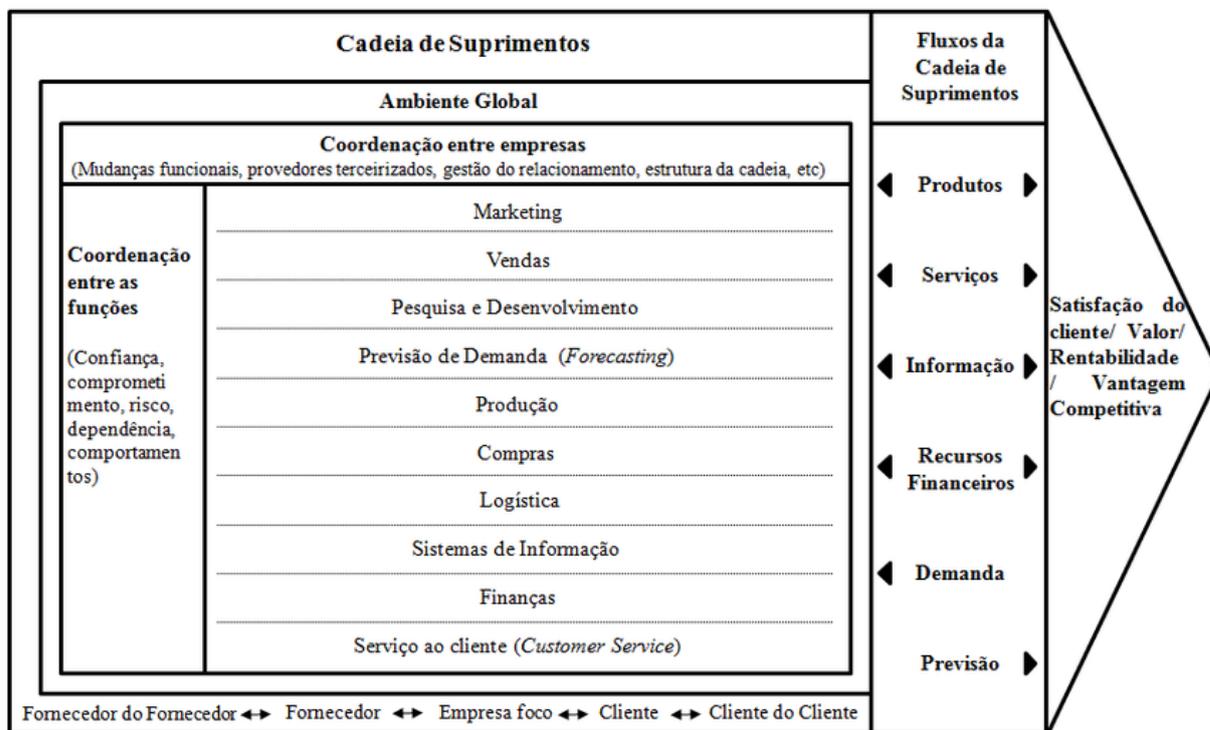
Figura 1- Cadeia de suprimentos Visão Integrada e Visão Global



Fonte: MachLine (2011)

A figura 1 expõe as etapas do conceito de CS no Brasil, isto é, na primeira fase nota-se a visão isolada da empresa, não integrando parceiros; na segunda, visão sistêmica, é o início da integração empresa-fornecedores, momento de melhoria na troca de informações e vantagens, porém se limitava aos parceiros diretamente ligados ao negócio; na terceira fase, o relacionamento interempresarial é ampliado, onde a empresa buscava informações sobre todo o caminho logístico, denominada visão integrada; por fim, a visão global, que diz respeito a busca pelo desenvolvimento de uma rede de abastecimento e uma maior preocupação com o cliente final, atribuindo maior valor aos feedbacks de cada integrante da cadeia de suprimentos. Neste sentido, Ballou (2006) propôs uma exemplificação de gerenciamento da CS, expressando a diferença entre logística integrada e gerenciamento da cadeia de suprimentos (GCS). Na Figura 2 é mostrado que outras áreas empresariais são integradas ao processo de tomada de decisão de uma cadeia produtiva.

Figura 2- Modelo de gerenciamento da cadeia de suprimentos



Fonte: Ballou (2006)

A CS tem como principal objetivo a satisfação do cliente/consumidor, objetivando agregar valor a cada processo que intermedia a relação entre a oferta e a demanda. Desta forma, todas as áreas devem ser integradas para melhorar a performance da CS. Um exemplo válido é o desenvolvimento da embalagem de um certo produto, áreas de vendas buscarão aspectos que afetem o consumidor em sua decisão de compra, já o setor de logística buscará aspectos que facilitem a guarda e transporte do mesmo produto. Esses dois processos procuram agregar valor à cadeia.

O GCS é a administração dos interesses presentes dentro de uma cadeia de suprimentos. Christopher (2010) compreende como uma rede organizada e conectada responsável por controlar, gerenciar e aperfeiçoar o fluxo de matéria.

Ribeiro e Gomes (2013) afirmam que a gestão da CS é o controle de insumos e dados oriundos do processo envolvendo gestores e a integração de um mix de empresas. A GCS, conforme a perspectiva de Orenstein, Ladik e Rainford (2016), é uma forma de vincular fornecedores, fábricas, armazéns e lojas, a fim de garantir que os produtos sejam entregues na quantidade correta, no lugar certo e no momento certo. Os autores sinalizam também a integração estratégica entre os participantes da

cadeia de suprimentos bem como as trocas de informações. Salieta-se que a forma como essas informações são obtidas e interpretadas podem caracterizar em vantagem competitiva. Em Fleury, Wanke e Figueiredo (2000) são destacadas características que contribuem para o sucesso do GCS, um destes destaques é a criação de um canal de informações capaz de conectar todos os elos da cadeia. Certas vantagens podem ser destacadas a partir desse ponto, como a otimização dos estoques da cadeia.

### 2.1.2 Gerenciamento de Riscos da Cadeia de Suprimentos

GRCS é o estudo que propõem investigar e discutir eventos que, em si ou através destes, tem capacidade de interromper ou onerar de algum modo uma cadeia de abastecimento.

Em uma rede de cadeia de suprimentos ou em uma cadeia de suprimentos fortemente acoplada, eventos críticos como greves, incêndios ou falência em um nó em uma cadeia de suprimentos podem levar a interrupções ou capacidade reduzida na rede de suprimentos. Comum para todos esses exemplos é que várias ameaças afetam a missão do sistema e levam a uma operação para “resolver as coisas” e recuperar a estabilidade (ASBJØRNSLETT, 2009).

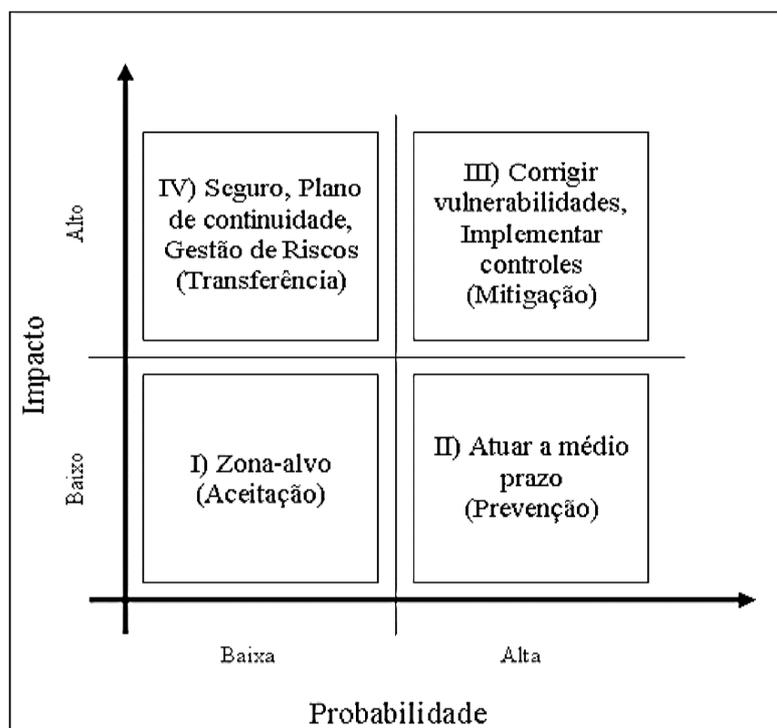
A gestão de riscos é fundamentada em atividades coordenadas para orientar e controlar uma organização em relação aos riscos (OLIVEIRA *et al.*, 2017). Aguiar, Tortato e Gonçalves (2012) explicam que as organizações e as CS são, por natureza, suscetíveis a eventos de riscos, e que deve ser levado em conta os riscos oriundos de fluxos interconectados de materiais, informações e finanças nas relações interempresariais. Craighead *et al.* (2007) propõem que o gerenciamento de risco da cadeia de suprimentos é categorizado em risco operacional e risco de evento de interrupção, isto é, eventos não planejados e imprevistos que interferem no fluxo tradicional dos bens e materiais inseridos em uma CS. Citando caso análogo, temos as oscilações de chuvas em regiões de plantio, onde excessos ou grandes estiagens podem afetar a safra desse *commodity*, desencadeando desequilíbrio entre oferta e demanda.

Norrman e Jansson (2004) sustentam que a relação de risco e recompensas é um fator importante para o bom desempenho das cadeias de suprimentos e sua confiabilidade, fazendo com que o gerenciamento desses riscos se torne um atributo

relevante para a competitividade e sobrevivências das cadeias. A probabilidade e o impacto de eventos ou condições inesperados de nível macro e/ou micro que influenciam adversamente qualquer parte de uma cadeia de suprimentos, levando a falhas ou irregularidades de nível operacional, tático ou estratégico (HO *et al.*, 2015).

Ao tratar da estrutura do gerenciamento de risco Janjua, Nawaz e Prior (2021) entendem que o GRCS envolve identificação de risco, avaliação de risco e avaliação de risco como alguns de seus componentes essenciais. Na visão de Tang e Tomlin (2008) o gerenciamento de risco ainda é acrescentado da análise de decisão (a interferência que as decisões sofrem pelos riscos em diferentes cenários), da mitigação do risco e do planejamento de contingência. Behzadi *et al.* (2018) pontuam que inicialmente o processo de identificação e avaliação de riscos deve ser abrangente, de tal modo que possa avaliar todos os tipos de riscos, incluindo os riscos ocultos, por causa de seus possíveis impactos disruptivos na CS. O diagrama 1 representa o diagrama para análise e avaliação dos riscos.

Diagrama 1- Análise e avaliação de riscos



Fonte: Asbjornslett (2008)

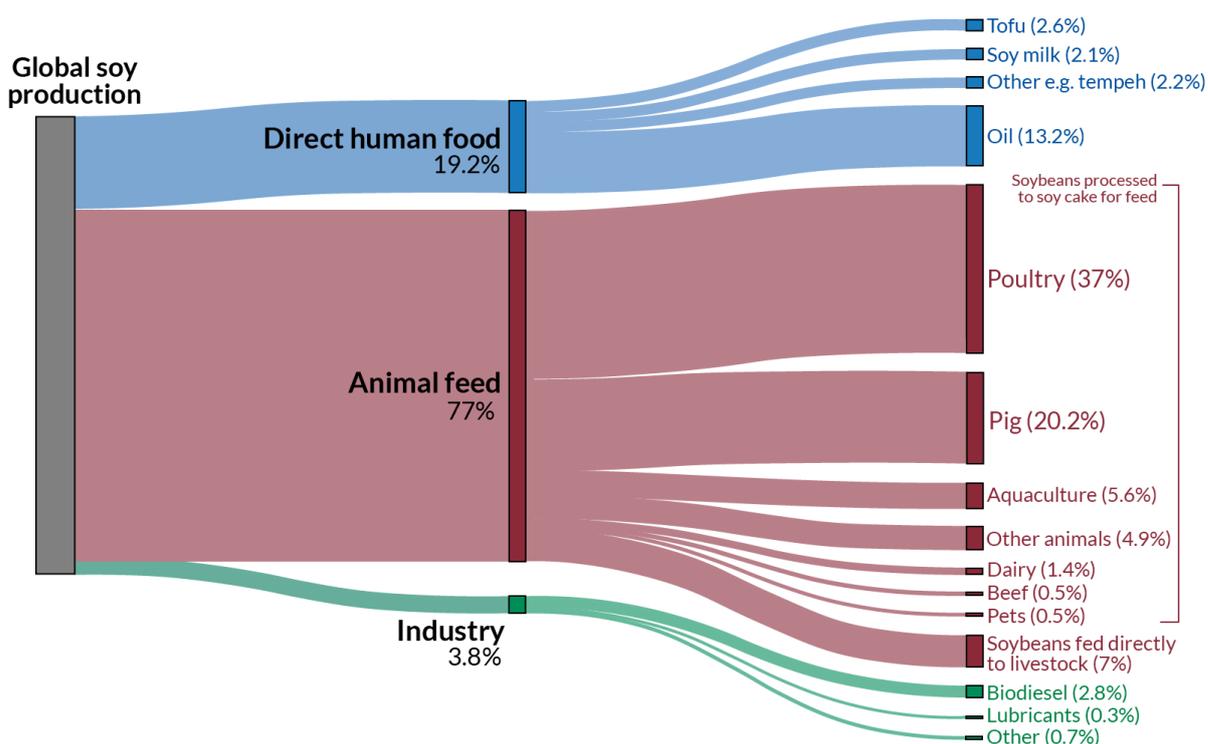
A figura 3 mostra a matriz de risco, um método frequentemente utilizado para análise de riscos. A partir dele é possível propor soluções apropriadas para os riscos

identificados, bem como estabelecer sua priorização, dado sua probabilidade de ocorrência e o potencial de impacto. Esse tipo de classificação auxilia na adoção de qual estratégia poderá ser aplicada diante do risco, por exemplo: transferir, mitigar, prevenir e aceitar os riscos, dentre outras.

### 2.1.3 Soja

A soja pode ser considerada um dos *commodities* essenciais frente a cadeia global. A soja é uma *commodity* agrícola amplamente consumida em muitas formas, como a soja integral, óleo de soja e mercados de farelo de soja (FEARNSIDE, 2001). Mais de três quartos (77%) da soja global é fornecida ao gado para produção de carne e laticínios. A maior parte do restante é usada para biocombustíveis, indústria ou óleos vegetais (RITCHIE e ROSER, 2021). O fato é representado na figura abaixo:

Figura 3- Alocação global da produção de soja entre 2017 e 2019



Data source: Food Climate Resource Network (FCRN), University of Oxford; and USDA PSD Database. OurWorldinData.org – Research and data to make progress against the world's largest problems.

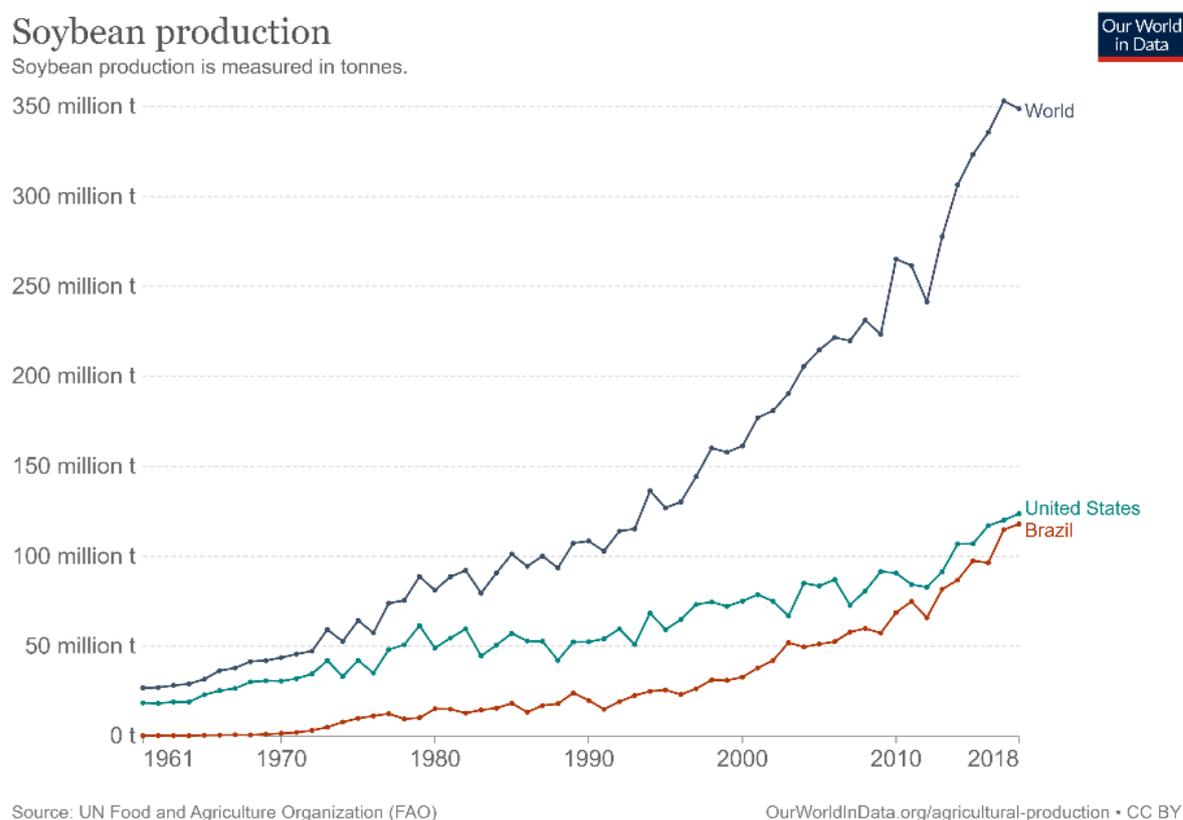
Licensed under CC-BY by the author Hannah Ritchie.

Fonte: OurWorldinData (2021)

A soja inteira pode ser processada para uso industrial (e.g., biocombustível e óleos comestíveis) ou triturados para produzir produtos alimentícios (e.g., leite de soja, ração animal e molho de soja) (WWF, 2014).

A WWF (2014) em seu estudo revelou que a produção da soja é crescente nos últimos 30 anos, com sua produção mundial passando de quase 100 milhões de toneladas para 350 milhões de toneladas. Especificamente, Hirakuri (2020) afirma que a produção brasileira de soja apresentou uma taxa geométrica de crescimento anual de 6,2% entre as safras agrícolas 2000/2001 e 2017/2018, o que fez a quantidade colhida mais do que triplicar, saltando de 38,4 milhões para 119,3 milhões de toneladas. Para ele, dois elementos tiveram grande importância: área e produtividade. O gráfico 1 traz a comparação da produção mundial, dos EUA e do Brasil.

Gráfico 1- Produção da soja mensurada em toneladas



Fonte: OurWorldinData (2021)

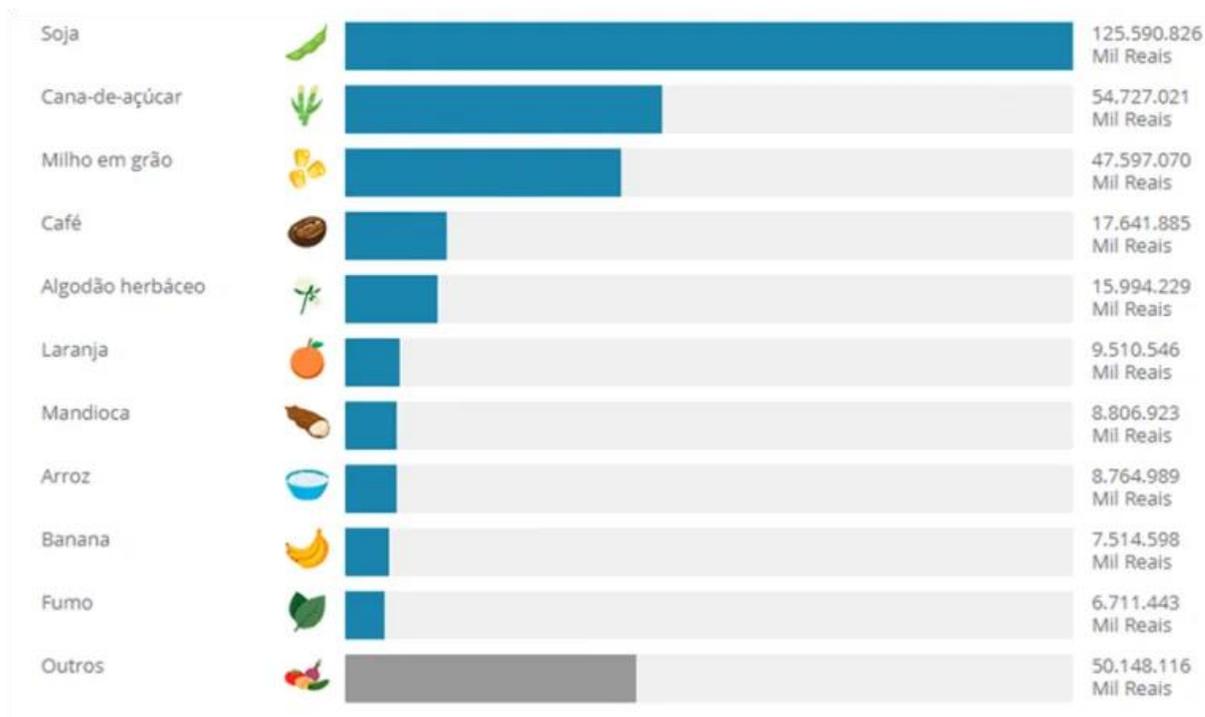
Em 2020 o Brasil alcançou 126 milhões de toneladas, o maior volume de soja produzido no mundo, superando os EUA. O país está entre os maiores exportadores da *commodity*, 84 milhões de toneladas em 2020 foram embarcados (CANAL AGRO, 2022).

Entre as safras 2014/2015 e 2017/2018, as condições climáticas foram favoráveis para grande parte dos principais países produtores de soja, sobretudo Estados Unidos e Brasil, propiciando produções mundiais substanciais. Nesse cenário, os preços da soja em grão recuaram

significativamente em 2015, mantendo relativa estabilidade até dezembro de 2018. (HIRAKURI, 2020)

No gráfico 2, é mostrado o volume financeiro da soja em comparação a outros produtos agrícolas.

Gráfico 2- Ranking: Agricultura, Valor da produção (2019)



Fonte: IBGE (2019)

A liderança da soja na agricultura brasileira se deve principalmente pelo retorno econômico e versatilidade do grão, que pode ser utilizado pela indústria (STOLLER, 2021). Pode-se afirmar que o valor de produção da soja é maior que dos outros três *commodities* seguintes, expondo a importância da soja para o agronegócio brasileiro.

Para 2029, a projeção é de aumento de 32% da produção, 22% do consumo e 41% das exportações (APROSOJA BRASIL, 2020). Nota-se que cada vez mais a soja aumentará sua presença no agronegócio nacional e global impulsionando sua cadeia produtiva. Tais projeções assinalam a significância desse *commodity*, em vista disso é pertinente que as organizações envolvidas nessa cadeia produtiva adquiram ferramentas que as auxiliem em seu planejamento de abastecimento.

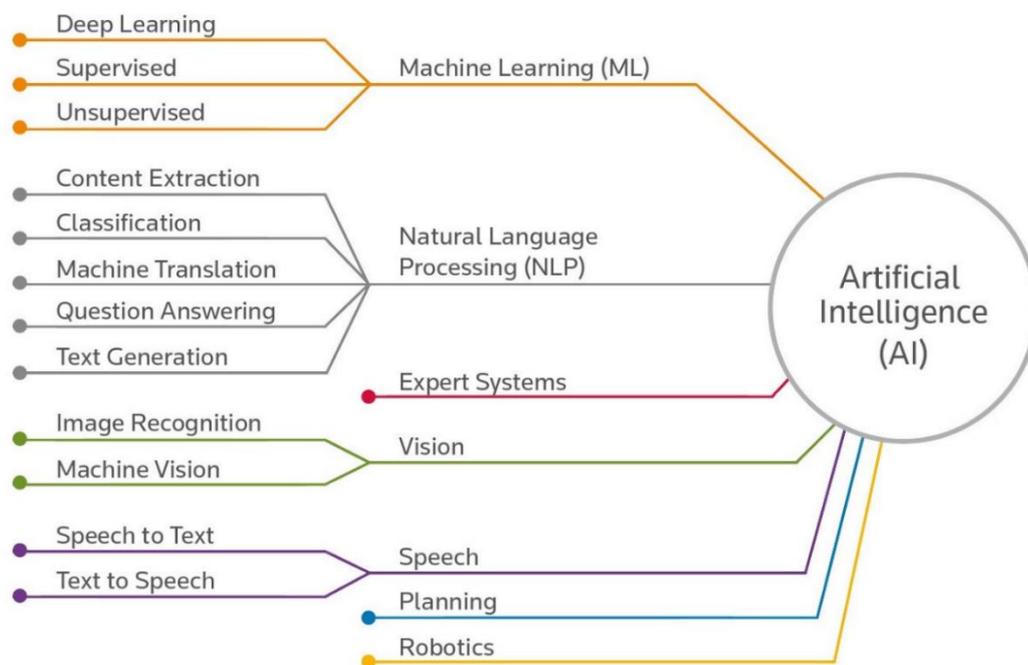
#### 2.1.4 Inteligência artificial e Aprendizado de máquinas

A Inteligência Artificial (IA) evoluiu nas últimas décadas de uma simples partida de xadrez feita pelo supercomputador da IBM Deep Blue em 1996, até numa modernização da medicina ajudando em diagnósticos mais precisos.

De acordo com autores como Bhattad e Jain (2020), Dos e Gomes (2010) e Enholm *et al.* (2021), a IA é uma forma de inteligência criada que pode ser reproduzida em uma máquina e que se refere à capacidade de dar ao computador habilidades que normalmente requerem inteligência humana. A IA é construída por meio de teorias e técnicas que permitem a criação de máquinas capazes de simular inteligência. A IA é caracterizada por sua capacidade de interpretar corretamente dados externos, aprender com esses dados e usar esses aprendizados para atingir metas e tarefas específicas por meio de adaptação flexível. Ela é capaz de analisar adequadamente os dados e gerar soluções a partir deles, proporcionando benefícios a quem a utiliza.

Russel e Norvig (2003) colocam que o objetivo da IA é desenvolver modelos computacionais que podem realizar tarefas que requerem inteligência humana. Fan *et al.* (2020) traçam um paralelo entre os objetivos da IA e da neurociência, afirmando que possuem objetivos complementares, dado que entender como o cérebro processa informações, toma decisões e interage com o ambiente serve de subsídio para uma maneira direta de desenvolver a IA. A figura 4 traz um mapa das subdisciplinas da inteligência artificial.

Figura 4- Subdisciplinas da IA



Fonte: Thomson Reuters (2016)

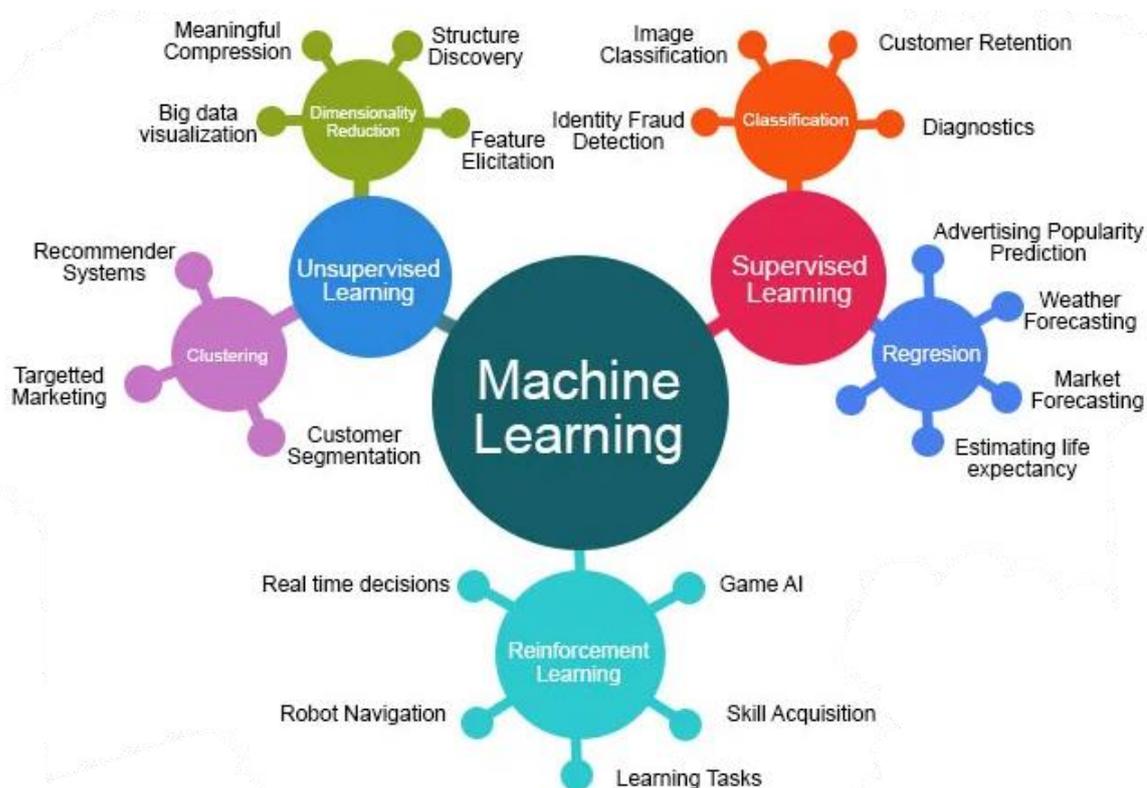
A figura 4 apresenta as principais disciplinas derivadas da IA. No atual contexto, pode ser percebido que a IA com área de conhecimento é multidisciplinar, ou como afirma Howard (2019), é um grande campo da ciência transdisciplinar. Russel e Norvig (2014) incluem lógica, estatística, psicologia cognitiva, teoria da decisão, neurociência, linguística, cibernética e engenharia da computação.

Ao tratar a subdisciplina aprendizado de máquinas (ML, sigla em inglês), Mohri, Rostamizadeh e Talwalkar (2018) este é um termo abrangente para vários métodos diferentes para alcançar a IA. Howard (2019) simplifica como uma subdisciplina da IA que permite que os computadores aprendam com os dados. Jordan e Mitchell (2015) observam que ML pode ser a mais importante ferramenta de IA para obter insights cognitivos, fazer previsões e apoiar a tomada de decisões a partir de um computador. Segundo Wang, Huang e Zhang (2019), o objetivo é treinar uma máquina com os dados apresentados e fazer deduções a partir deles, criando a possibilidade de identificar associações e realizar previsões com base no que foi aprendido.

Para Enholm *et al.* (2021) e Wang, Huang e Zhang (2019) o aprendizado de máquinas possui quatro tipos de algoritmos, adicionando o Aprendizado Semi-supervisionado, entretanto a maioria dos autores frequentemente citam apenas três

tipos de categorias de algoritmos de aprendizado de máquinas, como representado abaixo:

Figura 5- Categorias de Aprendizado de Máquina



Fonte: Educba.com

A figura 5 esquematiza essas categorias de aprendizado, denominados Aprendizado Supervisionado, Aprendizado Não Supervisionado e Aprendizado Reforçado. Em cada modelo estão aglomerados algoritmos que seguem a metodologia de cada tipo de aprendizado.

No Aprendizado Supervisionado (*Supervised Learning*) o conjunto de dados é rotulado, ou seja, para cada característica ou variável independente, há um dado-alvo correspondente que usáramos para treinar o modelo (PEDAMKAR, 2019). O sistema então identifica padrões a partir dos dados de treinamento e infere suas próprias regras a partir dos dados rotulados (AFIOUNI-MONLA, 2019). Em outras palavras, treinamos a máquina com a entrada e a saída correspondente, posteriormente com base no treinamento, a máquina prevê a saída utilizando o conjunto de dados do treinamento.

Antagônico ao anterior, o Aprendizado Não Supervisionado pode ser utilizado com o objetivo de descobrir padrões ocultos no conjunto de dados, com aplicações proeminentes como agrupamento automático, detecção de anomalias e mineração de associação (SCHMIDT *et al.*, 2020). Nesse tipo de aprendizado, o valor alvo não é inserido no treinamento. O sistema tem que analisar a estrutura dos dados de treinamento e suas propriedades estatísticas para resolver o problema (AFIOUNI-MONLA, 2019). Esse mesmo autor comenta que o modelo pode encontrar novas relações que não foram detectadas pelos humanos que o operam. Um método comum de aprendizado não supervisionado é o agrupamento ou agrupamento de dados com base em estruturas ocultas identificadas nele pelo algoritmo (CHOY *et al.*, 2018).

O Aprendizado Reforçado é adaptado da teoria fundamental da aprendizagem, o modelo aprende em cada ação realizada. Há recompensas recebidos por decisões corretas e penalidades por erros cometidos. Pedamkar (2019) avalia que isso permite aprender os padrões e tomar decisões mais precisas sobre dados desconhecidos. Um computador usando aprendizado por reforço está gerando seus próprios dados de treinamento por meio de experimentação e otimização das saídas (TADDY, 2019).

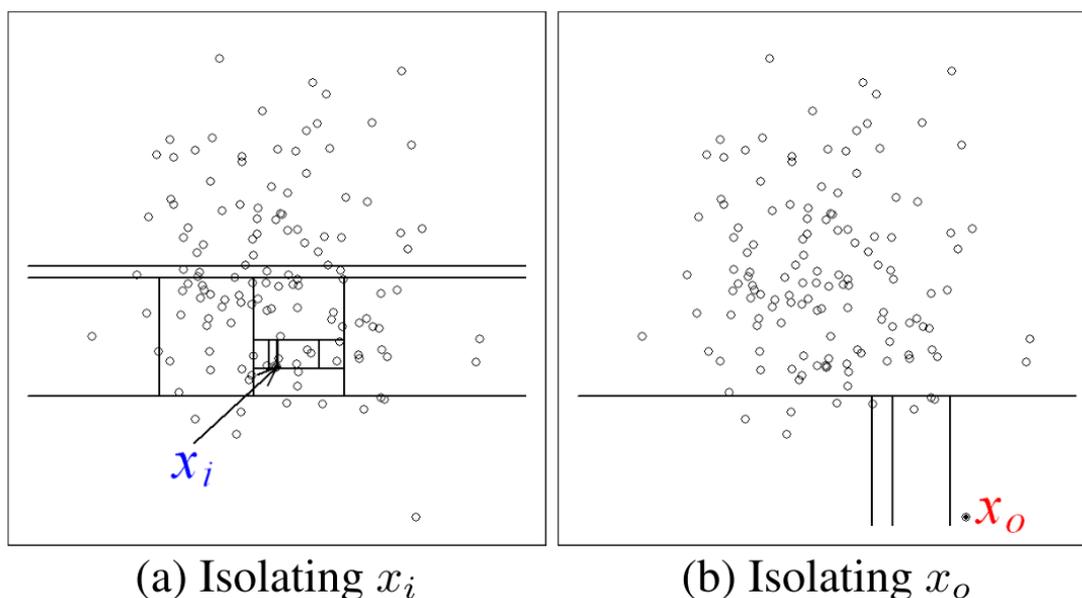
As vantagens proporcionadas pela aplicação de modelos de aprendizado de máquinas aceleram o entendimento sobre áreas e fenômenos pouco compreendidos pelo homem, isso devido principalmente a capacidade de processamento de dados de uma máquina e velocidade em feedbacks. Esse estudo utilizou um modelo de aprendizagem não supervisionado, especificamente voltado para detecção de anomalias.

#### 2.1.5 *Isolation Forest*

Proposto por Fei Tony Liu, Kai Ming Ting e Zhi-Hua Zhou em 2008, o *Isolation Forest* (também chamado de *iForest*) é um algoritmo de detecção de anomalias que utiliza a construção de florestas aleatórias de árvores de decisão para isolar observações anômalas, isto é, separar uma instância do restante das instâncias.

O algoritmo baseia-se no entendimento de que as observações anormais são mais suscetíveis ao isolamento, o que significa que requerem menos partições aleatórias para separá-las de outras observações. A figura 6 expressa como essas partições acontecem.

Figura 6- Exemplo de isolamentos através de particionamento aleatório



Fonte: Tony Liu; Ming Ting; Zhou (2008)

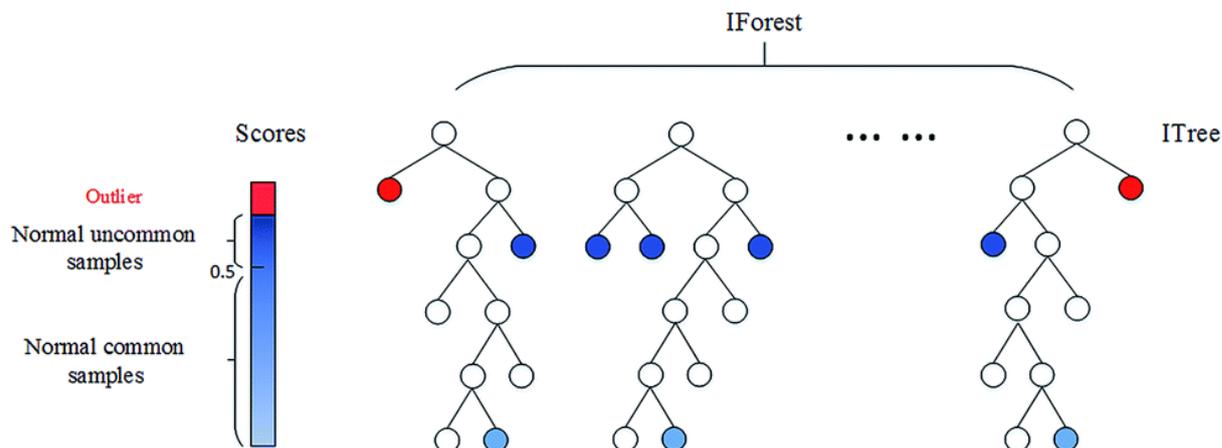
Na figura é apresentado o isolamento de  $x_i$ , o qual demanda de muitos particionamentos aleatórios para ser isolado, se tratando assim de um dado normal. Já no caso de  $x_o$ , poucas partições já produzem o efeito de isolamento, nesse contexto  $x_o$  é tido como anômalo.

Aplica-se duas propriedades quantitativas de anomalias, as quais são dados que consistem em menos instâncias e que possuem valores de atributos distintos das instâncias normais, tornando-as mais passíveis ao isolamento, pois são poucos e diferentes. O *iForest* não utiliza medidas de distância ou densidade para detectar anomalias, o que ajuda a reduzir o custo computacional.

Segundo Liu, Ming e Zhou (2008) o *iForest* a) identifica anomalias como pontos com comprimentos de caminho mais curtos; e b) possui várias árvores atuando como 'especialistas' para localizar diferentes anomalias. Neste método há apenas duas variáveis (parâmetros para o modelo): o número de árvores a serem construídas e o tamanho da subamostragem. O modelo isola anomalias em vez de perfis de instâncias normais, ele constrói um conjunto de árvores isoladas (denominadas iTrees) para um conjunto de dados determinado. O primeiro estágio constrói árvores isoladas usando subamostras do conjunto de treinamento. O segundo estágio usa as árvores isoladas para executar as instâncias de teste e obter a pontuação de anomalia para cada

instância. A árvore isolada é uma árvore binária, onde cada nó da árvore tem exatamente zero ou dois nós filhos. A figura 7 exemplifica esse processo.

Figura 7- Construção *iForest* aprendida para um conjunto de dados



Fonte: Donghwa Kim (2018)

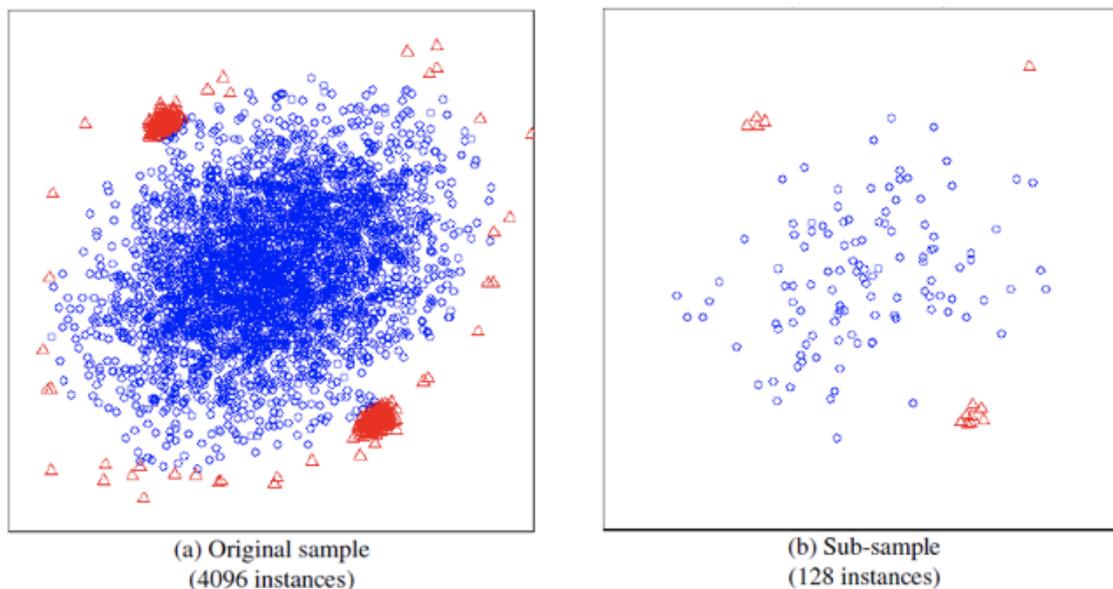
O *iForest* funciona em duas fases: fase de treino e fase de testes. Durante a primeira fase, o algoritmo constrói um conjunto de árvores isoladas, conhecido como *iTrees*, formulando um *iForest*. Cada árvore é construída seguindo o seguinte processo: (i) Desenhe uma amostra uniforme dos dados (ii) Selecione um ponto de divisão  $p$  aleatoriamente e um atributo  $q$  (iii) Divida a amostra de dados (iv) Repita as etapas 2 e 3, até que uma certa altura de árvore predefinida seja alcançada, todos os pontos de dados na amostra têm o mesmo valor ou apenas um ponto de dados permanece na amostra para dividir. (REGAYA, FADLI e AMIRA, 2021)

O segundo estágio passa cada ponto de dados por cada árvore aleatória construída e calcula o valor discrepante correspondente  $s(x)$  que varia entre 0 e 1. Os pontos de dados com valores inferiores a 0,5 são rotulados normalmente e recebem um valor de 1. Caso contrário, os pontos de dados são considerados potencialmente anômalos e os pontos com valores próximas a 1 são sinalizados como anomalias atribuindo um valor de -1.

As árvores isoladas são capazes de lidar com os efeitos de inundação e mascaramento. A primeira é referente à identificação incorreta de uma instância normal como anômala, isso devido à proximidade de instancias anômalas, nesse caso o número de partições necessárias para separar as anomalias aumenta, dificultando a distinção entre anomalias e instâncias normais. O outro efeito diz respeito a presença de muitas anomalias que escondem a sua própria presença, resultando em

uma camuflagem. Se o cluster de anomalias for grande e denso, também aumenta o número de partições para separar cada anomalia. A figura 8 apresenta um exemplo.

Figura 8- Efeitos de inundação e mascaramento em um conjunto de dados



Fonte: Tony Liu; Ming Ting; Zhou (2008)

Na figura 8 mostra um conjunto de dados original e a outra uma amostra desse conjunto de dados. Os agrupamentos de anomalias são claramente identificáveis na subamostra, uma vez que os clusters de anomalias se tornam menores, o que os torna mais fáceis de identificar.

Devido à propriedade única de árvores isoladas, o *iForest* pode construir um modelo parcial com uma subamostra que atenua os efeitos aleatórios de inundação e cobertura. As razões para tal é que a subamostra controla o tamanho dos dados, o que ajuda o *iForest* a isolar melhor os outliers; e cada árvore de isolamento pode ser especializada porque cada subamostra contém outliers diferentes ou até mesmo nenhum.

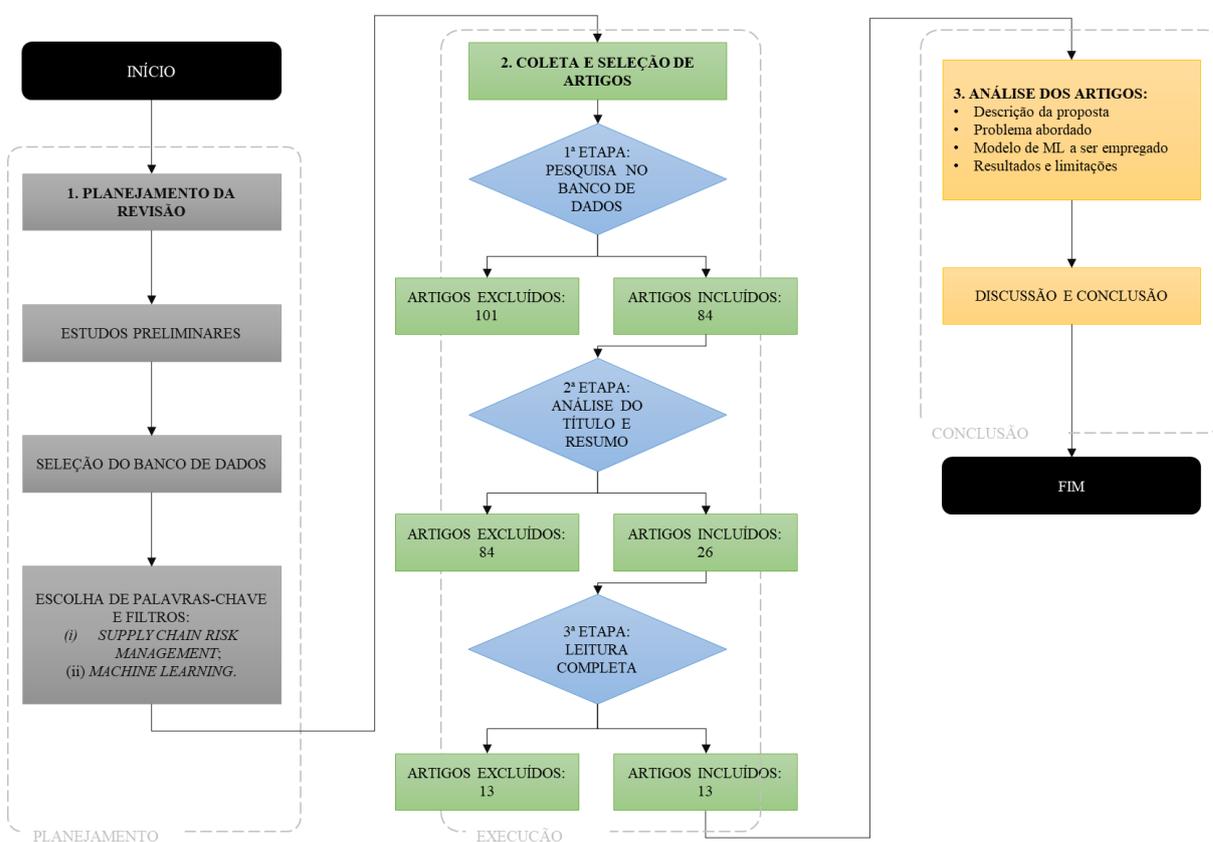
## 2.2 REVISÃO DA LITERATURA

Esta subseção apresenta uma revisão da literatura referente ao uso da aplicação de técnicas de aprendizado de máquina no contexto do GRCS. O banco de dados utilizado é o *Web of Science* da Clarivate, escolhido pela qualidade verificada de seu acervo, que fornece um fator de impacto associado às publicações periódicas.

De fato, tal métrica serve como guia na avaliação e seleção das contribuições dos artigos selecionados, um dos objetivos desse trabalho. Inicialmente, o presente trabalho considerou artigos, revisões e acesso antecipado, escritos na língua inglesa, sendo excluídos artigos publicados em outras línguas. O intervalo de busca das publicações considera os últimos 10 anos (i.e., 2012 e 2022).

O fluxograma 1 traz o fluxograma da metodologia de pesquisa empregada nessa revisão de literatura. Sua estrutura é dividida em três etapas, o planejamento da revisão, a coleta e seleção dos artigos e a análise dos artigos selecionados. Os processos internos são detalhados no decorrer desse trabalho.

Fluxograma 1- Etapas da revisão

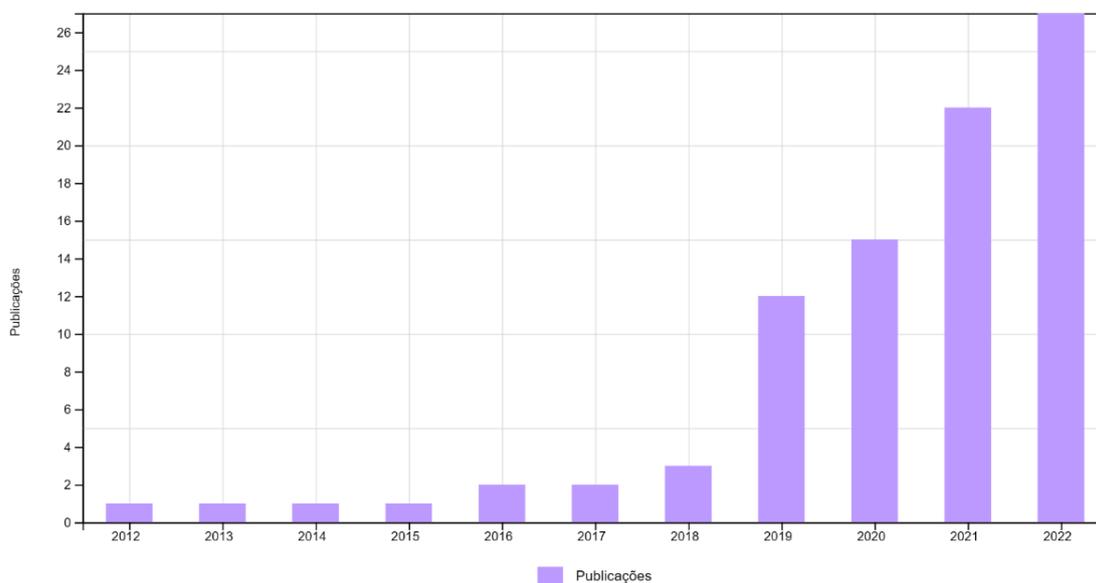


Fonte: O Autor (2023)

As palavras-chave escolhidas para realizar busca no banco de dados foram (i) *'supply chain risk management'* e (ii) *'machine learning'*. A busca resultou em um total de 4.748 referências relacionadas à primeira palavra-chave. Ao combinar a pesquisa com a segunda palavra-chave, *machine learning*, foi encontrada uma amostra de 101 publicações. O gráfico 3 apresenta o número de citações e publicações ao longo do

tempo destes documentos. Através dela é possível perceber uma escalada de interesse sobre esse tema nos últimos anos, principalmente no último quadriênio.

Gráfico 3- Número de citações e publicações ao longo do tempo



Fonte: Web of Science (2023)

Na etapa seguinte realizou-se uma triagem para selecionar publicações cuja abordagens estão diretamente alinhadas aos propósitos desse estudo. Pelo fato desta revisão enfatizar aplicações, 14 artigos caracterizados como revisões bibliográficas foram excluídos da análise, resultando em um total de 87 trabalhos.

### 2.2.1 Classificação e seleção dos estudos

Com o apoio do software VosViewer, realizou-se uma análise das palavras-chave com mais ocorrência dentre das 87 publicações. Através de um mapa de redes, apresentado na Figura 9, pode-se visualizar as palavras-chaves com maior ocorrência. *Machine learning* é a que possui maior ocorrência (36), o que é esperado uma vez que é o foco da revisão. Os links mais fortes desta palavra-chave são com os termos *management*, *supply chain*, *big data analytics*, *risk* e *model*.



como o país com mais documentos (26), seguida pelos Estados Unidos (15) e Inglaterra (9).

Para a etapa de seleção, foram lidos os títulos e resumos dos artigos para verificação de quais trabalhos se enquadrariam no escopo da pesquisa. Permaneceram 26 documentos no total. Uma última triagem foi feita baseado na leitura completa dos artigos, gerando uma nova seleção de apenas 13 documentos, os quais são inteiramente associados com o objetivo deste trabalho. Os fatores que influenciaram a inclusão ou exclusão dos artigos analisados nessa revisão foram a afinidade com a proposta desse trabalho e a disponibilidade para leitura.

Tabela 1- Relação dos artigos selecionados

REFERÊNCIA	TÍTULO DA PUBLICAÇÃO	DESCRIÇÃO DO ESTUDO	MODELO APLICADO
Janjua, NK; Nawaz, F; Prior, DD	A fuzzy supply chain risk assessment approach using real-time disruption event data from Twitter	O estudo desenvolveu uma metodologia para identificar eventos de interrupção da cadeia de suprimentos usando feeds do Twitter em tempo real.	Conditional Random Field (CRF); Fuzzy
Melancon, GG; Grangier, P; Prescott-Gagnon, E; Sabourin, E; Rousseau, LM	A Machine Learning-Based System for Predicting Service-Level Failures in Supply Chains	Propôs um sistema de aprendizado de máquina que prevê falhas de nível de serviço com algumas semanas de antecedência e alerta os planejadores.	Gradient-boosted decision trees (GBDTs)
Lau, H; Tsang, YP; Nakandala, D; Lee, CKM	Risk quantification in cold chain management: a federated learning-enabled multi-criteria decision-making methodology	É proposto um sistema de avaliação de risco multicritério habilitado para aprendizagem federada.	FL-enabled multi-criteria risk evaluation system (FMRES)
Jomthanachai, S; Wong, WP; Lim, CP	An Application of Data Envelopment Analysis and Machine Learning Approach to Risk Management	A motivação para este estudo é que a combinação das abordagens DEA e ML oferece uma escolha flexível e realista no gerenciamento de riscos.	Data Envelopment Analysis (DEA); Failure mode and effect analysis (FMEA); Artificial neural network (ANN)
Salamai, AA; El-kenawy, EM; Abdelhameed, I	Dynamic Voting Classifier for Risk Identification in Supply Chain 4.0	Este artigo propôs um classificador de votação para identificar os riscos operacionais na cadeia de suprimentos 4.0	Sine Cosine Dynamic Group (SCDG)
Han, CL; Zhang, Q	Optimization of supply chain efficiency management based on machine learning and neural network	Este estudo constrói um modelo de gerenciamento de riscos da cadeia de suprimentos baseado em aprendizado e rede neural.	Back-propagation (BP) neural network
Mahato, PK; Narayan, A	Robust Supply Chains with Gradient Boosted Trees	Desenvolveu um pipeline de aprendizado de máquina com árvores de decisão impulsionadas por gradiente para mitigar falhas de nível de serviço nas cadeias de suprimentos.	Gradient Boosted Decision Trees (GBDT); LightGBM
Baryannis, G; Dani, S; Antoniou, G	Predicting supply chain risks using machine learning: The trade-off between performance and interpretability	Este trabalho primeiro elaborou uma estrutura para prever os riscos da cadeia de suprimentos, depois explorou o trade-off entre desempenho preditivo e interpretabilidade.	Support-vector machine (SVM); Decision tree
Cavalcante, IM; Frazzon, EM; Forcellini, FA; Ivanov, D	A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing	Elaborou uma técnica híbrida, combinando simulação e aprendizado de máquina para aplicações de suporte à tomada de decisão baseada em dados na seleção de fornecedores resilientes.	k-nearest neighbors (k-NN); Logistic Regression (LR)
Brintrup, A; Pak, J; Ratiney, D; Pearce, T; Wichmann, P; Woodall, P; McFarlane, D	Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing	Neste artigo, discutiu-se a aplicação da análise de dados na previsão de interrupções na cadeia de suprimentos de primeiro nível usando dados históricos.	Random Forest (RF)

Hassan, AP	Enhancing Supply Chain Risk Management by Applying Machine Learning to Identify Risks	Esta investigação utiliza aprendizado de máquina para encontrar riscos em documentos textuais.	Natural Language Processing (NLP)
Lee, CH; Yang, HC; Wei, YC; Hsu, WK	Enabling Blockchain Based SCM Systems with a Real Time Event Monitoring Function for Preemptive Risk Management	O trabalho combinou um método de detecção de eventos em tempo real usando dados coletados do Twitter e tecnologia blockchain para monitoramento de eventos para melhorar a visibilidade do sistema da cadeia de suprimentos e tomar medidas preventivas para evitar riscos.	Burst (Burst deTecton)
Athaudage, GNP; Perera, HN; Sugathadasa, PTRS; De Silva, MM; Herath, OK	Modelling the impact of disease outbreaks on the international crude oil supply chain using Random Forest regression	Explorar os fatores influentes que afetam o COSC internacional em termos de consumo, produção e preço. E desenvolver um modelo para prever o preço internacional do petróleo bruto durante surtos de doenças.	Random Forest (RF)

Fonte: O Autor (2023)

A aplicação de modelos de *machine learning* em problemas de GRCS é um assunto pouco explorado, todavia nos últimos anos recebeu maior atenção dos pesquisadores. Este tópico buscou revisar a literatura para entender como os modelos de ML estão sendo aplicados e em quais contextos. Uma observação importante se faz necessária, não é competência desse estudo julgar a qualidade da contribuição científica dos estudos incluído ou dos excluídos, apenas verificar as aplicações e resultados de ML no contexto de SCRM, resumindo suas propostas e métodos aplicados, servindo assim de norte para estudos futuros.

A utilização de apenas um banco de dados serviu como um filtro que qualificasse ainda mais esse estudo, todavia, vale ressaltar que as limitações de quantidade de artigos disponíveis podem ter impedido este trabalho de observar estudos que contenham propostas promissoras. Há também o fato de que existem trabalhos acadêmicos não disponibilizados ou publicados, desde modo, esses exemplos são considerados fatores delimitadores aos objetivos deste estudo.

Aqui, na maior parte dos estudos foram utilizados algoritmos de classificação, aplicados principalmente para detectar possíveis riscos a CS. Alguns artigos propuseram inovações, como por exemplo Salamai, El-Kenawy e Abdelhameed (2021), que propôs um classificador de votação fundamentado em um algoritmo de otimização SCDG. Algo comum foi a não disponibilização do banco de dados utilizado, como é o caso do estudo de Brintrup *et al.* (2020), que utilizou o ERP da empresa estudada, onde habitualmente existe questões acerca do sigilo destes dados.

### 3 ESTUDO DE CASO

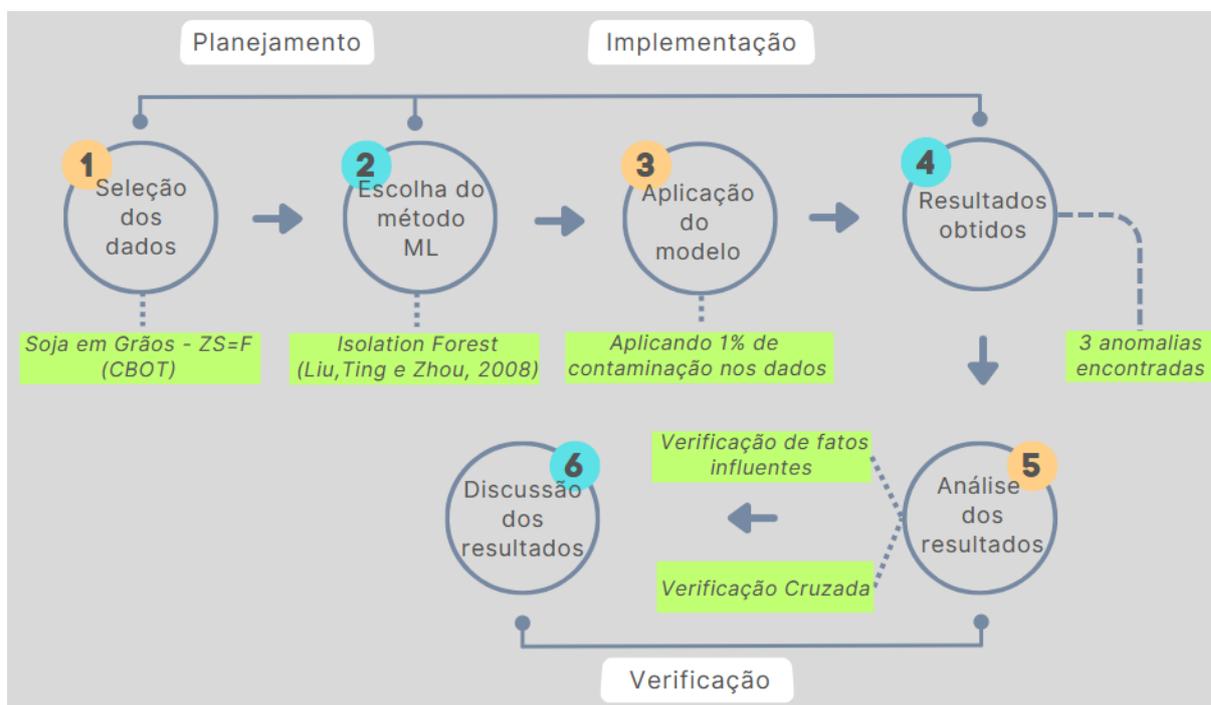
#### 3.1 PROCEDIMENTOS METODOLÓGICOS

Nesta seção encontra-se a caracterização do estudo, origem dos dados utilizados, bem como a síntese das etapas que compõem o presente trabalho.

##### 3.1.1 Caracterização da pesquisa

Este estudo distribuiu-se em 6 etapas sequenciais. A primeira etapa diz respeito a seleção do banco de dados, em seguida a escolha do modelo de detecção de anomalias, a terceira etapa é a aplicação do modelo escolhido. Na quarta etapa é apresentado os resultados obtidos na implementação do modelo. A quinta e sexta etapa são organizadas os processos de análise e discussão dos resultados. abaixo é mostrado o diagrama da metodologia desse estudo.

Diagrama 2- Metodologia do estudo



Fonte: O Autor (2023)

Ressalta-se que na etapa de análise, foi feita uma checagem da relação das anomalias encontradas e as prováveis causas. É importante explicar que a análise da causalidade das anomalias parte da interpretação de acontecimentos relevantes dentro do contexto da soja em escala mundial.

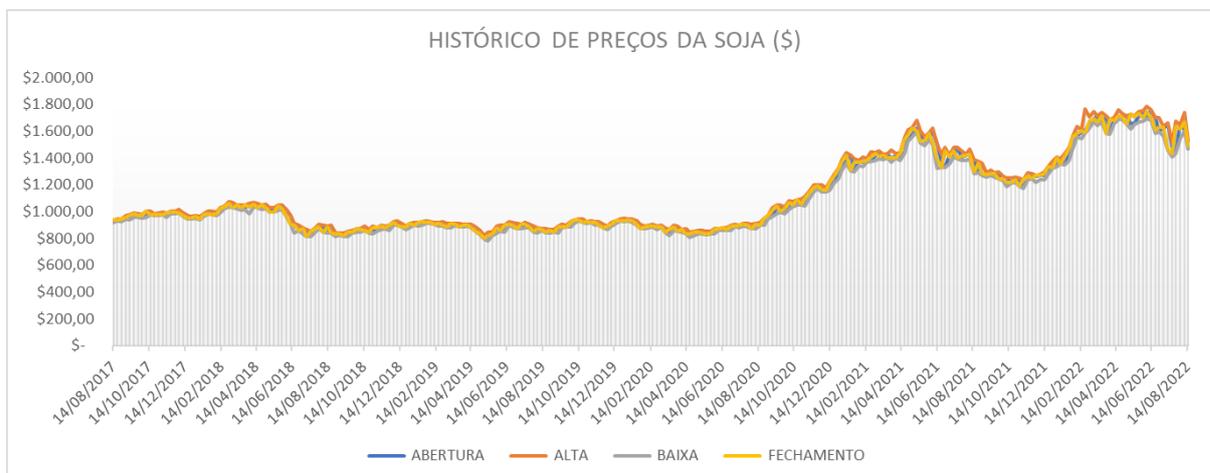
A premissa desse estudo é identificar eventos de interrupção na cadeia produtiva da soja. Para alimentar o algoritmo de aprendizado de máquinas, foram utilizados dados do mercado financeiro. A proposta é aplicar um modelo de aprendizado não supervisionado para que o algoritmo verifique anomalias e encontre padrões a fim de que possa prever possíveis riscos futuros. De forma a fornecer insights em tempo hábil para tomada de decisões dentro da cadeia do *commodity*. Essa definição foi dada pelo entendimento que as marcações de mercado reagem a vários fatores, principalmente em situações em que existe a possibilidade de risco é relevante.

### 3.1.2 Origem e descrição dos dados

Os dados utilizados são o histórico dos últimos 5 anos da marcação de preço do *commodity* da soja em grãos, cujo ticket é  $ZS = F$ , vinculado ao Chicago Board of Trade (CBOT). Os valores foram extraídos do Yahoo Finance, um site amplamente utilizado por investidores e estudiosos do ramo econômico. O histórico de preços foi intervalado semanalmente, pois eventos capazes de impactar o valor de um *commodity* tão importante não influenciaria o fechamento de preços de alguns dias dentro de uma semana, isso significa que eventos disruptivos não acabam num mesmo dia. Os respectivos fechamentos de preços, estão entre os anos de 2017 e 2022.

A base de dados possui 262 semanas, datadas de 14 de agosto de 2017 a 15 de agosto de 2022. Cada semana contém as informações sobre abertura, altas, baixas e fechamento de preços, bem como o volume de ordens emitidas. A precificação do histórico está em dólar. O gráfico abaixo apresenta a marcação de mercado da Soja.

Gráfico 4- Histórico de preços da Soja em dólares



Fonte: O Autor (2023)

Na primeira semana observada, o fechamento foi de \$937,50, e na última semana o fechamento foi \$1.494,00, crescimento do valor do *commodity* de mais de 50% do valor inicial dentro do período analisado. O maior fechamento registrado foi de \$1.745,50 em 06/06/2022, o menor valor de fechamento se deu em 06/05/2019, marcando \$797,00.

### 3.2 APLICAÇÃO DO MÉTODO *ISOLATION FOREST*

Os dados são uma série temporal onde utilizamos apenas os preços de fechamento no primeiro momento para simplificação, observando que o volume também é importante, mas seja ele positivo ou não, será refletido no preço de fechamento. A título de exemplo, uma grande quantidade de compras fará com que o preço seja elevado, logo o volume normalmente refletirá no fechamento da semana.

A precificação de uma *commodity* possui comportamento diferente de ações de empresas. Por não ter tantas especulações, não é frequente observar oscilações bruscas dos preços. A expectativa era de que eventos, notícias, ou informações sobre safras e cadeia produtiva da soja por exemplo, pudessem influenciar diretamente o preço da *commodity*.

A decisão por aprendizado de máquina não supervisionado se deu pelo perfil de detecção de anomalias, uma vez que o aprendizado de máquina supervisionado requer treinamento das saídas para aprender com os dados presentes, e na detecção

de anomalias nós não compreendemos como e de qual forma ocorrem as anomalias. Neste caso não foi interessante optar por aprendizado supervisionado, ressaltando que existem modelos de detecção de anomalias que utilizam essa abordagem. O método escolhido foi a *Isolation Forest*, um método inovador e por ter sido desenvolvido especificamente para detecção de anomalias. O principal motivo da escolha do método se deve ao fato de ele requer um menor custo computacional. Empregou-se a implementação do *iForest* proposta por Ambonati (2017), cuja problemática foi envolvimento da detecção de falhas na medição de temperatura.

A amostra foi a cotação da soja contendo 262 preços de fechamento. O *iForest* exige que seja definido o parâmetro *contamination*. Este parâmetro tradicionalmente pode ser definido por um limiar chamado de *outliers\_fraction*, que é utilizado para determinar a fração de pontos de dados que serão considerados como anomalias, a biblioteca Scikit-learn, Pedregosa et al. (2011), também dispõe de uma função que calcula automaticamente o parâmetro. Para a implementação desse estudo assumiu-se o valor de 0.01 para o parâmetro *contamination*. Dois fatores embasam essa decisão: a) a variação do preço da soja apresenta um comportamento relativamente estável e previsível ao longo do tempo; b) o estudo busca detectar anomalias que sejam derivadas de eventos com capacidade disruptiva dentro da cadeia da soja, logo, um valor tão pequeno de contaminação serve como filtro para que o modelo detectasse as principais anomalias presentes no banco de dados. Abaixo, a tabela 2 demonstra o treinamento do modelo:

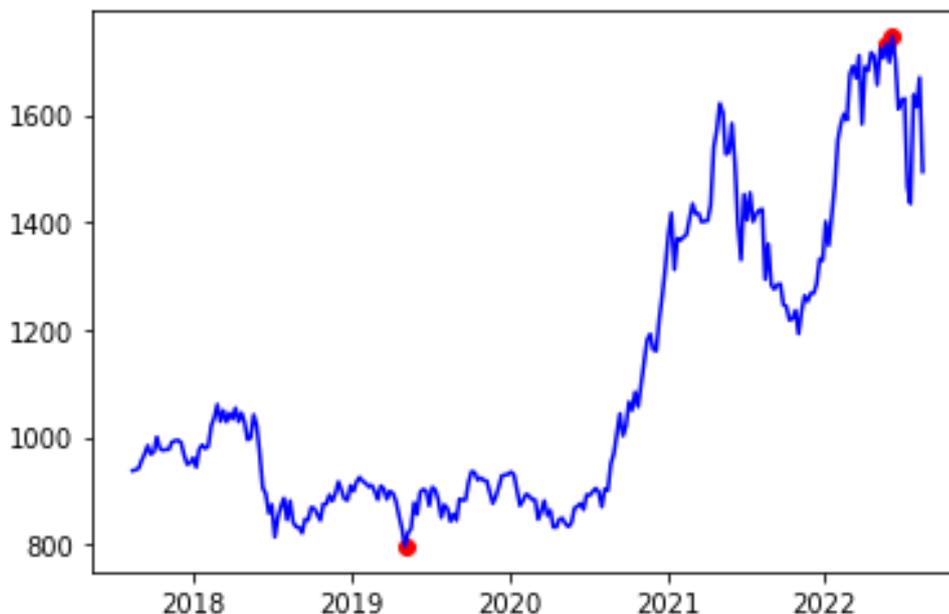
Tabela 2- Etapas de treinamento *iForest*

- |   |   |
|---|---|
| 1 | : Carregue o conjunto de dados do arquivo CSV (df)                      |
| 2 | : Defina a fração de anomalias esperada na amostra (0.1)                |
| 3 | : Separe as colunas dos dados de entrada das colunas das classes (X, y) |
| 4 | : Crie uma instância do algoritmo Isolation Forest (model)              |
| 5 | : Treine o modelo com os dados de entrada (model.fit(X))                |
| 6 | : Realize a detecção de anomalias nos dados de entrada (y_pred)         |
| 7 | : Adicione uma nova coluna ao banco (df['anomaly25'])                   |
| 8 | : Imprima as previsões de anomalia                                      |
| 9 | : Fim   |

Fonte: O Autor (2023)

Após rodar o modelo foram detectadas 3 anomalias dentro desse intervalo. As anomalias estavam datadas nos dias 6 de maio de 2019, 23 de maio de 2022 e 06 de julho de 2022. O gráfico 5 apresenta as anomalias detectadas pelo modelo.

Gráfico 5- Anomalias detectadas pelo *iForest*



Fonte: O Autor (2023)

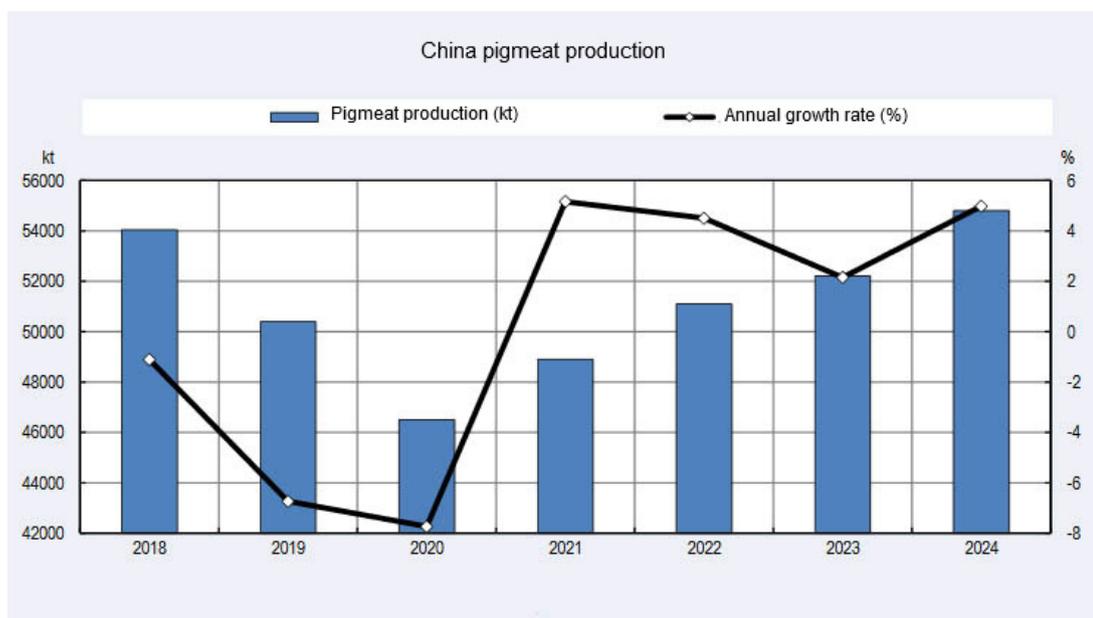
Detectadas possíveis anomalias inicia-se o processo de verificação. Esse processo levou em consideração notícias ou informações capazes de impactar o preço da soja. Filtramos a busca para os países que mais exportam soja no mundo, como Brasil, EUA, China, Argentina, Ucrânia, Rússia, dentre outros importantes. Ressalta-se que seria um erro apenas coletar notícias dos dias específicos das possíveis anomalias, tornando pertinente a coleta de informações também dos dias anteriores, os quais dentro da semana teriam capacidade de influenciar no fechamento daquela respectiva semana.

Iniciando pela anomalia mais antiga da amostra, 2019 foi marcado pela guerra de sanções entre Estados Unidos e China, o que provocou desconforto nos mercados globais. Hirtzer e Mario (2019) explicam que os futuros da soja e do milho caíram após tuítes de Donald Trump que ameaçava uma escalada da guerra comercial EUA-China, frustrando produtores americanos que esperavam uma resolução rápida. A expectativa era de um fim rápido para as tarifas agrícolas da China, o que ocasionou em excesso de estoques nos mercados agrícolas. A Associação Brasileira das

Indústrias de Óleos Vegetais (Abiove) elevou a estimativa de produção de soja do Brasil, entretanto acrescentou que estimava queda nas exportações, Gomes (2019) elencou que como reflexo da maior produção e embarques mais fracos, a Abiove elevou sua estimativa de estoques finais.

A informação que mais afetou o mercado global foi o surto de peste suína africana na China. O gráfico 6 demonstra o que ocorreu com a produção de carne suína em 2019.

Gráfico 6- Produção de carne suína da China



Fonte: pig333.com

“A crise sem precedentes na suinocultura chinesa, provocada pela ocorrência da Peste Suína Africana, está tumultuando o mercado mundial de proteína animal. Com mais de 1 milhão de animais já sacrificados e abatidos e os outros 150 milhões a caminho do mesmo destino, a produção pode cair até 30% no país até 2020, segundo estimativas do Rabobank (ARANHA, 2019). Essa notícia afetou os futuros da soja porque a China é a maior importadora da *commodity* no mundo, uma das principais finalidades é a utilização como alimentos na criação dos porcos.

A segunda data a qual ocorreu a detecção de anomalia é 23/05/2022. As principais notícias são relacionadas as articulações geopolíticas, com é apresentado na Tabela 3, que traz essas informações resumidas.

Tabela 3- Principais notícias referentes a segunda anomalia

**Biden iniciará novo bloco econômico da Ásia-Pacífico com uma dúzia de aliados.**

O objetivo é um ressurgimento da influência dos EUA na região, especialmente em relação à China, após a saída do presidente Donald J. Trump da Parceria Transpacífico.

<https://www.whitehouse.gov/briefing-room/statements-releases/2022/05/23/statement-on-indo-pacific-economic-framework-for-prosperity/>

**O presidente dos EUA, Joe Biden, diz pela primeira vez que estaria disposto a usar a força para defender Taiwan contra uma invasão chinesa.**

As perspectivas econômicas 'escureceram', alertam líderes empresariais e governamentais em Davos

<https://www.reuters.com/markets/europe/economic-storm-looming-business-government-leaders-warn-davos-2022-05-23/>

**Sinais de alerta nas perspectivas econômicas da China à medida que o COVID-19 se espalha.**

Os sinais de desaceleração econômica em Xangai e Guangdong são especialmente preocupantes, considerando o papel descomunal que desempenham na economia geral da China.

<https://thediplomat.com/2022/05/warning-signs-in-chinas-economic-outlook-as-covid-19-spreads/>

**FMI adverte contra a fragmentação econômica da guerra na Ucrânia.**

Desfazer a integração tornaria o mundo mais pobre e mais perigoso. As nações devem reduzir as barreiras comerciais, diversificar as cadeias de abastecimento

<https://www.bloomberg.com/news/articles/2022-05-23/imf-warns-against-global-economic-fragmentation-from-ukraine-war?leadSource=verify%20wall>

**Temores de estagnação aumentam, com 77% dos investidores prevendo tempestade econômica**

<https://www.foxbusiness.com/economy/stagflation-fears-rising-majority-investors-forecasting-economic-storm>

Fonte: O Autor (2023)

O contexto dessa segunda data é marcado por um cenário mundial mais delicado. 2022 iniciou com o conflito bélico entre Ucrânia e Rússia, que afetou diretamente o mercado de diversas *commodities*, principalmente o trigo. Derivado dos anos anteriores, EUA e China com suas tensões comerciais quase entraram em conflito bélico quando a China ameaçou invadir Taiwan. Além disso, os EUA propuseram um novo bloco econômico para frear a influência da China na Ásia. O resultado foram movimentações econômicas que tem fragmentado o mercado

mundial, gerando incertezas e temores. Por fim também houve a preocupação de outra onda do vírus da covid-19, o qual obrigou a China a decretar lockdown em regiões específicas.

Os fatos supracitados afetaram as economias globais principalmente por estarem relacionadas com as maiores nações do planeta. É provável que não só as marcações de preços da soja tenham sido influenciadas por essas notícias devido sua magnitude.

A última anomalia detectada ocorreu 2 semanas após a anterior, dia 06 de junho de 2022, por consequência, o cenário mundial é o mesmo, contendo acontecimentos derivados. Abaixo estão listadas as principais notícias:

- “Deixada de fora do acordo Indo-Pacífico, a China avança em direção ao maior acordo comercial do mundo” (TAN, 2022);
- “Rússia pede ação coordenada do BRICS contra riscos econômicos globais” (REUTERS, 2022);
- “Biden oferecerá novo marco econômico para a América Latina” (NICHOLS, 2022);
- “USDA envia US\$ 700 milhões para instalações de etanol e biodiesel afetadas pela pandemia” (VONDRACEK, 2022).

As primeiras notícias são articulações políticas com enfoque econômico, logo o impacto desses fatos sobre o mercado em geral é difícil de ser mensurado. Todavia o último tópico é divergente pois se refere a políticas públicas de fomento a produção de biocombustíveis. Como descrito na própria matéria, tal medida afetará positivamente os plantios de milho e soja.

### 3.3 VERIFICAÇÃO CRUZADA

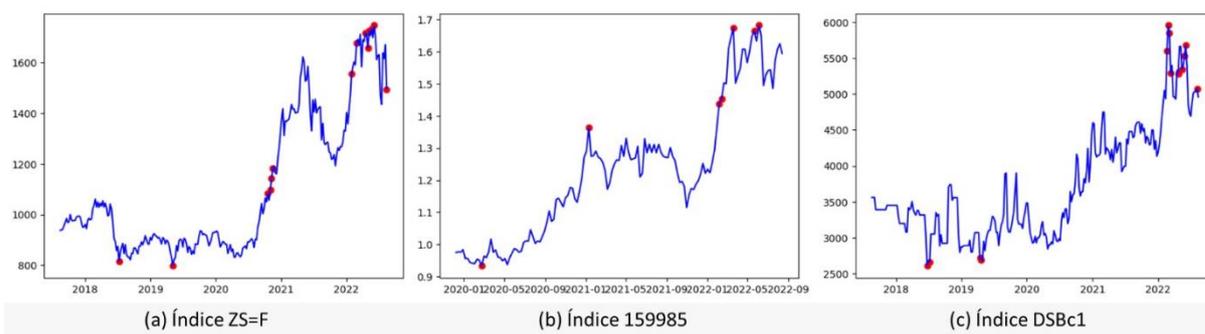
Como análise adicional, adotou-se uma segunda validação das anomalias detectadas. Nessa etapa a implementação do *iForest* foi realizada em outro dois bancos de dados relativos as cotações da soja em diferentes lugares. O objetivo é investigar se o modelo iria detectar anomalias nos mesmos períodos (ou próximos) da primeira implementação.

Utilizamos os índices No 2 Soybeans Futures (DSBc1) - Dalian *Commodity* Exchange (DCE) e ChinaAMC Feed&Soybean Meal Futures (159985) – Shenzhen, ambos pertencendo as bolsas chinesas. Os valores dos índices foram obtidos através do portal Investing.com. Sobre o índice DSBc1, He e Wang (2011) apontam que este tem como objetivo conectar a China e mercados futuros internacionais de soja e aumentar o impacto das demandas da China no mercado internacional. Devido à volatilidade dos índices e do sentimento dos investidores chineses, os quais podem diferir do sentimento dos investidores do índice americano (ZS=F), houve uma ponderação sobre o parâmetro de contaminação, alterando para 5% dos dados.

O ChinaAMC Feed&Soybean Meal Futures (159985) é o índice mais recente, sendo seu início de marcação a partir de 2020. Apenas DSBc1 não é precificado em dólar americano, todavia as transações internacionais da soja são negociadas em dólar, assim é esperado que o preço da *commodity* seja afetado no mercado interno pelos preços internacionais.

Na implementação do índice ZS=F, 14 anomalias foram detectadas; para o índice DSBc1, 13 anomalias detectadas; finalmente, para o 159985, foram encontradas 7 anomalias. O gráfico 7 apresenta o histórico dos índices com suas respectivas anomalias detectadas pelo algoritmo proposto.

Gráfico 7- Ocorrência de anomalias na verificação cruzada.



Fonte: O Autor (2023)

Ao fazer um cruzamento de informações (detecções) entre os três bancos de dados, foram observadas a ocorrência de anomalias nos das duas novas bases em datas similares as da primeira implementação. A Tabela 4 apresenta as datas das possíveis anomalias, correlacionando a ocorrência em cada uma das três bases de dados analisadas.

Tabela 4- Quadro de anomalias detectadas

DSBc1	ZS=F	159985
05/06/2022	15/08/2022	05/06/2022
29/05/2022	06/06/2022	22/05/2022
08/05/2022	23/05/2022	20/03/2022
01/05/2022	09/05/2022	13/02/2022
24/04/2022	02/05/2022	06/02/2022
13/03/2022	18/04/2022	10/01/2021
06/03/2022	28/02/2022	23/02/2020
27/02/2022	31/01/2022	
20/02/2022	16/11/2020	
21/04/2019	09/11/2020	
14/04/2019	02/11/2020	
08/07/2018	19/10/2020	
24/06/2018	06/05/2019	
	09/07/2018	

— Intervalo até 8 dias  
 - - - Intervalo entre 8 e 15 dias  
 ■ Correlação entre anomalias  
 ■ Anomalias 1ª implementação

Fonte: O Autor (2023)

Dentre essas anomalias detectadas na primeira implementação (margeadas em azul), apenas a datada em 06/05/2019 não possui uma anomalia de data similar nas outras aplicações do modelo, entretanto existe uma anomalia relativamente próxima dentro do índice DSBc1 na semana 21/04/2019. É possível que essas anomalias possam refletir ao mesmo evento, pois o DSBc1 refere-se aos futuros da soja na China, e a informação mais relevante no período da anomalia foi a queda das importações de soja devido ao abate de milhões de porcos por conta de contaminação. Naturalmente o mercado interno sobre o impacto antes do mercado externo, dito isso a anomalia nos futuros chineses acontece pouco antes da detectada nos futuros norte-americanos.

A validação através dos dados de marcação da soja de outros mercados traz mais robustez aos resultados obtidos pelo modelo aplicado. Assim, foi avaliado a generalização das detecções do modelo em diferentes contextos e regiões. A utilização de dados de marcação de soja de outros mercados ajuda a mitigar o viés de um único conjunto de dados, fornecendo uma visão mais abrangente e confiável.

Essa validação cruzada permite identificar tendências e padrões que podem ser relevantes para o aprimoramento do modelo, tornando-o mais preciso e adaptável.

### 3.4 DISCUSSÕES DOS RESULTADOS

Diante dos resultados e suas respectivas contextualizações, é factível afirmar que o modelo tem eficácia em suas detecções de anomalias. Os três dados anômalos possuíam fatores externos capazes de manipular os fechamentos do preço da soja. Na crise de proteína suína, os estoques aumentariam, de forma que a demanda estava em projeção de queda, com isso, os preços também caíram. Nos outros dois casos, a escassez poderia ser o problema, invertendo assim a lógica do comportamento, oferta limitada e demanda crescente, resultou em preços atingindo picos. Na verificação cruzada a incorporação desses dados de marcação adicionais fortalece a credibilidade e a utilidade do modelo aplicado na previsão da produção de soja.

O modelo permite alterar sua sensibilidade mudando os valores do parâmetro *contamination*, assim a depender das modificações realizadas, ele iria detectar mais ou menos das possíveis anomalias. Tal análise de sensibilidade foi descartada em primeiro momento por não ser a problemática principal desse estudo.

Uma observação interessante a ser apontada é o não surgimento de anomalias nas datas das fases iniciais da pandemia do Covid-19. Vale ressaltar que a dinâmica dos preços da soja mudou após a metade de 2020, subindo suas cotações. É possível que tal mudança no comportamento dos preços seja decorrente da pandemia do Covid-19. Esta discussão é relevante dado que pode ocasionar em mais explicações de como a pandemia afetou o agronegócio e em qual escala, e se isso interferiu a precificação da soja.

## 4 CONCLUSÃO

A aplicação de modelos de aprendizado de máquina a problemas de GRCS é um tema ainda pouco explorado, mas que tem recebido crescente atenção dos pesquisadores nos últimos anos. De fato, percebe-se que dentro dessa temática não existiam muitos estudos com aplicações de ML para detecção de anomalias. O presente estudo conseguiu resultados promissores sobre o uso de aprendizado não supervisionado, detectando dados anômalos nas precificações da soja.

O *iForest* apresentou bons resultados após a implementação. As anomalias encontradas possuem conexão com fatores externos, o que corrobora com a eficácia do modelo. A verificação cruzada trouxe maior robustez aos resultados encontrados, pois índices diferentes apresentaram resultados bastantes similares.

Ainda assim, pode-se destacar limitações relacionadas ao estudo. Por exemplo, apesar de existir uma conexão entre as anomalias e os fatores externos para os dados encontrados, esse estudo não pode afirmar que foram esses os fatores específicos que de fato influenciaram a precificação da soja e produziram distorções, pois existem muitas variáveis que poderiam ser veladas em conta. O preenchimento desta lacuna pode servir de inspiração para novas produções científicas que busquem consolidar ainda mais utilização das marcações de mercado como input para detecção de eventos disruptivos.

Como pesquisas futuras, será buscada a identificação e exploração de outros bancos de dados de rápida sensibilidade (resposta) a um evento disruptivo, o que resultaria numa aceleração do tempo de detecção deste evento em potencial cuja capacidade pode afetar uma cadeia de suprimentos. Além disso, outros modelos de aprendizado como Local Outlier Factor (LOF) e One-Class SVM podem ser analisados e comparados na detecção de anomalias nas series de soja. Bem como, pode-se investigar a verificação do nível médio da série da soja, em virtude da sua mudança de padrão após o período pandêmico.

## REFERÊNCIAS

- ABOUTORAB, H. *et al.* A reinforcement learning-based framework for disruption risk identification in supply chains. **Future Generation Computer Systems**, 2022. v. 126, p. 110–122. Disponível em: <<https://doi.org/10.1016/j.future.2021.08.004>>.
- AFIOUNI-MONLA, R. Organizational learning in the rise of machine learning. **40th International Conference on Information Systems, ICIS 2019**, 2019.
- AGUIAR, E. C.; TORTATO, U.; GONÇALVES, M. A. Riscos e gestão de riscos em cadeias de suprimentos: Uma síntese da literatura. **Espacios**, 2012. v. 33, n. 8, p. 3.
- AMBONATI, V. Unsupervised Anomaly Detection. **kaggle**, 2017. Disponível em: <<https://www.kaggle.com/code/victorambonati/unsupervised-anomaly-detection/notebook>>. Acesso em: 10 out. 2022.
- APROSOJA BRASIL. SOJA BRASILEIRA: HISTÓRIA E PERSPECTIVAS. **Associação Brasileira dos Produtores de Soja**, 2020. Disponível em: <<https://aprosojabrasil.com.br/comunicacao/blog/2020/08/27/brazilian-soybean-exports/>>. Acesso em: 23 ago. 2022.
- ARANHA, C. Peste suína na China afeta mercados de proteínas e soja no mundo. **Globo Rural**, 2019. Disponível em: <<https://globorural.globo.com/Noticias/Criacao/Suinos/noticia/2019/05/febre-suina-na-china-afeta-mercados-de-proteinas-e-soja-no-mundo.html>>. Acesso em: 13 jan. 2023.
- ASBJØRNSLETT, B. E. Assessing the Vulnerability of Supply Chains. 2009. p. 15–33. Disponível em: <[http://link.springer.com/10.1007/978-0-387-79934-6\\_2](http://link.springer.com/10.1007/978-0-387-79934-6_2)>.
- BALLOU, R. H. **Gerenciamento da Cadeia de Suprimentos/Logística Empresarial**. 5. ed. Porto Alegre: Bookman, 2006.
- BARYANNIS, G. *et al.* Supply chain risk management and artificial intelligence: state of the art and future research directions. **International Journal of Production Research**, 2019. v. 57, n. 7, p. 2179–2202. Disponível em: <<https://doi.org/10.1080/00207543.2018.1530476>>.
- BEHZADI, G. *et al.* Agribusiness supply chain risk management: A review of quantitative decision models. **Omega (United Kingdom)**, 2018. v. 79, p. 21–42.

Disponível em: <<http://dx.doi.org/10.1016/j.omega.2017.07.005>>.

BERTAGLIA, P. R. **Logística e Gerenciamento da Cadeia de Abastecimento**. São Paulo: Saraiva, 2009.

BHATTAD, P. B.; JAIN, V. Artificial Intelligence in Modern Medicine – The Evolving Necessity of the Present and Role in Transforming the Future of Medical Care. **Cureus**, 9 maio. 2020. Disponível em: <<https://www.cureus.com/articles/30942-artificial-intelligence-in-modern-medicine---the-evolving-necessity-of-the-present-and-role-in-transforming-the-future-of-medical-care>>.

BOWERSOX, D. J. *et al.* **Gestão Logística da Cadeia de Suprimentos**. 4. ed. Porto Alegre: AMGH Editora Ltda., 2014.

BRINTRUP, A *et al.* Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing. Univ Cambridge, Dept Engr, Inst Mfg, Cambridge CB3 0FS, England: **INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH**, 2020. v. 58, n. 11, p. 3330–3341.

BU, S. Logistics engineering optimization based on machine learning and artificial intelligence technology. **Journal of Intelligent & Fuzzy Systems**, 2 fev. 2021. v. 40, n. 2, p. 2505–2516. Disponível em: <<https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JIFS-189244>>.

CANAL AGRO. Quais são os principais produtores de grãos do mundo? 2022. Disponível em: <<https://summitagro.estadao.com.br/comercio-exterior/quais-sao-os-principais-produtores-de-graos-do-mundo/>>. Acesso em: 24 ago. 2022.

CANZIAN, F. Guerra e sanções aceleram desglobalização e põem dólar em xeque. 2022. Disponível em: <<https://www1.folha.uol.com.br/mercado/2022/04/guerra-e-sancoes-aceleram-desglobalizacao-e-poem-dolar-em-xeque.shtml>>. Acesso em: 19 abr. 2022.

CHOY, G. *et al.* Current Applications and Future Impact of Machine Learning in Radiology. **Radiology**, ago. 2018. v. 288, n. 2, p. 318–328. Disponível em: <<http://pubs.rsna.org/doi/10.1148/radiol.2018171820>>.

CHRISTOPHER, M. **Logística e gerenciamento da cadeia de suprimentos. Criando redes que agregam valor**. São Paulo: Thomson Learning, 2010.

CIMINI, C. *et al.* Smart Logistics and The Logistics Operator 4.0. **IFAC-PapersOnLine**, 2020. v. 53, n. 2, p. 10615–10620. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S2405896320335837>>.

CRAIGHEAD, C. W. *et al.* The Severity of Supply Chain Disruptions: Design Characteristics and Mitigation Capabilities. **Decision Sciences**, fev. 2007. v. 38, n. 1, p. 131–156. Disponível em: <<https://onlinelibrary.wiley.com/doi/10.1111/j.1540-5915.2007.00151.x>>.

DOS, D.; GOMES, S. Inteligência Artificial : Conceitos e Aplicações. 2010. p. 234–246.

ENHOLM, I. M. *et al.* Artificial Intelligence and Business Value: a Literature Review. **Information Systems Frontiers**, 25 ago. 2021. v. 974, n. August, p. 85–97. Disponível em: <<https://link.springer.com/10.1007/s10796-021-10186-w>>.

FAN, J. *et al.* From Brain Science to Artificial Intelligence. **Engineering**, 2020. v. 6, n. 3, p. 248–252. Disponível em: <<https://doi.org/10.1016/j.eng.2019.11.012>>.

FEARNSIDE, P. M. Soybean cultivation as a threat to the environment in Brazil. **Environmental Conservation**, 10 mar. 2001. v. 28, n. 1, p. 23–38. Disponível em: <[https://www.cambridge.org/core/product/identifier/S0376892901000030/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0376892901000030/type/journal_article)>.

FLEURY, P. F.; WANKE, P.; FIGUEIREDO, K. F. **Logística Empresarial: A Perspectiva Brasileira**. 1. ed. [S.l.]: Atlas, 2000.

GOMES, J. R. ABIOVE ELEVA EM 0,6% ESTIMATIVA DE PRODUÇÃO DE SOJA DO BRASIL. **Reuters**, 2019. Disponível em: <<https://www.reuters.com/>>.

HE, L.; WANG, R. A tale of two markets : Who can represent the soybean futures markets in China ? 2011. v. 5, n. 3, p. 826–832.

HIRAKURI, M. H. TECNOLOGIAS DE PRODUÇÃO DE SOJA. *In*: SEIXAS, C. D. S. *et al.* (Org.). [S.l.]: Embrapa, 2020, p. 348.

HIRTZER, M.; MARIO, P. Trump Trade Tweets Send Grain Markets Diving to 42-Year Low. 2019. Disponível em: <<https://www.bloomberg.com/news/articles/2019-05-06/grains-soy-trump-tariff-threat-to-china-drives-ag-prices-lower#xj4y7vzkg>>.

HO, W. *et al.* Supply chain risk management: a literature review. **International Journal of Production Research**, 18 ago. 2015. v. 53, n. 16, p. 5031–5069. Disponível em:

<<http://www.tandfonline.com/doi/full/10.1080/00207543.2015.1030467>>.

HOWARD, J. Artificial intelligence: Implications for the future of work. **American Journal of Industrial Medicine**, 2019. v. 62, n. 11, p. 917–926.

JANJUA, N. K.; NAWAZ, F.; PRIOR, D. D. A fuzzy supply chain risk assessment approach using real-time disruption event data from Twitter. Edith Cowan Univ, Sch Sci Comp & Secur, Perth, WA, Australia: **ENTERPRISE INFORMATION SYSTEMS**, 2021.

JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. **Science**, 17 jul. 2015. v. 349, n. 6245, p. 255–260. Disponível em: <<https://www.science.org/doi/10.1126/science.aaa8415>>.

KOSASIH, E. E.; BRINTRUP, Alexandra. A machine learning approach for predicting hidden links in supply chain with graph neural networks. **International Journal of Production Research**, 31 jul. 2021. p. 1–14. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/00207543.2021.1956697>>.

MACHLINE, C. Cinco décadas de logística empresarial e administração da cadeia de suprimentos no Brasil. **Revista de Administração de Empresas**, 2011. v. 51, n. 3, p. 227–231.

MOHRI, M.; ROSTAMIZADEH, A.; TALWALKAR, A. **Foundations of Machine Learning**. 2. ed. Cambridge: MIT Press, 2018.

NICHOLS, H. Biden to offer new economic framework for Latin America. **Axios**, 2022. Disponível em: <<https://www.axios.com/2022/06/06/biden-to-offer-new-economic-framework-for-latin-america>>. Acesso em: 28 nov. 2022.

NIMMY, S. F. *et al.* Explainability in supply chain operational risk management: A systematic literature review. **Knowledge-Based Systems**, jan. 2022. v. 235, p. 107587. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0950705121008492>>.

NORRMAN, A.; JANSSON, U. Ericsson's proactive supply chain risk management approach after a serious sub-supplier accident. **International Journal of Physical Distribution & Logistics Management**, 1 jun. 2004. v. 34, n. 5, p. 434–456. Disponível em: <<https://www.emerald.com/insight/content/doi/10.1108/09600030410545463/full/html>>.

>.

NOVAES, A. G. **Logística e gerenciamento da cadeia de distribuição**. Rio de Janeiro: [s.n.], 2001.

OLIVEIRA, U. R. DE *et al.* The ISO 31000 standard in supply chain risk management. **Journal of Cleaner Production**, maio. 2017. v. 151, p. 616–633. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0959652617304894>>.

ORENSTEIN, P.; LADIK, D.; RAINFORD, S. What are the Key Drivers of Future Supply Chains? **Journal of Accounting – Business & Management**, 2016. v. 23.

PEDAMKAR, P. Machine Learning Algorithms. 2019. Disponível em: <<https://www.educba.com/machine-learning-algorithms/>>. Acesso em: 13 set. 2022.

PEDREGOSA; ET AL. Scikit-learn: Machine Learning in Python. **JMLR 12**, 2011. Disponível em: <<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>>. Acesso em: 20 fev. 2023.

RAJESH, R. A grey-layered ANP based decision support model for analyzing strategies of resilience in electronic supply chains. **Engineering Applications of Artificial Intelligence**, 2020. v. 87, n. November 2019, p. 103338. Disponível em: <<https://doi.org/10.1016/j.engappai.2019.103338>>.

REGAYA, Y.; FADLI, F.; AMIRA, A. Point-Denoise: Unsupervised outlier detection for 3D point clouds enhancement. **Multimedia Tools and Applications**, 2021. v. 80, n. 18, p. 28161–28177.

REUTERS. Russia calls for coordinated BRICS action against global economic risks. 2022. Disponível em: <<https://www.reuters.com/markets/europe/russia-calls-coordinated-brics-action-against-global-economic-risks-2022-06-06/>>. Acesso em: 25 nov. 2022.

RIBEIRO, P. C. C.; GOMES, C. F. S. **Gestão da cadeia de suprimentos integrada à tecnologia da informação**. 4. ed. São Paulo: Thomson, 2013.

RITCHIE, H.; ROSER, M. “Forests and Deforestation”. **OurWorldInData.org**, 2021. Disponível em: <<https://ourworldindata.org/forests-and-deforestation>>.

RUSSEL, S. J.; NORVIG, P. **Artificial Intelligence: a modern approach**. 2. ed. New

Jersey: Pearson Education Inc, 2003.

RUSSEL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 3. ed. Harlow, UK: Pearson Education Inc, 2014.

SALAMAI, A. A.; EL-KENAWY, E. M.; ABDELHAMEED, I. Dynamic Voting Classifier for Risk Identification in Supply Chain 4.0. Jazan Univ, Community Coll, Jazan, Saudi Arabia: **CMC-COMPUTERS MATERIALS & CONTINUA**, 2021. v. 69, n. 3, p. 3749–3766.

SCAVARDA, L. F. R.; HAMACHER, S. Evolução da Cadeia de Suprimentos da Indústria Automobilística no Brasil. **RAC**, 2001.

SCHMIDT, R. *et al.* Value creation in connectionist artificial intelligence - a research agenda. **26th Americas Conference on Information Systems, AMCIS 2020**, 2020. p. 0–10.

STOLLER. A importância da cultura da soja. 2021. Disponível em: <<https://www.stoller.com.br/importancia-da-soja-para-a-agricultura-brasileira/>>. Acesso em: 24 ago. 2022.

TADDY, M. The Technological Elements of Artificial Intelligence. **The Economics of Artificial Intelligence**, 2019. p. 61–88.

TALLMAN, J. *et al.* Soybean. **Compendium of Transgenic Crop Plants**. Chichester, UK: John Wiley & Sons, Ltd, 2008, p. 1–82.

TAN, S.-L. Left out of the Indo-Pacific deal, China pushes toward the world's largest trade deal. **CNBC**, 2022. Disponível em: <<https://www.cnbc.com/2022/06/06/left-out-of-the-indo-pacific-deal-china-pushes-toward-rcep-trade-deal.html>>. Acesso em: 25 nov. 2022.

TANG, C.; TOMLIN, B. The power of flexibility for mitigating supply chain risks. **International Journal of Production Economics**, nov. 2008. v. 116, n. 1, p. 12–27. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0925527308002181>>.

TONY LIU, F.; MING TING, K.; ZHOU, Z.-H. Isolation Forest ICDM08. **Icdm**, 2008. Disponível em: <<https://cs.nju.edu.cn/zhoush/zhoush.files/publication/icdm08b.pdf%0Ahttps://cs.nju.edu.cn/zhoush/zhoush.files/publication/icdm08b.pdf?q=isolation-forest>>.

VOIGT, K.; MÜLLER, J. M. The Impact of Industry 4 . 0 on Supply Chains in Engineer-to-Order Industries The Impact of Industry 4 . 0 on Supply Chains in Engineer-to-Order Industries The on Supply Chains in Industries An Exploratory Study The Impact Impact of of Industry on Supply. **IFAC-PapersOnLine**, 2018. v. 51, n. 11, p. 122–127. Disponível em: <<https://doi.org/10.1016/j.ifacol.2018.08.245>>.

VONDRACEK, C. USDA sends \$700 million to pandemic-affected ethanol, biodiesel facilities. **StarTribune**, 2022. Disponível em: <<https://www.startribune.com/usda-sends-700-million-to-pandemic-impacted-ethanol-biodiesel-facilities/600179740/?refresh=true>>. Acesso em: 14 jan. 2023.

WANG, H.; HUANG, J.; ZHANG, Z. The impact of deep learning on organizational agility. In proceedings of the 40th International Conference on Information Systems (ICIS), Munich, Germany. 2019. Disponível em: <<https://aisel.aisnet.org/icis2019>>.

WWF. **The growth of soy: Impacts and solutions**. Gland: WWF International, 2014.